

Appendix: S3 Report on QSAR/QSPR models

Table of Content

Model name	Page
Prediction model of nanoparticles uptake by PaCa2 cells by MLR	9
Prediction model of nanoparticles uptake by PaCa2 cells by ANN	20
Methodology for developing structure-activity evaluation to identify combinations of physical features of nanomaterial that influence potential cell damage by MLR/LDA (TiO ₂ case)	31
Methodology for developing structure-activity evaluation to identify combinations of physical features of nanomaterial that influence potential cell damage by MLR/LDA (ZnO case)	37
Biological activity of manufactured nanoparticles by SVM	43
Nanoparticles uptake of PaCa2 cells model by MOE and kNN	51
Regression model to understand the aggregated ZVCN against E.Coli by MLR and a Quadratic model (Simplex centroid design)	62
Regression model to understand the aggregated ZVCN against E.Coli by MLR (Placket-Burman design)	68
Oxidative stress caused by metal oxides nanoparticles	74
Classification nanoSAR for the cytotoxicity of metal oxide nanoparticles for BEAS-2B by a Logistic Regression	82
Prediction of the Biological surface adsorption index (BSAI) on different NPs by MLR	88
Toxicological effects of Co-NPs/Co-ions on different cell types studied by KDD method: Decision tree (J48).	95
Experimental testing and computational modelling methodologies to study the cytotoxicity of metal oxide nanoparticles in E. coli	102
Model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	109
Consensus model to predict the Nanoparticles uptake of PaCa2 cells(it includes kNN, SVM, NB in the final consensus model)	117
Metal oxide toxicity prediction through their energy bands and their solubility, by decision trees	129
Toxicity prediction of decorated nanotubes by GFA (Study in a variety set of descriptors and endpoints.)	136
Metal oxide NPs toxicity classification by SVM	144
Bioactivity response (active/inactive) classification of coated iron oxide NPs by NBC	152
Predictive model of TiO ₂ NPs damage on membrane cell by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	160
Toxic Effect (Active/Inactive) prediction on embryonic zebrafish due wide type of nanomaterials by kNN under ABMiner software	168
Toxic Effect (Active/Inactive) prediction on embryonic zebrafish due wide type of nanomaterials and Weightened EZ Metric score as Endpoint by kNN under ABMiner software	174
Prediction model for nanoparticles uptake of PaCa2 cells by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	180

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure. (PMNs - RT case)	191
Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure. (PMNs - RF case)	197
Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure. (MAC - RT case)	204
Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure. (MAC - RF case)	210
Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure. (LDH -RT case)	216
Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure. (LDH - RF case)	223
Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure. (TP - RT case)	230
Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure. (TP - RF case)	237
Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions by a perturbational approach and LDA	244
Extended version of : Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions by a perturbational approach and LDA	253
Prediction model of nanoparticles uptake by PaCa2 cells byMold2 and kNN	267
The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles by Random Forest(E. Coli case)	278
The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles by Random Forest(HaCaT case)	286
Prediction of the Induction of apoptosis by metal oxide nanoparticles by MLR (SMA -MLR- case)	294
Prediction of the Induction of apoptosis by metal oxide nanoparticles by BRANN (SMA - BRANN- case)	302
Cellular uptake by CLIO NPs in several types of cells by BRANN (HUVEC case)	310
Cellular uptake by CLIO NPs in several types of cells by MLREM (PaCa2 case)	321
Ensemble Learning for predicting biological activity of diverse nanomaterials (cell apoptosis - Regression- case)	332
Ensemble Learning for predicting biological activity of diverse nanomaterials (cell apoptosis - Classification- case)	340
Ensemble Learning for predicting biological activity of diverse nanomaterials (PaCa2 - Regression- case)	348
Ensemble Learning for predicting biological activity of diverse nanomaterials (PaCa2 - Classification- case)	359
Ensemble Learning for predicting biological activity of diverse nanomaterials (E. Coli case)	370
Ensemble Learning for predicting biological activity of diverse nanomaterials(Toxicity prediction of decorated nanotubes)	377
Prediction model of nanoparticles uptake by PaCa2 cells by MLR plus PLS	385
Predictive model of Mutagenicity of fullerene C60 by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	396

Ecotoxicity model under three organo-coated silver NPs by GLM (E. coli case)	402
Ecotoxicity model under three organo-coated silver NPs by GLM (D. Magna case)	408
Cytotoxicity of metal oxide to bacteria E.Coli models by Periodic table-based descriptors and stepwise-MLR	414
Cytotoxicity of metal oxide to bacteria E.Coli models by Periodic table-based descriptors and PLS	420
cytotoxicity of SiO ₂ NPs on human lung fibroblast model by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	427
Photo-induced toxicity of metal oxide NPs to E. Coli by MLR (dark condition case)	434
Photo-induced toxicity of metal oxide NPs to E. Coli by MLR (Photo-induced (light) case)	440
quasi-QSPR model for photocatalytic decolourization rate constants by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	447
quasi-QSAR model for cellular viability by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	454
Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS (Arginase:iNOS case)	461
Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS (IL-10 case)	467
Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS (TNF- α case)	473
Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS (Cathepsin_3day case)	480
Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS (Cathepsin_7day case)	486
Metal Oxide NPs toxicity towards BEAS-2B and RAW 264.7 by RF and causal inference interpretation (BEAS-2B case)	492
Metal Oxide NPs toxicity towards BEAS-2B and RAW 264.7 by RF and causal inference interpretation (RAW264.7 case)	500
Toxicological modelling of diverse nanomaterials using the embryonic zebrafish (EZ) metric toxicity by multiple non-linear regression	508
Prediction of multiple antibacterial profiles of nanoparticles by a perturbational approach and LDA	514
Toxicity model of engineered zinc oxide nanoparticles to embryonic zebrafish by PCA and Kriging estimation	522
Modelling of membrane disruption mediated by ZnO and TiO ₂ NPs by linear regression method (MLR case for TiO ₂ and ZnO)	528
Modelling of membrane disruption mediated by TiO ₂ NPs by linear regression method (MLR case for TiO ₂)	535
Modelling of membrane disruption mediated by ZnO NPs by linear regression method (MLR case for ZnO)	542


Modelling of membrane disruption mediated by ZnO and TiO ₂ NPs by non-linear regression method (SVM-linear case for TiO ₂ and ZnO)	549
Modelling of membrane disruption mediated by TiO ₂ NPs by non-linear regression method (SVM-linear case for TiO ₂)	555
Modelling of membrane disruption mediated by ZnO NPs by non-linear regression method (SVM-linear case for ZnO)	561
Modelling of membrane disruption mediated by ZnO and TiO ₂ NPs by non-linear regression method (SVM-radial case for TiO ₂ and ZnO)	567
Modelling of membrane disruption mediated by TiO ₂ NPs by non-linear regression method (SVM-radial case for TiO ₂)	573
Modelling of membrane disruption mediated by ZnO NPs by non-linear regression method (SVM-radial case for ZnO)	579
Modelling of membrane disruption mediated by ZnO and TiO ₂ NPs by non-linear regression method (RBFNN case for TiO ₂ and ZnO)	585
Modelling of membrane disruption mediated by TiO ₂ NPs by non-linear regression method (RBFNN case for TiO ₂)	591
Modelling of membrane disruption mediated by ZnO NPs by non-linear regression method (RBFNN case for ZnO)	597
Modelling of membrane disruption mediated by ZnO and TiO ₂ NPs by non-linear regression method (GRNN case for TiO ₂ and ZnO)	603
Modelling of membrane disruption mediated by TiO ₂ NPs by non-linear regression method (GRNN case for TiO ₂)	609
Modelling of membrane disruption mediated by ZnO NPs by non-linear regression method (GRNN case for ZnO)	615
Modelling of membrane disruption mediated by ZnO and TiO ₂ NPs by classification tree method (Classification tree case for TiO ₂ and ZnO)	621
Toxicity model for metal oxide NPs by PLS	627
Cytotoxicity of metal oxide nanoparticles on E. coli prediction by nano-read-across (t-HCA)	633
Cytotoxicity of metal oxide nanoparticles on HaCaT cell line prediction by nano-read-across (t-HCA)	640
Prediction of Toxicity of metal oxide NPs to E. Coli with/without photo-inducing by SMILES-based optimal descriptor and Monte Carlo technique	646
modelling cell membrane damage of metal oxide nanoparticles to BEAS-2B by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	655
Prediction model of nanoparticles uptake by PaCa2 cells by GA-PLS plus MLR	664
NP-cell association based on corona proteins and physicochemical properties by SVR	675
NP-cell association based on corona proteins and physicochemical properties by MLR	684
Predictive model of Mutagenicity of multi-walled carbon-nanotubes by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	694
Predictive model of Mutagenicity of multi-walled carbon-nanotubes and fullerene by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)	702
Toxicity model of metal oxide nanoparticles to human keratinocyte cell line (HaCaT) by GA-MLR	710
Predictive model of Gold nanoparticles exocytosis on macrophages by PLSR	717
Metal oxide toxicity classification by GPTree	723

Prediction model of nanoparticles uptake by PaCa2 cells by GPtree - 1st pool of descriptors -	730
Prediction model of nanoparticles uptake by PaCa2 cells by GPtree - 2nd pool of descriptors -	741
Cytotoxicity keratinocyte (HaCaT) cell line model of metal oxide nanoparticles to human by GPtree	752
Predictive model of Gold nanoparticles exocytosis on macrophages by GPtree	759
Predictivity model ZnO nanoparticles effects on Cellular viability by MLREM	764
Predictivity model ZnO nanoparticles effects on Cellular viability by BRANNLP	770
Predictivity model ZnO nanoparticles effects on cell membrane damage by MLREM	776
Predictivity model ZnO nanoparticles effects on cell membrane damage by BRANNLP	782
Predictivity model ZnO nanoparticles effects on Oxidative Stress by MLREM	788
Predictivity model ZnO nanoparticles effects on Oxidative Stress by BRANNLP	794
Protein binding to CNTs by consensus model of kNN, SVM and RF	800
Cytotoxicity model of decorated CNTs by consensus model of kNN, SVM and RF	812
Size and cytotoxicity of chitosan/streptokinase prediction by ANN	824
Cytotoxicity of chitosan/streptokinase prediction by ANN	829
Cytotoxicity of chitosan/streptokinase prediction by ANN with size as solely dependent variable	834
Metal-based nanoparticle hazard classification by FT(Data set I)	839
Metal-based nanoparticle hazard classification by C4.5 Decision tree(Data set I)	847
Metal-based nanoparticle hazard classification by RT(Data set I)	854
Metal-based nanoparticle hazard classification by Simple CART(Data set I)	862
Metal-based nanoparticle hazard classification by FT(Data set II)	869
Metal-based nanoparticle hazard classification by C4.5 Decision tree(Data set II)	876
Metal-based nanoparticle hazard classification by RT(Data set II)	883
Metal-based nanoparticle hazard classification by Simple CART(Data set II)	891
Metal-based nanoparticle hazard classification by FT(Data set III)	898
Metal-based nanoparticle hazard classification by C4.5 Decision tree(Data set III)	905
Metal-based nanoparticle hazard classification by RT(Data set III)	911
Metal-based nanoparticle hazard classification by Simple CART(Data set III)	918
Metal-based nanoparticle hazard classification on Danio rerio by FT	924
Metal-based nanoparticle hazard classification on Danio rerio by C4.5 Decision tree	931
Metal-based nanoparticle hazard classification on Daphnia Magna by FT	938
Metal-based nanoparticle hazard classification on Daphnia Magna by C4.5 Decision tree	945
Metal-based nanoparticle hazard classification on Pseudokirchneriella subcapitata by FT	952
Metal-based nanoparticle hazard classification on Pseudokirchneriella subcapitata by C4.5 Decision tree	959
Metal-based nanoparticle hazard classification on Staphylococcus aureus by C4.5 Decision tree	966
Predicting metal oxide Nps toxicity to E. Coli cell line by PLS and pEC50_HaCaT toxicity as descriptor (Nano-QTTR)	972
Predicting metal oxide Nps toxicity to HaCaT cell line by PLS and pEC50_E.Coli toxicity as descriptor (Nano-QTTR)	980
Predicting metal oxide Nps toxicity to E. Coli cell line by optimal descriptors and CORAL software(Model 1 - without size descriptor case)	988

Predicting metal oxide Nps toxicity to E. Coli cell line by optimal descriptors and CORAL software(Model 2 - with size descriptor case)	995
Predicting metal oxide Nps toxicity to HaCaT cell line by optimal descriptors and CORAL software(Model 3 - without size descriptor case)	1002
Predicting metal oxide Nps toxicity to HaCaT cell line by optimal descriptors and CORAL software(Model 4 - with size descriptor case)	1009
Predicting metal oxide Nps toxicity to E. Coli cell line by MLR	1016
Predicting metal oxide Nps toxicity to E. Coli and HaCaT by SMILES-based optimal descriptor and Monte Carlo technique and CORAL software	1023
Bioactivity response (active/inactive) classification of coated iron oxide NPs by Decision tree (J48)	1031
Prediction model of nanoparticles uptake by PaCa2 cells by SVR	1039
Prediction model of nanoparticles uptake by PaCa2 cells by MLR	1050
Predicting the solubility of C-60 in various solvents by ANN	1061
Prediction of C60 solubilities from solvent molecular structures MLR case	1067
Prediction of C60 solubilities from solvent molecular structures CNN case	1073
Prediction of C60 solubilities from solvent molecular structures (CNN case 2)	1079
Predicting C60 solubility in organic solvents (alkanes) by MRA	1085
Predicting C60 solubility in organic solvents (alkyl halides) by MRA	1090
Predicting C60 solubility in organic solvents (alcohols) by MRA	1096
Predicting C60 solubility in organic solvents (cycloalkanes) by MRA	1101
Predicting C60 solubility in organic solvents (alkylbenzenes) by MRA	1106
Predicting C60 solubility in organic solvents (aryl halides) by MRA	1112
Predicting C60 solubility in organic solvents (Aliphatic) by MRA	1118
Predicting C60 solubility in organic solvents (Aliphatic_2) by MRA	1124
Predicting C60 solubility in organic solvents (Aromatics) by MRA	1130
Predicting C60 solubility in organic solvents (Aliphatic_3) by MRA	1136
Predicting solubility of C60 fullerene by LSER approach (Stepwise-MLR) - Solvents at 298K	1141
Predicting solubility of C60 fullerene by LSER approach (Stepwise-MLR) - Solvents at 303K	1148
Predicting solubility of C60 fullerene by LSER approach (Stepwise-MLR) All solvents	1154
QSPR model to predict the solubility of C60 various solvents; Linear model	1160
QSPR model to predict the solubility of C60 various solvents, Nonlinear model	1166
Predicting Young's modulus by correlation weighting by SMILES-based optimal descriptor and Monte Carlo technique	1172
Predicting C60 solubility in organic solvents by SMILES-based optimal descriptor and Monte Carlo technique	1178
Predicting C60 solubility in organic solvents by SMILES-based optimal descriptor and Monte Carlo technique	1184
Model for predicting water solubility for CNT based on the chiral vector by MLR	1190
Model for predicting octanol water partition coefficient for CNT based on the chiral vector by MLR	1195
Predicting C60 solubility in organic solvents by SMILES-based optimal descriptor and Monte Carlo technique	1200

Predicting C60 solubility in organic solvents by means of a molecular-based model GA-MLR model	1207
Predicting C60 solubility in organic solvents by means of a molecular-based model FFNN model	1213
Binding interactions between fullerene inhibitors and HIV-1 PR residues using molecular docking and molecular dynamics simulations, by 3D QSAR applying PLS	1219
Predicting C60 solubility in organic solvents by InChI-based optimal descriptors and Monte-Carlo optimization	1225
Predicting C60 - C70 solubility in chlorobenzene by SMILES-based optimal descriptor and Monte Carlo technique	1232
Predicting C60 solubility in organic solvents based on quantum-chemical and topological descriptors by GA-MLR	1238
Young's modulus prediction in CNTs (taking into account vacancies) by PLS	1244
Poisson's ratio prediction in CNTs (taking into account vacancies) by PLS	1250
Young's modulus prediction in CNTs (taking into account vacancies plus methyl functionalization) by PLS	1256
Poisson's ratio prediction in CNTs (taking into account vacancies plus methyl functionalization) by PLS	1262
LSER model to predict the Solubility of fullerene C60 in various solvents by stepwise-MLR	1268
Predicting phase behavior of amphiphilic nanostructured nanoparticle drug delivery vehicles by BRANN (phytantriol batch 1 case)	1274
Predicting phase behavior of amphiphilic nanostructured nanoparticle drug delivery vehicles by BRANN (phytantriol batch 2 case)	1280
Predicting phase behavior of amphiphilic nanostructured nanoparticle drug delivery vehicles by BRANN (monoolein case)	1286
Predicting phase behavior of amphiphilic nanostructured nanoparticle drug delivery vehicles by BRANN (Myverol [monoolein with impurities] case)	1292
Predicting dispersion of SWNTs in different organic solvents by MLR	1298
Predicting dispersion of SWNTs in different organic solvents by ANN	1304
Predicting dispersion of SWNTs in different organic solvents by PLS	1310
Predict efficiency of methylene blue (MB) adsorption onto CuO-NP-AC by PCA-MLR	1316
Predict efficiency of methylene blue (MB) adsorption onto CuO-NP-AC by PCA-LSSVM	1323
NP-cell association based on corona proteins by PLSR	1330
Predicting dispersion of SWNTs in different organic solvents by GA-MLR	1339
NP-cell association based on physicochemical properties of protein corona by MLR	1345
Predicting C60 - C70 solubility in chlorobenzene by GA-MLR	1355
Consensus model to predict Zeta potential of metal oxide nanoparticle by GA-MLR	1361
Predicting C60 - C70 solubility in chlorobenzene by PLS	1368
NP-cell association based on corona proteins and physicochemical by machine learning approaches (MLR case)	1374
NP-cell association based on corona proteins and physicochemical by machine learning approaches (k-NN case)	1383
NP-cell association based on corona proteins and physicochemical by machine learning approaches (GRegNN case)	1392

NP-cell association based on corona proteins and physicochemical by machine learning approaches (RBFNN case)	1401
NP-cell association based on corona proteins and physicochemical by machine learning approaches (CPANN case)	1410
NP-cell association based on corona proteins and physicochemical by machine learning approaches (SVM-radial case)	1419
NP-cell association based on corona proteins and physicochemical by machine learning approaches (SVM-linear case)	1428
NP-cell association based on corona proteins and physicochemical by machine learning approaches (PLS case)	1437
NP-cell association based on corona proteins and physicochemical by machine learning approaches (PPR case)	1446
NP-cell association based on corona proteins and physicochemical by machine learning approaches (EARTH case)	1455
NP-cell association based on corona proteins and physicochemical by machine learning approaches (RF-6 case)	1464
NP-cell association based on corona proteins and physicochemical by machine learning approaches (RF-150 case)	1473

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells by MLR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Mehdi Ghorbanzadeh

m.ghorbanzade@umz.ac.ir

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software package:

Ghorbanzadeh M, Fatemi MH, Karimpour M. 2012. Modelling the cellular uptake of magnetofluorescent nanoparticles in pancreatic cancer cells: a quantitative structure activity relationship study. Ind Eng Chem Res 51:10712–18.

(MLR case)

<http://doi.org/10.1021/ie3006947>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3.Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- number of donor atoms for H-bonds (N and O) (nHDon)
- Geary autocorrelation of lag 1 weighted by van der Waals volume (GATS1v)
- 3D-MoRSE-signal 29/unweighted (Mor29u)
- D total accessibility index/weighted by Sanderson electronegativity (De)
- 3D-MoRSE-signal 14/unweighted (Mor14u)
- Mean electrotopological state (Ms); 6

4.4.Descriptor selection:

Hyperchem v7 was applied to construct all the molecular structures (geometry optimization by the Austin Model 1) and Dragon software was used to generate an initial set of descriptors.

Calculated descriptors were analyzed for the existence of a constant or near constant values, and descriptors with low variation were removed.

Multiple Linear Regression MLR

- Self-organising mapping network was used to eliminate redundant descriptors

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

90/6

Descriptor: Chemical ratio :1:15

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

The first approach was based on the ranges of individual descriptors used for the building model:

$2.23 \leq \text{pM} < 4.44$

$0 \leq \text{nHDon} \leq 8$

$0.52 \leq \text{GATS1v} \leq 1.73$

$-0.64 \leq \text{Mor28u} \leq 0.35$

$0.36 \leq \text{De} \leq 0.62$

$-1.04 \leq \text{Mor14u} \leq 4.24$

$1.05 \leq \text{Ms} \leq 5.27$

For specific details see (in the publication) Table 2

The second one it was the verification by the leverage approach and Williams plot.

$h^* < 0.263$

(it should be 0.233)

Any data detected as outlier.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

90 Metal Oxide

List: $(\text{Fe}_2\text{O}_3)_n(\text{Fe}_3\text{O}_4)_m$

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride
 4 3,3-Dimethyldihydrofuran-2,5-dione
 Furan-2,5-dione
 3-Methylfuran-2,5-dione
 7 3,4-Dimethylfuran-2,5-dione
 Hexanoic anhydride
 3-Methyldihydrofuran-2,5-dione
 5,5'-Carbonylbis(2-benzofuran-1,3-dione)
 5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride

5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid

NCCCCC(N)C(O)=O

Amino(4-chlorophenyl)acetic acid

NC(C(O)=O)c1ccc(Cl)cc1

2-Aminopropanoic acid

2-Amino-5-carbamimidamidopentanoic acid

2-Aminobutanedioic acid

2,5-Diamino-5-oxopentanoic acid

2-Aminopentanedioic acid

2-Amino-3-(1Himidazol-4-yl)propanoic acid

2-Amino-4-(methylsulfanyl)butanoic acid

2-Amino-3-phenylpropanoic acid

Dihydrofuran-2,5-dione

Acetic anhydride

3-Methylidenedihydrofuran-2,5-dione

1,4-Dioxane-2,6-dione

2-Benzofuran-1,3-dione

(2,5-Dioxotetrahydrofuran-3-yl)acetic acid

4,7-Difluoro-2-benzofuran-1,3-dione

{Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

Overall size (volume weighted) in aqueous solution.

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

SMILES notation for all the coating are reported in Table S1 of publication's supplementary material.

6.6.Pre-processing of data before modelling:

The compounds in the dataset were randomly divided into training, internal and external data sets through diversity analysis (through the concept of the distance to have a homogeneous distribution of the training and test sets)

6.7.Statistics for goodness-of-fit:

$R_{(MLR,Train)}=0.782$

$R_{(MLR,Predict)}=0.755$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

$Q^2_{MLR}=0.577$

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

10 Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine

2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione

1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

Overall size (volume weighted) in aqueous solution.

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

SMILES notation for all the coating are reported in Table S1 of publication's supplementary material.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Proposed ANN vs the MLR.

Lack in rigurosity of the external validation methodology, predictive power of the model and the treatment of the statistical obtained data.

R_(MLR,Train/Ext_test) : square root of regression coefficient

Q²_MLR: leave-many-out cross validation correlation coefficient

MLR :Multiple Linear Regression

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

(already reported in this table)

Fourches, D. et al., 2010. Quantitative nanostructure-activity relationship modelling. *ACS nano*, 4(10), pp.5703–12 (Case Study 2)

10.Summary (JRC QSAR Model Database)
10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - number of donor atoms for H-bonds (N and O) (nHDon)

- Geary autocorrelation of lag 1 weighted by van der Waals volume (GATS1v)
- 3D-MoRSE-signal 29/unweighted (Mor29u)
- D total accessibility index/weighted by Sanderson electronegativity (De)
- 3D-MoRSE-signal 14/unweighted (Mor14u)
- Mean electrotopological state (Ms),MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells by ANN
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells by ANN

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Mehdi Ghorbanzadeh

m.ghorbanzade@umz.ac.ir

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software package:

Ghorbanzadeh M, Fatemi MH, Karimpour M. 2012. Modelling the cellular uptake of magnetofluorescent nanoparticles in pancreatic cancer cells: a quantitative structure activity relationship study. *Ind Eng Chem Res* 51:10712–18.

(ANN case)

<http://doi.org/10.1021/ie3006947>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3. Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLP-NN: Multilayered Perceptron Neural Network

4.3.Descriptors in the model:

- number of donor atoms for H-bonds (N and O) (nHDon)
- Geary autocorrelation of lag 1 weighted by van der Waals volume (GATS1v)
- 3D-MoRSE-signal 29/unweighted (Mor29u)
- D total accessibility index/weighted by Sanderson electronegativity (De)
- 3D-MoRSE-signal 14/unweighted (Mor14u)
- Mean electrotopological state (Ms); 6

4.4.Descriptor selection:

Hyperchem v7 was applied to construct all the molecular structures (geometry optimization by the Austin Model 1) and Dragon software was used to generate an initial set of descriptors.

Calculated descriptors were analyzed for the existence of a constant or near constant values, and descriptors with low variation were removed.

Multiple Linear Regression MLR

- Self-organising mapping network was used to eliminate redundant descriptors

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

90/6

Descriptor: Chemical ratio :1:15

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The first approach was based on the ranges of individual descriptors used for the building model:

$2.23 \leq \text{pM} < 4.44$

$0 \leq \text{nHDon} \leq 8$

$0.52 \leq \text{GATS1v} \leq 1.73$

$-0.64 \leq \text{Mor28u} \leq 0.35$

$0.36 \leq De \leq 0.62$

$-1.04 \leq Mor14u \leq 4.24$

$1.05 \leq Ms \leq 5.27$

For specific details see (in the publication) Table 2

The second one it was the verification by the leverage approach and Williams plot.

$h^* < 0.263$

(it should be 0.233)

Any data detected as outlier.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

90 Metal Oxide

List: $(Fe_2O_3)_n(Fe_3O_4)_m$

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione
 Furan-2,5-dione
 3-Methylfuran-2,5-dione
 7 3,4-Dimethylfuran-2,5-dione
 Hexanoic anhydride
 3-Methyldihydrofuran-2,5-dione
 5,5'-Carbonylbis(2-benzofuran-1,3-dione)
 5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione

Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
 NCCCCC(N)C(O)=O

Amino(4-chlorophenyl)acetic acid
NC(C(=O)O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

Overall size (volume weighted) in aqueous solution.

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

SMILES notation for all the coating are reported in Table S1 of publication's supplementary material.

6.6.Pre-processing of data before modelling:

The compounds in the dataset were randomly divided into training, internal and external data sets through diversity analysis (through the concept of the distance to have a homogeneous distribution of the training and test sets)

6.7.Statistics for goodness-of-fit:

R_(MLP-NN,Train)=0.934

R_(MLP-NN,Int_test)=0.943

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

$Q^2_{MLP-NN}=0.655$

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

10 Metal Oxide

List

$(Fe_2O_3)_n(Fe_3O_4)_m$

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic dianhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.02,6]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine

2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione

2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

Overall size (volume weighted) in aqueous solution.

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

SMILES notation for all the coatings are reported in Table S1 of publication's supplementary material.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R_{(MLP-NN, Ext_test)} = 0.945$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Proposed ANN vs the MLR.

Lack in rigour of the external validation methodology, predictive power of the model and the treatment of the statistical obtained data.

$R_{(MPL-NN, Predict/Train/Ext_test)}$: regression coefficient

Q^2_{MPL-NN} : leave-many-out cross validation correlation coefficient

MLP-NN: Multilayered perceptron Neural Network

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

(already reported in this table)

Fourches, D. et al., 2010. Quantitative nanostructure-activity relationship modelling. *ACS nano*, 4(10), pp.5703–12 (Case Study 2)

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - number of donor atoms for H-bonds (N and O) (nHDon)

- Geary autocorrelation of lag 1 weighted by van der Waals volume (GATS1v)
- 3D-MoRSE-signal 29/unweighted (Mor29u)
- D total accessibility index/weighted by Sanderson electronegativity (De)
- 3D-MoRSE-signal 14/unweighted (Mor14u)
- Mean electrotopological state (Ms),MLP-NN: Multiplelayered Perceptron Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Methodology for developing structure-activity evaluation to identify
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Methodology for developing structure-activity evaluation to identify combinations of physical features of nanomaterial that influence potential cell damage by MLR/LDA
(TiO₂ case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Christie Sayes

csayes@cvm.tamu.edu

2.6. Date of model development and/or publication:

2010

2.7. Reference(s) to main scientific papers and/or software package:

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

(TiO₂ case)

<http://doi.org/10.1111/j.1539-6924.2010.01438.x>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

$y < 0,99$ --> Dense cell membrane
 $0,99 < y < 1,09$ --> Normal cell membrane
 $1,09 < y < 1,25$ --> Leaky cell membrane
 $1,25 < y$ --> Disrupted cell membrane
 LDA classifies in Dense and Disrupted

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Apply first the MLR: Multiple Linear Regression
 once realized that the correlation coefficient (R^2) is not enough to model the data,
 it was applied LDA: Linear Discrimination Analysis.

4.3.Descriptors in the model:

- Size in water
- Concentration [mg/L]
- Zeta Potential [mV]; 3

4.4.Descriptor selection:

Normalization and standardizing the data by individual variances.

PCA and correlation matrix to identify multicollinearity of initial descriptors.

In the LDA case: Check all possible combinations of 1, 2 or 3 variables/descriptors. Select combination which minimizes the resubstitution error.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

24/3

Descriptor: Chemical ratio :01:08

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

We can apply the range of the final descriptors in the training data.

Metal oxides related with: TiO₂

For specific details see (in the publication) Table I

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

24 Metal Oxide

List: TiO₂

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: Anastase/Rutile crystal structures

To reduce particle settlement, Tween 20 (~1% v/v) was added to each

nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

24 measures, combination of:

Engineered Size (30, 45, 125)

2 x Concentration(25, 50, 100, 200)

No splitting data. Considered as not enough experimental data.

For LDA, only the data classified in the two classes (Dense and Disrupted) were chosen to be used.

6.7.Statistics for goodness-of-fit:

MLR : $0.15 < R^2 < 0.7$

LDA: $E_{sub} = 0$

Internal validation: Resubstitution error: error rate obtained from training data

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA Metal Oxide

ListTiO₂Shape:NACoating:NASize(nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

Anastase/Rutile crystal structures

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1. Comments:

This publication is focused in the methodology, more than in the final results, with huge future work specified in the text.

Mechanistic Interpretation

The presence or absence of Zeta potential as descriptor in both sets is argued.

R^2 : correlation coefficient

MLR: multiple linear regression

LDA: linear discrimination analysis

CCM: culture cell media

Esub: Resubstitution error

PBS: phosphate buffered saline

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Immortalized rat L2 lung epithelial cells

and


rat lung alveolar macrophages, QSAR, - Size in water

- Concentration [mg/L]

- Zeta Potential [mV], Apply first the MLR: Multiple Linear Regression

once realized that the correlation coefficient (R^2) is not enough to model the data, it was applied LDA: Linear Discrimination Analysis.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Methodology for developing structure-activity evaluation to identify
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Methodology for developing structure-activity evaluation to identify combinations of physical features of nanomaterial that influence potential cell damage by MLR/LDA
(ZnO case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Christie Sayes

csayes@cvm.tamu.edu

2.6. Date of model development and/or publication:

2010

2.7. Reference(s) to main scientific papers and/or software package:

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

(ZnO case)

<http://doi.org/10.1111/j.1539-6924.2010.01438.x>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

$y < 0,99$ --> Dense cell membrane
 $0,99 < y < 1,09$ --> Normal cell membrane
 $1,09 < y < 1,25$ --> Leaky cell membrane
 $1,25 < y$ --> Disrupted cell membrane
 LDA classifies in Dense and Disrupted

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Apply first the MLR: Multiple Linear Regression

once realized that the correlation coefficient (R^2) is not enough to model the data,

it was applied LDA: Linear Discrimination Analysis.

4.3.Descriptors in the model:

- Size in water
- Concentration [mg/L]
- Size in CCM; 3

4.4.Descriptor selection:

Normalization and standardizing the data by individual variances.

PCA and correlation matrix to identify multicollinearity of initial descriptors.

In the LDA case: Check all possible combinations of 1, 2 or 3 variables/descriptors. Select combination which minimizes the resubstitution error.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

18/3

Descriptor: Chemical ratio :01:06

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

We can apply the range of the final descriptors in the training data.

Metal oxides related with: ZnO

For specific details see (in the publication) Table III

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

18 Metal Oxide

List: ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 50, 60, 70, 1000, 1200, 1500

Size in water: 55-1283

Size in PBS: 158-2109

Size in CCM: 107-1578

Other info: To reduce particle settlement, they added Tween 20 (~1% v/v) to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

18 measures, combination of:

Engineered Size (50, 60, 70, 1000, 1200, 1500)

Concentration(25, 50, 100)

No splitting data. Considered as not enough experimental data.

For LDA, only the data classified in the two classes (Dense and Disrupted) were chosen to be used.

6.7.Statistics for goodness-of-fit:

MLR : $0.19 < R^2 < 0.49$

LDA: $E_{sub} = 0$

Internal validation: Resubstitution error: error rate obtained from training data

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA Metal Oxide

List

ZnO

Shape:NA

Coating:NA

Size(nm): Engineered Size: 50, 60, 70, 1000 , 1200, 1500

Size in water: 55-1283

Size in PBS: 158-2109

Size in CCM: 107-1578

Other properties:

To reduce particle settlement, they added Tween 20 (~1% v/v) to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

This publication is focused in the methodology, more than in the final

results, with huge future work specified in the text.

Mechanistic Interpretation

The presence or absence of Zeta potential as descriptor in both sets is argued.

R^2 : correlation coefficient

MLR: multiple linear regression

LDA: linear discrimination analysis

CCM: culture cell media

Esub: Resubstitution error

PBS: phosphate buffered saline

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Immortalized rat L2 lung epithelial cells
and


rat lung alveolar macrophages, QSAR, - Size in water

- Concentration [mg/L]

- Size in CCM, Apply first the MLR: Multiple Linear Regression

once realized that the correlation coefficient (R^2) is not enough to model the data,
it was applied LDA: Linear Discrimination Analysis.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Biological activity of manufactured nanoparticles by SVM
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Biological activity of manufactured nanoparticles by SVM

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Alexander Tropsha

alex_tropsha@unc.edu

2.6. Date of model development and/or publication:

2010

2.7. Reference(s) to main scientific papers and/or software package:

Fourches, D. et al., 2010. Quantitative nanostructure-activity relationship modelling. ACS nano, 4(10), pp.5703–12

Case Study 1

<http://doi.org/10.1021/nn1013484>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)

Murine RAW 264.7 leukemic monocyte/macrophage cell

3.2. Endpoint:

In vitro - Cytotoxicity - measured by different biological activities

3.3.Comment on endpoint:

Biological activity was defined as the arithmetic mean of in vitro tests on four different cell lines, Using four doses, and four different assays of cellular physiology.

The four assays measured

- (i) ATP content,
- (ii) Reducing equivalents,
- (iii) Caspase-mediated apoptosis,
- (iv) Mitochondrial membrane potential.

Biological activity profiles were recorded for the following concentrations of MNPs: 0.01, 0.03, 0.1, and 0.3 mg/mL for all iron-based MNPs; and 3, 10, 30, and 100 nM for the three quantum dot-based MNPs

Vector of 64 measurements reduced to one value, Z_{mean} .

Classifying the endpoint in class 1 ($Z_{\text{mean}} \geq -0.40$) or class 0 ($Z_{\text{mean}} < -0.40$)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM classification : Support vector machine classification
by WinSVM (version 1.1.8)

4.3.Descriptors in the model:

- nanoparticle size
- r1: spin-lattice Relaxivity
- r2: spin-spin Relaxivity
- zeta potential (surface charge); 4

4.4.Descriptor selection:

Experimental results. Due the disponibility of the descriptors, the initial 51 NPs were reduced to 44 NPs

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

35/4

Descriptor: Chemical ratio :1:11

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Defined as (DT), the Euclidean distance between the query NP and its

5-NN in the Training Set.

$DT = Y_{average} + Z \cdot \sigma$.

Yaverage: average of Euclidean distance.

Sigma: the Standard Deviation of the average.

Z: is a constant set to 0.5.

This means that if query compound is >1.5 times away from SD of the average of Euclidean distance, the query compound is considered out of domain.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

35 Metal

Metal Oxide

List: Fe₂O₃

Fe₃O₄

Shape: NA

Coating: Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂
 Cross-linked dextran :: Alexa Fluor 488
 Cross-linked dextran :: Alexa Fluor 750
 Cross-linked dextran :: FITC, R-COOH
 Cross-linked dextran :: biotin
 Cross-linked dextran :: FITC, COOH
 Cross-linked dextran :: Cy3.5
 Cross-linked dextran :: Cy5.5, protamine
 Cross-linked dextran :: Cy5.5, tat
 Cross-linked dextran :: Cy5.5
 Cross-linked dextran :: Cy5
 Cross-linked dextran :: Cy7
 Cross-linked dextran :: FITC
 Cross-linked dextran :: FITC, Glutamic acid
 Cross-linked dextran :: glycine
 Cross-linked dextran :: rhodamine, protamine
 Cross-linked dextran :: FITC, succinimidyl iodoacetate
 Cross-linked dextran :: Tat peptide
 Cross-linked dextran :: VT680
 Cross-linked dextran :: VT680, protamine
 Dextran :: NA
 Sucrose :: NA
 PVA :: COOH
 PVA :: Ethylene diamine
 PVA :: Ethylene diamine, VT680
 PVA :: protamine, rhodamine
 PVA :: L-arg8-COOH
 PVA :: COOH
 PVA :: AminoSPARK™680
 PVA :: PEG Ethylene diamine, AminoSPARK™680
 PVA :: Ethylene diamine, AminoSPARK™680
 PVA, PEG :: AngioSPARK™680- IVM
 PVA :: 15-mer peptide
 PVA :: L-arg7-COOH
 PVA :: Ethylene diamine, VT750
 PVA, PEG :: Ethylene diamine, VT750
 PVA :: D-arg7-COOH
 PVA, PEG :: NA
 Arabino-galactan :: NA
 Carboxymethyldextran :: NA
 Amphiphilic polymer – PEG :: NH₂
 Amphiphilic polymer :: COOH
Size (nm): 20-74

Other info: Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

6.6.Pre-processing of data before modelling:

5-fold external CV and 5-fold CV internal (modelling set) was applied. Data was divided 5 times into Modelling set (80% of data) and External test set (20% of data). Each MNP was included into a validation set only once, allowing us to calculate the overall external prediction accuracy.

Modelling sets were divided multiple times into test and training sets to build internal models. These models were evaluated with 5-fold CV. The models with good scores for training and test set were used to evaluate each external test set.

6.7.Statistics for goodness-of-fit:

Accuracy: 0.73

Sensitivity: 0.60

Specificity: 0.86

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

5-fold-CV applied

Y-randomization was applied and showed robustness of the models (Internal application).

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal

Metal Oxide

List

Fe₂O₃

Fe₃O₄

Shape:NA

Coating:Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750

Cross-linked dextran :: FITC, R-COOH

Cross-linked dextran :: biotin

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: Cy3.5

Cross-linked dextran :: Cy5.5, protamine

Cross-linked dextran :: Cy5.5, tat

Cross-linked dextran :: Cy5.5

Cross-linked dextran :: Cy5

Cross-linked dextran :: Cy7

Cross-linked dextran :: FITC

Cross-linked dextran :: FITC, Glutamic acid

Cross-linked dextran :: glycine

Cross-linked dextran :: rhodamine, protamine

Cross-linked dextran :: FITC, succinimidyl iodoacetate

Cross-linked dextran :: Tat peptide

Cross-linked dextran :: VT680

Cross-linked dextran :: VT680, protamine

Dextran :: NA

Sucrose :: NA

PVA :: COOH

PVA :: Ethylene diamine

PVA :: Ethylene diamine, VT680

PVA :: protamine, rhodamine

PVA :: L-arg8-COOH

PVA :: COOH

PVA :: AminoSPARK™680

PVA :: PEG Ethylene diamine, AminoSPARK™680

PVA :: Ethylene diamine, AminoSPARK™680

PVA, PEG :: AngioSPARK™680- IVM

PVA :: 15-mer peptide

PVA :: L-arg7-COOH

PVA :: Ethylene diamine, VT750

PVA, PEG :: Ethylene diamine, VT750

PVA :: D-arg7-COOH

PVA, PEG :: NA

Arabino-galactan :: NA

Carboxymethyldextran :: NA

Amphiphilic polymer – PEG :: NH₂

Amphiphilic polymer :: COOH

Size(nm): 20-74

Other properties:

Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Data was split into two classes: $Z \geq -0.40$ and $Z < -0.40$ (Being Z the average activity). Clustering was performed. NPs with the same core were clustered together.

They call external set the data which should be called test set, since when you use k-fold cross validation your tests set were be used as training set in other of the cycles, thus you have to consider it as an internal set.

(additionally , for the training set an other 5-foldCV was applied, but only the better result was provided for the test set of such cycle of the "external" 5-fold-CV)

The classification models were considered acceptable if:

$CCR_{CV} \geq 0.6$ and

$CCR_{test} \geq 0.6$

Where: $CCR = 0.5 * (Sensitivity + Specificity)$

MNP: Manufactured Nanoparticles

SVM: Support Vector Machine

CV: Cross Validation

CLIO: Cross-Linked Iron Oxide

Qt-dot: Quantum dots

PNP: Pseudocaged Nanoparticle

MION: Monocrystalline Iron Oxide

Nanoparticle

CCR: Correct Classification Rate.

9.2.Bibliography:

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21), 7387–7392. <http://doi.org/10.1073/pnas.0802878105>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)


Murine RAW 264.7 leukemic monocyte/macrophage cell, QSAR, - nanoparticle size

- r1: spin-lattice Relaxivity

- r2: spin-spin Relaxivity

- zeta potential (surface charge),SVM classification : Support vector machine classification by WinSVM (version 1.1.8)

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Nanoparticles uptake of PaCa2 cells model by MOE and kNN
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Nanoparticles uptake of PaCa2 cells model by MOE and kNN

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Alexander Tropsha

alex_tropsha@unc.edu

2.6. Date of model development and/or publication:

2010

2.7. Reference(s) to main scientific papers and/or software package:

Fourches, D. et al., 2010. Quantitative nanostructure-activity relationship modelling. ACS nano, 4(10), pp.5703–12

Case Study 2

<http://doi.org/10.1021/nn1013484>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3. Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4. Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

kNN: k-Nearest Neighbour

4.3.Descriptors in the model:

- Two-dimensional MOE descriptors calculated for the NPs decorations only (software CCP).
- Physical properties
- Surface areas
- Atom count
- Bond count
- Kier & Hall connectivity indices
- Kappa shape indices
- Adjacency and distance matrix descriptors
- Pharmacophore feature descriptors
- Molecular charges

For specific details see (in the publication) Table S4; 10

4.4.Descriptor selection:

Two-dimensional MOE descriptors (Molecular Operating Environment, commercial software distributed by Chemical computing Group).

The top-10 most frequently descriptors in each individual fold and the average frequency are listed in the supplementary material

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

87/10

Descriptor: Chemical ratio :10:87 ~ 1:9

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Defined as (DT), the Euclidean distance between the query NP and its 5-NN in the Training Set.

$DT = Y_{average} + Z \cdot \sigma$.

Yaverage: average of Euclidean distance.

Sigma: the Standard Deviation of the average.

Z: is a constant set to 0.5.

This means that if query compound is >1.5 times away from SD of the average of Euclidean distance, the query compound is considered out of domain.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

87 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione

3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(=O)O
 Amino(4-chlorophenyl)acetic acid
NC(C(=O)O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid

2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

SMILES notation for all the coating are reported in Table S1 of publication's supplementary material

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

5-fold external CV and 5-fold CV internal (modelling set) was applied. Data was divided 5 times into Modelling set (80% of data) and External test set (20% of data). Each MNP was included into a validation set only once, allowing us to calculate the overall external prediction accuracy.

6.7.Statistics for goodness-of-fit:

No AD:

$R^2=0.72$

MAE=0.18

With AD:

(80% molecules within AD):

$R^2=0.77$

MAE=0.17

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

5-fold CV validation.

Y-randomization was applied and showed robustness of the models (Internal application).

No statistically significant models were retrieved.(Not provided value)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape:NA

Coating:Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.02,6]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine

2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione

2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

SMILES notation for all the coatings are reported in Table S1 of publication's supplementary material

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

To enable model interpretation, they identified descriptors that occurred most frequently in kNN models with the highest prediction accuracy. The top-10 most frequently selected descriptors in each individual fold and the averaged frequency across five folds are listed in Supplementary Materials (SM_Tables S3 and S4).

They call external set the data which should be called test set, since when you use k-fold cross validation your test set were be used as training set in other of the cycles, thus you have to consider it as an

internal set.

R²: Correlation coefficient

MAE: Mean absolute Error

AD: Applicability Domain

CV: Cross Validation

kNN: k-Nearest Neighbour

9.2. Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC


10.3. Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - Two-dimensional MOE descriptors calculated for the NPs decorations only (software CCP).

- Physical properties
- Surface areas
- Atom count
- Bond count
- Kier & Hall connectivity indices
- Kappa shape indices
- Adjacency and distance matrix descriptors
- Pharmacophore feature descriptors
- Molecular charges

For specific details see (in the publication) Table S4, kNN: k-Nearest Neighbour

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression model to understand the aggregated ZVCN against E.Coli
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression model to understand the aggregated ZVCN against E.Coli by MLR and a Quadratic model

(Simplex centroid design)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Vishal Shah

ShahV@dowling.edu

2.6. Date of model development and/or publication:

2010

2.7. Reference(s) to main scientific papers and/or software package:

Rispoli, F., Angelov, A., Badia, D., Kumar, A., Seal, S., & Shah, V. (2010). Understanding the toxicity of aggregated zero valent copper nanoparticles against Escherichia coli. Journal of Hazardous Materials, 180(1-3), 212–216.

<http://doi.org/10.1016/j.jhazmat.2010.04.016>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as percentage of dead E. Coli population

3.3.Comment on endpoint:

To determine the toxicity :

E. coli, the culture was grown in nutrient broth medium overnight at 30 °C, 200 rpm. Nutrient broth was prepared in 100mM phosphate buffer with the pH under study and then diluted to a 5mM concentration. E. coli was diluted as described in Tables 1 and 2, and 1mL of the inoculum were added to 49mL of nutrient medium in 125mL Erlenmeyer flask containing ZVCN.

The flasks (with ZVCN and E. Coli already prepared) were incubated for 30 min at different temperatures and aeration rate. Serial dilution and plating was carried out at the end of the experiment and agar plates incubated at 30 °C for 24 h to determine the colony forming units (CFUs). The percent toxicity of nanoparticles was determined by comparing the number of CFU present in the media after the incubation as compared to the number of CFU at time zero.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

and

"Quadratic model"

4.3.Descriptors in the model:

- pH
- Temperature
- Aertion rate
- Concentration of nanoparticles
- Concentration of bacteria; 5

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

11/5

Descriptor: Chemical ratio :5:11 ~ 1:2

(10:11 ~ 1:1 in the "Quadratic model")

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

ZVCN with size verage of 25 nm

Range (2-60)

Temperature(°C) from 30 to 44

Aeration(rpm) from 0 to 400

Concentration(ppm) of nanoparticles from 0 to 1250

Concentration of bacteria (in % of v) from 0.1 to 10.1

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

11 Metal

List: ZVCN: zero valent copper Cu nanoparticle

Shape: Spherical

Coating: NA

Size (nm): Average of 25

Range of 2-60

Other info: Dynamic light scattering (DLS) was used to measure the rate of ZVCN aggregation at various temperatures and pH. Particle size measurements were carried out using a light scattering technique from Zeta Sizer Nano (Malvern Instruments). The instrument uses a 633nm wavelength laser to measure the size distribution of suspended particles.

Surface areas of 30-50 m²/g

6.6.Pre-processing of data before modelling:

All the data was obtained from experimental results, extra 3 experiments were developed to test the "Quadratic model"

6.7.Statistics for goodness-of-fit:

MLR: R²= 0.57

"Quadratic model":

R²=0.99

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

3 MMetal

List

ZVCN: zero valent copper Cu nanoparticle

Shape: Spherical

Coating: NA

Size(nm): Average of 25

Range of 2-60

Other properties:

Dynamic light scattering (DLS) was used to measure the rate of ZVCN aggregation at various temperatures and pH. Particle size measurements were carried out using a light scattering technique from Zeta Sizer Nano (Malvern Instruments). The instrument uses a 633nm wavelength laser to measure the size distribution of suspended particles.

Surface areas of 30-50 m²/g

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

"Quadratic model":

From external experiments (three): the average of difference of the predicted toxicity and the measured one is: 11% against 10.5% of standard error regression.

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

In this paper a linear regression model for few experimental results were developed, and a "Quadratic model" was also applied.

The last one was not clearly explained, and an overfitting could be present.

Due to some empty classification fields and the fact that the descriptors are not physicochemical or structural properties of the NP, it could be not considered as QSAR after all.

ZVCN: zero valent copper nanoparticle

MLR: multiple linear regression

R²: Correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - pH

- Temperature

- Aeration rate


- Concentration of nanoparticles

- Concentration of bacteria,MLR: Multiple Linear Regression

and

"Quadratic model"

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression model to understand the aggregated ZVCN against E.Coli
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression model to understand the aggregated ZVCN against E.Coli by MLR
(Placket-Burman design)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Vishal Shah

ShahV@dowling.edu

2.6. Date of model development and/or publication:

2010

2.7. Reference(s) to main scientific papers and/or software package:

Rispoli, F., Angelov, A., Badia, D., Kumar, A., Seal, S., & Shah, V. (2010). Understanding the toxicity of aggregated zero valent copper nanoparticles against Escherichia coli. Journal of Hazardous Materials, 180(1-3), 212–216.

<http://doi.org/10.1016/j.jhazmat.2010.04.016>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as percentage of dead E. Coli population

3.3. Comment on endpoint:

To determine the toxicity :

E. coli, the culture was grown in nutrient broth medium overnight at 30 °C, 200 rpm. Nutrient broth was prepared in 100mM phosphate buffer with the pH under study and then diluted to a 5mM concentration. E. coli was diluted as described in Tables 1 and 2, and 1mL of the inoculum were added to 49mL of nutrient medium in 125mL Erlenmeyer flask containing ZVCN.

The flasks (with ZVCN and E. Coli already prepared) were incubated for 30 min at different temperatures and aeration rate. Serial dilution and plating was carried out at the end of the experiment and agar plates incubated at 30 °C for 24 h to determine the colony forming units (CFUs).

The percent toxicity of nanoparticles was determined by comparing the number of CFU present in the media after the incubation as compared to the number of CFU at time zero.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- pH
- Temperature
- Aertion rate
- Concentration of nanoparticles
- Concentration of bacteria; 5

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

16/5

Descriptor: Chemical ratio :5:16 ~ 1:3

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

ZVCN with size verage of 25 nm

Range (2-60)

Temperature(°C) from 30 to 44

Aeration(rpm) from 0 to 400

Concentration(ppm) of nanoparticles from 0 to 1250

Concentration of bacteria (in % of v) from 0.1 to 10.1

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

16 Metal

List: ZVCN: zero valent copper Cu nanoparticle

Shape: Spherical

Coating: NA

Size (nm): Average of 25

Range of 2-60

Other info: Dynamic light scattering (DLS) was used to measure the rate of ZVCN aggregation at various temperatures and pH. Particle size measurements were carried out using a light scattering technique from Zeta Sizer Nano (Malvern Instruments). The instrument uses a 633nm wavelength laser to measure the size distribution of suspended particles.

Surface areas of 30-50 m²/g

6.6.Pre-processing of data before modelling:

All the data was obtained from experimental results, extra 3 experiments were developed to test the

"Quadratic model"

6.7.Statistics for goodness-of-fit:

$R^2=0.69$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal

List

ZVCN: zero valent cooper Cu nanoparticle

Shape:Spherical

Coating:NA

Size(nm): Average of 25

Range of 2-60

Other properties:

Dynamic light scattering (DLS) was used to measure the rate of ZVCN aggregation at various temperatures and pH. Particle size measurements

were carried out using a light scattering technique from Zeta Sizer Nano (Malvern Instruments). The instrument uses a 633nm wavelength laser to measure the size distribution of suspended particles.

Surface areas of 30-50 m²/g

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

In this paper a linear regression model for few experimental results were developed, and a "Quadratic model" was also applied.

The last one was not clearly explained, and an overfitting could be present.

Due the some empty classification fields and the fact that the descriptors are not physicochemical or structural properties of the NP, it could be not considered as QSAR after all.

ZVCN: zero valent cooper nanoparticle

MLR: multiple linear regression

R²: Correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - pH


- Temperature

- Aeration rate

- Concentration of nanoparticles

- Concentration of bacteria,MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Oxidative stress caused by metal oxides nanoparticles
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Oxidative stress caused by metal oxides nanoparticles

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Enrico Burello

Enrico.Burello@ec.europa.eu

2.6. Date of model development and/or publication:

2011

2.7. Reference(s) to main scientific papers and/or software package:

Burello, E., & Worth, A. P. (2011). A theoretical framework for predicting the oxidative stress potential of oxide nanoparticles. *Nanotoxicology*, 5(2), 228–235.

<http://doi.org/10.3109/17435390.2010.502980>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Intra- or extracellular redox reactions, creating an imbalance of oxidized levels in a cell

3.2. Endpoint:

ROS

3.3. Comment on endpoint:

The model uses reactivity descriptors to build the energy band structure of oxide nanoparticles,

assuming a particle diameter larger than 20–30 nm and no surface states in the band gap, and predicts their ability to induce an oxidative stress by comparing the redox potentials of relevant intracellular reactions with the oxides' energy structure

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

NA

4.3.Descriptors in the model:

- Eg: Band gap energy of bulk material (obtained from ΔH_f , enthalpy of formation, ionization potential, and electron affinity); 0

4.4.Descriptor selection:

Theoretical framework based on chemical hypothesis

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :1:6

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Metal oxides of sizes greater than 20-30nm within the range of parameters (descriptors) of the training set

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No
Chemical Name: not applicable
SMILES: not applicable
Formula: not applicable
INChI: not applicable
MOL file: not applicable
Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes
NP size: Yes
NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: Y2O3

Lu2O3

MgO

Al2O3

SiO2

Li2O

CaO

BaO

TiO

BeO

HfO2

SrO

Ti2O3

Sc2O3

K2O

VO

La2O3

Na2O

Cs2O

ZrO2

Er2O3

NbO

Ho2O3

Tb2O3

Dy2O3

Rb2O

Ce2O3

Gd₂O₃

Nd₂O₃

Yb₂O₃

MnO

GeO₂

Ga₂O₃

GeO

Eu₂O₃

V₂O₃

PbO

NiO

Ti₂O

ZnO

NbO₂

SnO₂

CdO

Cr₂O₃

In₂O₃

CoO

TiO₂

FeO

Fe₂O₃

Mn₂O

PbO

Shape: NA

Coating: NA

Size (nm): Greater than the range of 20-30

Other info: For TiO₂ anatase and rutile crystal structure

Size decided from theoretical assumptions:

"Band energies are calculated from the electronegativities of the constituent atoms and band gap values of oxides, assuming that nanoparticles do not have surface states in the band gap and behave like bulk materials – this approximation is valid if the particles diameter exceeds 20–30 nm (Auffan et al., 2009b)"

6.6.Pre-processing of data before modelling:

Titanium, copper, zinc and iron oxides were selected as case studies because of their relevancy as ingredients in consumer products and their large production volumes as well as because they potentially entail a variety of electron transfer processes that eventually trigger an oxidative stress response in vitro (Nel et al., 2006).

Band gaps of 64 untested metal oxides were predicted (For specific details see Fig 4 in the publication)

6.7.Statistics for goodness-of-fit:

$R^2=0.84$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA Metal Oxide

ListY₂O₃Lu₂O₃

MgO

Al₂O₃SiO₂Li₂O

CaO

BaO

TiO

BeO

HfO₂

SrO

Ti₂O₃Sc₂O₃

K₂O
 VO
 La₂O₃
 Na₂O
 Cs₂O
 ZrO₂
 Er₂O₃
 NbO
 Ho₂O₃
 Tb₂O₃
 Dy₂O₃
 Rb₂O
 Ce₂O₃
 Gd₂O₃
 Nd₂O₃
 Yb₂O₃
 MnO
 GeO₂
 Ga₂O₃
 GeO
 Eu₂O₃
 V₂O₃
 PbO
 NiO
 Tl₂O
 ZnO
 NbO₂
 SnO₂
 CdO
 Cr₂O₃
 In₂O₃
 CoO
 TiO₂
 FeO
 Fe₂O₃
 Mn₂O
 PbO

Shape:NA

Coating:NA

Size(nm): Grater than the range of 20-30

Other properties:

For TiO₂ anatase and rutile crystal structure

Size decided from theoretical assumptions:

"Band energies are calculated from the electronegativities of the constituent

atoms and band gap values of oxides, assuming that nanoparticles do not have surface states in the band gap and behave like bulk materials – this approximation is valid if the particles diameter exceeds 20–30 nm (Auffan et al., 2009b)"

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

This is more a theoretical model based on chemical theory rather than a QSAR that is based on observed data

The model is built on data for 6 NPs and extrapolation is done for 64 NPs

ROS: Reactive oxygen species. Despite of being a normal product of biological metabolism, ROS levels can increase dramatically under stress conditions (it is also known as oxidative stress), and damage the cell structures.

R²: Correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell,

Intra- or extracellular redox reactions, creating an imbalance of oxidized levels in a cell, QSAR, - Eg:
Band gap energy of bulk material (obtained from ΔH_f , enthalpy of formation, ionization potential, and electron affinity), NA

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Classification nanoSAR for the cytotoxicity of metal oxide nanoparticles
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Classification nanoSAR for the cytotoxicity of metal oxide nanoparticles for BEAS-2B by a Logistic Regression

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yoram Cohen

yoram@ucla.edu

2.6. Date of model development and/or publication:

2011

2.7. Reference(s) to main scientific papers and/or software

package:

Liu, R., Rallo, R., George, S., Ji, Z., Nair, S., Nel, A. E., & Cohen, Y. (2011). Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles. *Small*, 7(8), 1118–1126.

<http://doi.org/10.1002/smll.201002366>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Transformed bronchial epithelial cells (BEAS-2B)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as Percentage of damaged cells by Propidium Iodide uptake of BEAS-2B

3.3. Comment on endpoint:

The cytotoxicity induced in BEAS-2B cells exposed to nanoparticle concentrated in the range of

0.375–200 mg*L⁽⁻¹⁾ was assessed by measuring plasma-membrane leakage quantified by high throughput screening (HTS) of the Propidium Iodide (PI) uptake with the results quantified in terms of the percentage of membrane-damaged cells.

The cytotoxicity-screening assay was carried out using a set of six 384 well plates containing both cells in a BEGM medium exposed to nanoparticles in a range of concentrations and unexposed cells. In order to improve the reliability of toxic response identification, replicate samples and controls were used within each plate to estimate experimental variability.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Logistic Regression

4.3.Descriptors in the model:

- Atomization energy (EMeO (kcal eqv – 1))
- The period of metal atom (PMe)
- Nanoparticle primary size (d (nm))
- Nanoparticle mass concentration (θv); 4

4.4.Descriptor selection:

Leave-One-Out cross-validation (LOO) to calculate the accuracy and take into account the number of false positives. In order to select the best set of parameters an internal validation was performed.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

54/4

Descriptor: Chemical ratio :4:54 ~1:13

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Metal oxides of sizes 8-19nm with densities in the range of 2.2–7.22 g*cm⁽⁻³⁾ within the range of parameters (descriptors) of the training set

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

54 Metal Oxide

List: Al₂O₃

CeO₂

Co₃O₄

TiO₂

ZnO

CuO

SiO₂

Fe₃O₄

WO₃

Shape: Spherical

Coating: NA

Size (nm): 8-19

Other info: Specified:

TiO₂ (80% anatase and 20% rutile)

SiO₂ (amorphous)

9 NPs at 9 different concentrations, plus one more for both cases of Fe₃O₄ and WO₃

The nanoparticles were with primary sizes in the range of 8–19 nm and densities in the range of 2.2–7.22 g/cm³. The nanoparticles' surface charge was determined via zeta-potential measurements (ZetaPALS, Brookhaven

Instruments Corporation, Holtsville, NY) in water and as a function of pH, and these measurements also served to determine the isoelectric point (IEP , the pH at which a nanoparticle suspension has zero zeta potential) of the nanoparticles. All measurements were conducted using 1.5 mL of 50 mg L – 1 aqueous nanoparticle dispersion and for each measurement fi ve replicate runs of 10 cycles were collected.

6.6.Pre-processing of data before modelling:

Data for the Fe₃O₄ , WO₃ , and SiO₂ nanoparticles (Table 2 in the publication) were reserved for model validation, with the remaining six used for model training(Al₂O₃ ,CeO₂, Co₃O₄, TiO₂, ZnO, CuO).

6.7.Statistics for goodness-of-fit:

Accuracy: 100%

False Negatives: 0

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

29 Metal Oxide

List

Al₂O₃

CeO₂

Co₃O₄

TiO₂

ZnO

CuO

SiO₂

Fe₃O₄

WO₃

Shape:Spherical

Coating:NA

Size(nm): 8-19

Other properties:

Specified:

TiO₂ (80% anatase and 20% rutile)

SiO₂ (amorphous)

9 NPs at 9 different concentrations, plus one more for both cases of Fe₃O₄ and WO₃

The nanoparticles were with primary sizes in the range of 8–19 nm and densities in the range of 2.2–7.22 g/cm³. The nanoparticles' surface charge was determined via zeta-potential measurements (ZetaPALS, Brookhaven Instruments Corporation, Holtsville, NY) in water and as a function of pH, and these measurements also served to determine the isoelectric point (IEP, the pH at which a nanoparticle suspension has zero zeta potential) of the nanoparticles. All measurements were conducted using 1.5 mL of 50 mg L⁻¹ aqueous nanoparticle dispersion and for each measurement five replicate runs of 10 cycles were collected.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Accuracy: 100%

False Negatives: 0

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Strictly standardized mean difference (SSMD) and Maximum likelihood estimation (MLE) was used to identify and label measurable cytoToxic Effect (Active/Inactive)s, for the experimental results. (SI, Table S2 in the publication supplementary information).

As they recognize:

"A significantly larger data set would be needed in order to expand the applicability domain and increase the confidence and the reliability of the results"

A widespread Mechanistic Interpretation is present, which improves the reliability of the model.

BEAS-2B: Transformed bronchial epithelial cells

SSMD: Strictly standardized mean difference

MLE: Maximum likelihood estimation

LOO: Leave-One-Out Cross-Validation

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Cell, Transformed bronchial epithelial cells (BEAS-2B), QSAR, - Atomization energy (EMeO (kcal eqv – 1))

- The period of metal atom (PMe)
- Nanoparticle primary size (d (nm))
- Nanoparticle mass concentration (θ_v), Logistic Regression

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction of the Biological surface adsorption index (BSAI)
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction of the Biological surface adsorption index (BSAI)
on different NPs by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Jim E. Riviere

jim_riviere@ncsu.edu

2.6. Date of model development and/or publication:

2011

2.7. Reference(s) to main scientific papers and/or software package:

Xia, X. R., Monteiro-Riviere, N. A., Mathur, S., Song, X., Xiao, L., Oldenberg, S. J., ... Riviere, J. E. (2011). Mapping the surface adsorption forces of nanomaterials in biological systems. ACS Nano, 5(11), 9074-9081

Chen, R., Zhang, Y., Monteiro-Rivier

<http://doi.org/10.1021/nn203303c>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

$\log(k)$

k: adsorption coefficient

3.3.Comment on endpoint:

The adsorption coefficient of a set of compounds with relevant biology activity on different nanomaterials was experimental measured.

The data process for obtaining the nanodescriptors was briefly introduced using multiwalled carbon nanotubes (MWCNT) with diameters of 40 nm and carboxyl (COOH) surface derivatives as an example.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- V: Lipophilicity interaction
- β : Hydrogenbond basicity
- α : Hydrogenbond acidity
- π : Pipolarity/polarizability
- R: lone-pair electrons; 5

4.4.Descriptor selection:

The descriptor were selected from knowledge-based reason.

Trying to represent the surface adsorption forces

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

28/5

Descriptor: Chemical ratio :5:28 ~ 1:6

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

$h^* = 0.64$

Any outlier was detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

28 Carbon-based

List: MWCNT40nm-COOH

Shape: Fiber

Coating: chlorobenzene

ethylbenzene

p-xylene

bromobenzene

propylbenzene

4-chlorotoluene

phenol

benzonitrile

4-fluorophenol

benzyl alcohol

iodobenzene

acetophenone

3-methylphenol

methyl benzoate
 4-chloroanisole
 phenethyl alcohol
 3-methylbenzyl alcohol
 4-ethylphenol
 3,5-dimethylphenol
 ethyl benzoate
 methyl 2-methylbenzoate
 naphthalene
 3-chlorophenol
 4-nitrotoluene
 4-chloroacetophenone
 3-bromophenol
 1-methylnaphthalene
 nitrobenzene

Size (nm): Diameter: 40

Other info: The adsorption coefficients of the probe compounds were measured using a solid phase microextraction (SPME) and gas chromatography with mass spectrometry (GC/MS) method.

MWCNT40nm-COOH information provided by the supplier: Carboxylated MWCNT 30-50 nm OD(TEM), 10-20 μm in length, 95% purity, SSA 60 m^2/g and 0.73 wt% COOH

(supplier)Timesnano.com

For specific details see (in the publication) Table S2 for the rest nanomaterials where the model was tested.

6.6.Pre-processing of data before modelling:

It was used 28 compounds as training data and 12 extra compound for an external validation.

For an internal validation the training data was splitted several times to apply LOO and LMO.

6.7.Statistics for goodness-of-fit:

$R^2=0.95$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{\text{LOO}} = 0.923$

$Q^2_{\text{MLO}(25\%)} = 0.908$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

12 MCarbon-based

List

MWCNT40nm-COOH

Shape:Fiber

Coating:chlorobenzene

ethylbenzene

p-xylene

bromobenzene

propylbenzene

4-chlorotoluene

phenol

benzonitrile

4-fluorophenol

benzyl alcohol

iodobenzene

acetophenone

3-methylphenol

methyl benzoate

4-chloroanisole

phenethyl alcohol

3-methylbenzyl alcohol

4-ethylphenol

3,5-dimethylphenol

ethyl benzoate

methyl 2-methylbenzoate

naphthalene

3-chlorophenol
 4-nitrotoluene
 4-chloroacetophenone
 3-bromophenol
 1-methylnaphthalene
 nitrobenzene

Size(nm): Diameter: 40

Other properties:

The adsorption coefficients of the probe compounds were measured using a solid phase microextraction (SPME) and gas chromatography with mass spectrometry (GC/MS) method.

MWCNT40nm-COOH information provided by the supplier: Carboxylated MWCNT 30-50 nm OD(TEM), 10-20 µm in length, 95% purity, SSA 60 m²/g and 0.73 wt% COOH

(supplier)Timesnano.com

For specific details see (in the publication) Table S2 for the rest nanomaterials where the model was tested.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$Q^2_{\text{ext}}=0.78$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

For visual presentation and a clear comparison of diverse nanomaterials, the five-dimensional information can be reduced to two-dimensional via principal component analysis of the five nanodescriptors.

A two-dimensional plot of the two principal components (CP-1 versus

CP-2) is shown in publication's Figure 4. The 16 nanomaterials can be roughly clustered in strong/medium/weak adsorption nanomaterials.

The percentage of the explained variance should be reported for the components of the PCA, in order to evaluate the quality of the plot.

The same model was fitted for a different set of NPs in a posterior work (almost the same authors were involved in) where also Artificial Neural Network was tested for the same descriptors. Reference: Chen, R., Zhang, Y., Monteiro-Riviere, N. A., & Riviere, J. E. (2016). Quantification of nanoparticle pesticide adsorption: computational approaches based on experimental data. *Nanotoxicology*, 10(8), 1118–1128. <http://doi.org/10.1080/17435390.2016.1177745>

R²: Correlation coefficient

Q²_LOO: Leave-One-Out Cross-Validation correlation coefficient

Q²_LMO: Leave-Many-Out Cross-Validation correlation coefficient

Q²ext: external validation coefficient

MLR: Multiple Linear Regression

BSAI: Biological su

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - V: Lipophilicity interaction


- β: Hydrogenbond basicity

- α: Hydrogenbond acidity

- π: Dipolarity/polarizability

- R: lone-pair electrons, MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Toxicological effects of Co-NPs/Co-ions on different cell types studied
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Toxicological effects of Co-NPs/Co-ions on different cell types studied by KDD method: Decision tree (J48).

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Rafi Korenstein

korens@post.tau.ac.il

2.6. Date of model development and/or publication:

2011

2.7. Reference(s) to main scientific papers and/or software

package:

Horev-Azaria, L., Kirkpatrick, C. J., Korenstein, R., Marche, P. N., Maimon, O., Ponti, J., ... Villiers, C. (2011). Predictive toxicology of cobalt nanoparticles and ions: Comparative in vitro study of different cellular models using methods of knowledge d

<http://doi.org/10.1093/toxsci/kfr124>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Lung (A549 and NCIH441 cell lines)

Liver (HepG2 cell line)

Kidney (MDCK cell line)

Intestine (caco-2 TC7 cell line)

Primary mouse Dendritic Cells.

3.2.Endpoint:

In vitro - Cytotoxicity - measured as percentage of cellular viability

3.3.Comment on endpoint:

Different toxicity threshold (% of Cellular viability) were tested: accuracies of 86.6% for viability decrease threshold larger than 20%(>EC20), 87.1%for viability decrease threshold of 25% (>EC25), and 87.3% for viability decrease threshold of 30% (>EC30). No statistical significance between them was considered.

Dose-response curves were examined employing 3-(4,5-dimethylthiazol- 2-Yl)-2,5-diphenyltetrazolium bromide test (MTT), neutral red (NR), and Alamar blue as end point assays following 48- and 72-h exposures.

Because cobalt NPs undergoes dissolution in aqueous media, we determined the dose-response curves for Co-ions, employing cobalt chloride for the same end points.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

DT: Decision Tree

4.3.Descriptors in the model:

- Cell type (lung, liver, kidney, intestine, and the immune system)
- Particle type (Co-NPs or Co-ions)
- Concentration of Particle type
- Exposure time (48h or 72h); 4

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

2896/4

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper

Co-NPs from 20 nm to 500 nm and Co-ions

On those Cell type (lung, liver, kidney, intestine, and the immune system)

With concentration up to 1 mM

Exposure times of 48h and 72h

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

2896 Metal

List: Co NPs

Co ions

Shape: Hexagonal crystal system

Coating: NA

Size (nm): 10 and 50.

Other info: About 20% of the raw data was ignored due to unreasonable data variation.

From Co NPs and Co ions were obtained 2896 instances due to the different concentrations, exposure times and the target cell type

To investigate size, size distribution, morphology, agglomeration, and chemical composition in the range of nanometers and the crystal structure in the range of Armstrong, they used a field emission scanning electronic

microscope (SEM), the Zeiss 1540 EsB, and a conventional JEOL 3010 operating at 297 kV equipped with a LaB6 cathode, post-column Gatan imaging filter, and a 1-K slow-scan charge-coupled device (CCD) camera.

Hexagonal single crystals can be identified from the high-resolution transmission electron microscopy (HRTEM)

The chemical composition of the NPs was confirmed by taking an electron energy loss spectrum, which shows the characteristic O K-edge and Co L_{2,3}-edge.

The morphology of Co-NPs was also characterized in water and in complete cell culture medium by SEM/energy dispersive using x-ray technique and their size distribution by particles tracking analysis (NanoSight LM20 Nanoparticles Analysis System, Salisbury, UK). SEM analysis showed NPs aggregates in both cases. However, a small population of single dispersed NPs was detected by single particle tracking analysis revealing a size distribution ranging between 20 and 500 nm with a peak at 80 nm.

A table containing the main characteristics of Co-NPs is given in the Supplementary table SD1

6.6.Pre-processing of data before modelling:

In order to evaluate the classifier, they trained in 10-fold cross-validation mode. This was carried out by splitting the data set into 10 groups, using 9 of the groups for training and the 10th for validation, repeating this process 10 times. This method gives robust result for model validation (in essence it does the validation 10 times, where each time the test set is randomly chosen).

6.7.Statistics for goodness-of-fit:

EC20:

Accuracy: 86,6 %

- Class Toxic:

TP: 92,5 %

FP: 26 %

Recall (or Sensitivity): 92.5 %

Precision: 88.4 %

Fmesure: 90.4 %

- Class NonToxic:

TP: 74 %

FP: 0.75 %

Recall (or Sensitivity): 74 %

Precision: 82.2

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross-validation applied to generate the final model

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal

List

Co NPs

Co ions

Shape:Hexagonal crystal system

Coating:NA

Size(nm): 10 and 50.

Other properties:

About 20% of the raw data was ignored due to unreasonable data variation.

From Co NPs and Co ions were obtained 2896 instances due to the different concentrations, exposure times and the target cell type

To investigate size, size distribution, morphology, agglomeration, and chemical composition in the range of nanometers and the crystal structure in the range of Armstrong, they used a field emission scanning electronic microscope (SEM), the Zeiss 1540 EsB, and a conventional JEOL 3010 operating at 297 kV equipped with a LaB6 cathode, post-column Gatan imaging filter, and a 1-K slow-scan charge-coupled device (CCD) camera.

Hexagonal single crystals can be identified from the high-resolution transmission electron microscopy (HRTEM)

The chemical composition of the NPs was confirmed by taking an electron energy loss spectrum, which shows the characteristic O K-edge and Co L2,3-

edge.

The morphology of Co-NPs was also characterized in water and in complete cell culture medium by SEM/energy dispersive using x-ray technique and their size distribution by particles tracking analysis (NanoSight LM20 Nanoparticles Analysis System, Salisbury, UK). SEM analysis showed NPs aggregates in both cases. However, a small population of single dispersed NPs was detected by single particle tracking analysis revealing a size distribution ranging between 20 and 500 nm with a peak at 80 nm.

A table containing the main characteristics of Co-NPs is given in the Supplementary table SD1

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

There are not structural or molecular descriptors, thus could be not considered as QSAR model.

The data was treated also by a Naive Bayes model, getting 78.5% of accuracy (worse than the applied decision tree)

"In order to be able to discriminate between the direct toxicological effect of Co-NPs and their indirect effect due the release of Co-ions arising from the dissolution of the NPs, we determined the extent of release of Co-ions from Co-NPs"

Theoretical KDD (Knowledge Discovery from Data) was explained deeply .

Deeply explained the Mechanistic Interpretation of the results.

The same group, with the same methodology, develop the same model with a different set of NPs in a posterior work:

Horev-Azaria, L., Baldi, G., Beno, D., Bonacchi, D., Golla-Schindler, U., Kirkpatrick, J. C., ... Korenstein, R. (2013). Predictive Toxicology of cobalt ferrite nanoparticles: Comparative in-vitro study of different cellular models using methods of knowledge discovery from data. *Particle and Fibre Toxicology*, 10(1). <http://doi.org/10.1186/1743-8977-10-32>

TP: True Positive

FP: False Positive

J48: open source Java implementation of the C4.5 (an algorithm used to generate a decision tree) in the Weka data mining tool

LMO: Leave-Many-Out cross-validation

EC20 : concentration of a drug, antibody or toxica

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Lung (A549 and NCIH441 cell lines)

Liver (HepG2 cell line)

Kidney (MDCK cell line)

Intestine (caco-2 TC7 cell line)

Primary mouse Dendritic Cells.


, QSAR, - Cell type (lung, liver, kidney, intestine, and the immune system)

- Particle type (Co-NPs or Co-ions)

- Concentration of Particle type

- Exposure time (48h or 72h),DT: Decision Tree

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Experimental testing and computational modelling methodologies to
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Experimental testing and computational modelling methodologies to study the cytotoxicity of metal oxide nanoparticles in E. coli

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Jerzy Leszczynski

jerzy@icnanotox.org

2.6. Date of model development and/or publication:

2011

2.7. Reference(s) to main scientific papers and/or software package:

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ... Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, 6(3), 175–178.

<http://doi.org/10.1038/nnano.2011.10>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/EC50)

3.3. Comment on endpoint:

Determined the cytotoxicity of the metal oxide nanoparticles in terms of EC50 (concentration which cytotoxicity reduces bacteria viability up to 50%) based on the curve fitting least squares procedure. For experimental testing protocol see: Publication's supplementary information (Section 1)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- ΔH_{Me+} : represents the enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure.; 1

4.4.Descriptor selection:

The calculations of the descriptors have been performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package

GA (Genetic Algorithm) applied to the model building algorithm to select the most relevant descriptors among the initial ones.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

10/1

Descriptor: Chemical ratio :1:10

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Was verified with leverage approach and Williams plot. (For specific details see the publication's Figure S12 in Supplementary material)

$h^* = 0.6$

Any outlier was detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

10 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of publication's

supplementary material) to be used on the calculations .

6.6.Pre-processing of data before modelling:

The splitting algorithm was as follows:

- (1). 13 metal oxides for which toxicity data had been either taken from the previous paper, or they had been tested in Batch I were sorted based on decreasing toxicity.
- (2). In a next step they were split into two sets: the training set (T) and the validation set (V1) in a way ensured that the points from V1 were evenly distributed within the range of the toxicity of the training set compounds (T). We utilized the following pattern of splitting: T-T-V1-T-T-T-V1-T-T-T-V1-T-T.
- (3). Finally, three additional compounds tested in Batch II and La₂O₃ were additionally included in the validation set (those compounds are indicated with V2).

We split the data in an above discussed way because of three reasons:

- (i) to ensure that the compounds V1 are evenly distributed within the range of toxicity log (1/EC₅₀),
- (ii) to have both experimental batches represented in the validation set, whereas only compounds from the Batch I were used for training,
- (iii) to include to the validation set some additional compounds (V2) having toxicity not necessarily within the range of the training set (this would be impossible, if we have merged compounds from Batch I and II together and then labeled every third compound as a member of the validation set). Indeed, observed toxicity of CoO was higher than toxicity of the most toxic compound in the training set (ZnO).

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.85$$

$$RMESC = 0.20$$

t values and p-values for model's coefficients were computed (see Supplementary material, section 2.4)

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$$Q^2_{cv} = 0.77$$

$$RMSECV = 0.24$$

Y-scrambling test was applied. See supplementary material (section 2.5)

Confirmed the significance of the QSAR

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable
 INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA
 NP size: Yes
 NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

7 MMetal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape:NA

Coating:NA

Size(nm): 15-90

Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of publication's supplementary material) to be used on the calculations .

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$$Q^2_{\text{ext}} = 0.83$$

$$\text{RMSEP (V1+V2)} = 0.19$$

(Not relevant statistical difference)

$$\text{RMSEP (V1)} = 0.07$$

$$\text{RMSEP (V2)} = 0.19$$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

All quantum-mechanical calculations were performed using the PM6 method as implemented in MOPAC 2009 (semi-empirical method).

Because the size of the metal oxides nanoparticles we tested (15 - 90 nm) was too large even to perform calculations at the semi-empirical level it was necessary to simplify the molecular models used for calculations of the descriptors.

Mechanistic Interpretation was proposed, which improves the reliability of the model.

Consistent paper in all the QSAR fields, as well described in OECD document for Validation QSAR.

Supplementary material is a great example of QSAR methodology.

GA-MLR: Genetic Algorithm and Multiple Linear Regression

LOO: Leave-One-Out Cross-Validation

R^2 : correlation coefficient

Q^2_{cv} : leave-one-out cross-validation correlation coefficient

Q^2_{ext} : correlation coefficient for external validation

RMSEC:

9.2.Bibliography:

Experimental data plus previous published work:

Hu, X., Cook, S., Wang, P. & Hwang, H. M. In vitro evaluation of cytotoxicity of engineered metal oxide nanoparticles. Sci. Total Environ. 407,

3070–3072 (2009).

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - ΔH_{Me+} : represents the enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure.,MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software

package:

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., Puzyn, T., Leszczynska, D., & Leszczynski, J. (2012). Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli. Chemosphere, 89(9), 10

<http://doi.org/10.1016/j.chemosphere.2012.05.077>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/EC50)

3.3.Comment on endpoint:

Determined the cytotoxicity of the metal oxide nanoparticles in terms of EC50 (concentration which cytotoxicity reduces bacteria viability up to 50%) based on the curve fitting least squares procedure.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

Three SMILES attributes :

- '[' : Each non-hydrogen atom is specified independently by its atomic symbol enclosed in square brackets, []

- '=' : double bond

- 'O' : oxygen

; 3

4.4.Descriptor selection:

SMILES-based optimal descriptors and Monte-Carlo optimization with a threshold value from 0 to 2 (2 was the best) by the CORAL software.

(see section "2.2 Optimal descriptors")

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

11/3

Descriptor: Chemical ratio :3:11 ~ 1:4

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

It could be defined as metal oxides from 15-90 nm, and with metal oxide nanoparticles which have a close related SMILES structures with those in the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

11 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of publication's supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package

6.6.Pre-processing of data before modelling:

Six random splits into the training and test sets were examined (6 splits).

Two principles were used in order to prepare these splits:

- (i) ranges of pEC50 for training and test set are as equivalent as possible
- (ii) percentage of splits identity is as small as possible.

(For specific details see (in the publication) Table 1)

6.7.Statistics for goodness-of-fit:

Split1:

$$R^2 = 0.7407$$

$$RMSE = 0.234$$

Split2:

$$R^2 = 0.8217$$

$$RMSE = 0.232$$

Split3:

$$R^2 = 0.8214$$

$$RMSE = 0.170$$

Split4:

$$R^2 = 0.7779$$

$$RMSE = 0.261$$

Split5:

$$R^2 = 0.8171$$

$$RMSE = 0.207$$

Split6:

$$R^2 = 0.8377$$

$$RMSE = 0.190$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Y-randomization:

Split1:

(c)R²_p = 0.807

Split2:

(c)R²_p = 0.816

Split3:

(c)R²_p = 0.753

Split4:

(c)R²_p = 0.857

Split5:

(c)R²_p = 0.779

Split6:

(c)R²_p = 0.742

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

6 M Metal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂
 TiO₂
 V₂O₃
 Y₂O₃
 Bi₂O₃
 In₂O₃
 Sb₂O₃
 SiO₂
 ZrO₂
 CoO
 NiO
 Cr₂O₃
 La₂O₃

Shape:NA

Coating:NA

Size(nm): 15-90

Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of publication's supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Split1:

$R^2 = 0.9402$

RMSE= 0.204

Split2:

$R^2 = 0.9648$

RMSE= 0.236

Split3:

$R^2 = 0.8363$

RMSE= 0.337

Split4:

$R^2 = 0.9468$

RMSE= 0.139

Split5:

$R^2 = 0.9260$

RMSE= 0.270

Split6:

$R^2 = 0.8487$

RMSE= 0.294

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The aim of the present study is evaluation of the CORAL as a tool for the QSAR analysis of the toxicity of metal oxide nanoparticles.

The obtained model is a linear regression, the interesting fact is how are selected and managed the descriptors to generate one variable form them for the linear regression.

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

CORAL: CORrelation And Logic

SMILES: Simplified Molecular Input Line Entry Specification

pEC50: the minus logarithm of the concentration of the metal oxide nanoparticles effects the reduction of bacteria viability of 50%

R²: correlation coefficient

R

9.2.Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ...

Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nature Nanotechnology, 6(3), 175–178. 10.1038/nnano.2011.10

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, Three SMILES attributes :

- '[' : Each non-hydrogen atom is specified independently by its atomic symbol enclosed in square brackets, []


- '=' : double bond

- 'O' : oxygen

,Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Consensus model to predict the Nanoparticles uptake of PaCa2 cells
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Consensus model to predict the Nanoparticles uptake of PaCa2 cells
(it includes kNN, SVM, NB in the final consensus model)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Chun Wei Yap

phayapc@nus.edu.sg

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software package:

Chau, Y. T., & Yap, C. W. (2012). Quantitative Nanostructure-Activity Relationship modelling of nanoparticles. RSC Advances, 2(22), 8489–8496.

<http://doi.org/10.1039/c2ra21489j>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3. Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

The cellular uptakes in PaCa2 for the 105 nanoparticles were ranged from 170 to 27 542 nanoparticles per cell. A total of 56 nanoparticles with cellular uptake of more than 5000 nanoparticles per cell were considered to have good/moderate (henceforth referred to as good for brevity) cellular uptake (positive class), while 49 nanoparticles with cellular uptake of less than 5000 nanoparticles per cell were considered to have poor cellular uptake (negative class)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

From 2100 candidate models that were developed 5 candidate models were used in the final model:

Consensus model with:

3 kNN

1 SVM

1 NB

(for more details, see section "2.3.2 Generating candidate models" in the publication)

4.3.Descriptors in the model:

(From 679 descriptors, generated by software PaDELDescriptor v2.8, finally was used as initial descriptors 376. Removed which not showed variance for all the nanoparticles)

- Number of CH2 groups
- Number of primary nitrogens
- Number of secondary nitrogens
- Number of tertiary nitrogens
- Number of Halogens (fluorine, bromine, iodine)
- Number of Sulphurs
- Number of fused rings
- Number of Hydrogen bonding; 8

4.4.Descriptor selection:

A total of 679 1D, 2D chemical descriptors were calculated using PaDEL- Descriptor v2.8 software. After removing those descriptors that showed no variance for all the nanoparticles, 367 chemical descriptors were retained

Performed a randomisation process on the entire set of 367 molecular descriptors such that 100 different pools with varying numbers of descriptors was developed.

Forward selection was applied to each pool of descriptors. A single descriptor that best correlated with the dependent property was first identified and the next most contributing descriptor was then added in the subsequent steps. The selection was stopped when the addition of a descriptor did not improve the model's performance

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

105/8

Descriptor: Chemical ratio :8:105 ~ 1:13

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Each candidate model was built with an applicability domain (AD)

defined using the multiple threshold method proposed by Fumera et al :

They adopted the multiple thresholds method whereby two thresholds were determined for each candidate model. If the confidence value is greater than the higher threshold value, the nanoparticle will be predicted as having good cellular uptake. Conversely, if the confidence value is smaller than the lower threshold value, the nanoparticle will be predicted as having poor cellular uptake. When the confidence value lies between the two threshold values, the nanoparticle will be considered as out of the AD of the model and its degree of cellular uptake will not be predicted.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

105 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione

4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione

Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride

3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione

Dibenz(c,e)oxepin-5,7-dione

6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone

Lauric anhydride

1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid

5-Methyl-2-benzofuran-1,3-dione

4-Nitro-2-benzofuran-1,3-dione

1H-isochromene-1,3(4H)-dione

Dihydro-2H-pyran-2,6(3H)-dione

4,4'-Ethane-1,2-diylmorpholine-2,6-dione

2H-3,1-benzoxazine-2,4(1H)-dione

1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione

4-Methyldihydro-2Hpyran-2,6(3H)-dione

4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione

2,5-Dioxotetrahydrofuran-3,4-diyl diacetate

4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione

Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate

2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

In order to ensure the quality and accuracy of the data, the 3D structure of each compound was generated by converting the SMILES strings of compounds given in Fourches et al., into 3D structures and then manually inspected and compared with the structures provided by Weissleder et al., For further information see section : 2.1.2 Cleaning up dataset

6.6.Pre-processing of data before modelling:

To determine the predictive performance using five-fold cross validation, the dataset was first divided into 5 different portions. Four portions were used to form a training set to develop a model, while the remaining portion formed the testing set to determine the predictive performance of the model. This step was repeated until each subset was being used as the testing set. The five-fold cross validation predictive performance was then calculated as the average predictive performance on these five testing sets.

6.7.Statistics for goodness-of-fit:

Final consensus

SE = 98.2 %

SP = 76.6 %

MCC = 0.777

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

5-fold-cross validation was performed for each model of the consensus model.

(Average)

SE = 86.7 %

SP = 67.3 %

MCC = 0.559

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride
 Pentafluoropropanoic anhydride
 4,3,3-Dimethyldihydrofuran-2,5-dione
 Furan-2,5-dione
 3-Methylfuran-2,5-dione
 7,3,4-Dimethylfuran-2,5-dione
 Hexanoic anhydride
 3-Methyldihydrofuran-2,5-dione
 5,5'-Carbonylbis(2-benzofuran-1,3-dione)
 5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione

Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid

2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

In order to ensure the quality and accuracy of the data, the 3D structure of each compound was generated by converting the SMILES strings of compounds given in Fourches et al., into 3D structures and then manually inspected and compared with the structures provided by Weissleder et al., For further information see section : 2.1.2 Cleaning up dataset

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The dataset from Weissleder et al., 2005, was checked and also compared with the data from Fourches, D. et al., 2010, converting the SMILES strings of the compounds into 3D structures with

The Logistic regression was also included in the possible models, but it was not any of them selected for the final consensus model.

The final consensus model also performed well in predicting the properties of the 14 nanoparticles which have significant cellular uptake (as defined by Weissleder et al. 19)

Briefly mention of the Mechanistic Interpretation of the results.

kNN: k Nearest Neighbour

NB: Naïve Bayes

SVM: Support Vector Machine

SE : Sensitivity

SP: Specificity

MCC: Matthews correlation coefficient

SMILES: Simplified Molecular Input Line Entry Specification

9.2. Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

(already reported in this table)

Fourches, D. et al., 2010. Quantitative nanostructure-activity relationship modelling. *ACS nano*, 4(10), pp.5703–12 (Case Study 2)

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, (From 679 descriptors, generated by software PaDELDescriptor v2.8, finally was used as initial descriptors 376. Removed which not showed variance for all the nanoparticles)

- Number of CH2 groups
- Number of primary nitrogens
- Number of secondary nitrogens
- Number of tertiary nitrogens
- Number of Halogens (fluorine, bromine, iodine)
- Number of Sulphurs
- Number of fused rings
- Number of Hydrogen bonding, From 2100 candidate models that were developed 5 candidate models were used in the final model:

Consensus model with:


3 kNN

1 SVM

1 NB

(for more details, see section "2.3.2 Generating candidate models" in the publication)

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal oxide toxicity prediction through their energy bands and their
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal oxide toxicity prediction through their energy bands and their solubility, by decision trees

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Andre E. Nel

anel@mednet.ucla.edu

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software package:

Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kim, C., Liu, R., ... Nel, A. E. (2012). Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano*, 6(5), 4349–4368.

<http://doi.org/10.1021/nn3010087>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human bronchial epithelial cells (BEAS-2B)

and

Rat alveolar macrophage cells (RAW264.7)

3.2. Endpoint:

In vitro - Cytotoxicity - measured by the area under the curve of dose-response on HTS assay

3.3.Comment on endpoint:

The multi-parameter fluorescence assay by HTS, which quantitatively assesses changes in ROS production (DCF and Mitoses red fluorescence), intracellular calcium flux (Fluo-4 fluorescence), mitochondrial membrane potential (JC-1 fluorescence), and surface membrane permeability (PI uptake) in BEAS-2B and RAW 264.7 cells.

Established dose-response curves using nonparametric smoothing splines and summarized each trajectory with the area under the estimated dose-response curve. They related cytotoxicity (as measured by area under the curve) to conduction energy (Ec) and metal dissolution (dissolution in BEGM) in a regression tree model.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

DT: Decision Tree

4.3.Descriptors in the model:

- Ec : Computed Valence band energy
- Metal Dissolution in BEGM (%); 2

4.4.Descriptor selection:

Partially based on previous studies:

Burello, E., & Worth, A. P. (2011)

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

24/2

Descriptor: Chemical ratio :1:12

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper

It should be considered as applicability domain the range of descriptor values of Metal oxide NPs in size range of 10 - 200 nm

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

24 Metal Oxide

List: Al₂O₃

CuO

CeO₂

Co₃O₄

CoO

Cr₂O₃

Fe₂O₃

Fe₃O₄

Gd₂O₃

HfO₂

In₂O₃

La₂O₃

Mn₂O₃

NiO

Ni₂O₃

Sb₂O₃

SiO₂

SnO₂

R-TiO₂

WO₃

Y₂O₃

Yb₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): 10-100

Other info:

exceptions:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details see (in the publication) Table S1 for Crystalline structure information for metal oxide nanoparticles.

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

6.6.Pre-processing of data before modelling:

There is not an splitting data for testing, only the leave-one-out cross-validation (LOO) error was computed to minimize the complexity of the model.

6.7.Statistics for goodness-of-fit:

NA

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

LOO was applied, values not provided

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal Oxide

ListAl₂O₃

CuO

CeO₂Co₃O₄

CoO

Cr₂O₃Fe₂O₃Fe₃O₄Gd₂O₃HfO₂In₂O₃La₂O₃Mn₂O₃

NiO

Ni₂O₃Sb₂O₃SiO₂SnO₂R-TiO₂WO₃Y₂O₃Yb₂O₃

ZnO

ZrO₂

Shape:NA

Coating:NA

Size(nm): 10-100

Other properties:

exceptions:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details see (in the publication) Table S1 for Crystalline structure information for metal oxide nanoparticles.

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The main goal of the work was to demonstrate the theoretical approach about the toxicity of metal oxide nanoparticles through their bands position, hence their availability of reduction/oxidation reaction on cells, by experimental results (which was reasoned in extense way). At the end they also applied their data to generate a model, (which display the toxicological impact of the descriptors), but it lets the model be in a second plane with not a deeply performance work neither checking about the reliability.

It was used principal components analysis (PCA) to facilitate the interpretation of how individual assays contribute to the definition of the total toxicity profile and to understand how individual assays are related to each other.

NP: nanoparticle

LOO: leave-one-out cross-validation

BEGM: Bronchial Epithelial Growth Medium

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:


Cell, Human bronchial epithelial cells (BEAS-2B)

and

Rat alveolar macrophage cells (RAW264.7), QSAR, - Ec : Computed Valence band energy

- Metal Dissolution in BEGM (%), DT: Decision Tree

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Toxicity prediction of decorated nanotubes by GFA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Toxicity prediction of decorated nanotubes by GFA
(Study in a variety set of descriptors and endpoints.)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yufeng J. Tseng

yjtseng@csie.ntu.edu.tw

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Shao, C.-Y., Chen, S.-Z., Su, B.-H., Tseng, Y. J., Esposito, E. X., & Hopfinger, A. J. (2013). Dependence of QSAR models on the selection of trial descriptor sets: A demonstration using nanotoxicity endpoints of decorated nanotubes. Journal of Chemical In

<http://doi.org/10.1021/ci3005308>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

THP-1 (human monocytic cell line) differentiate into Macrophages

3.2. Endpoint:

In vitro - Cytotoxicity - measured as cellular viability by determining the mitochondrial

dehydrogenases' activity

3.3.Comment on endpoint:

Also there was a study of immune response of macrophages and binding proteins to nanotubes as other endpoints, further, the sum over all the endpoints was been evaluated.

To evaluate the acute cytotoxicity (Cell) of the DNC library in macrophages, Zhou et al., used the WST-1 assay. Cellular viability was measured by determining the mitochondrial dehydrogenases' activity, in the presence and absence of the nanotube-decorator complexes and expressed on a 0 to 100 scale (100 most "toxic"). The immune response was measured by treating macrophages with DNC for 24 h in a solution of lipopolysaccharide

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

The final descriptors of the selected model:

- * - eAll-NP227
- * - eAll-NP404
- * - eAll-NP127
- * - eAll-All83
- ** - vsurf_CW7

* e(X·YM) is the Mth eigenvalue representation of the pair interaction distances (Angstroms) between all atom/group types X and Y averaged over the conformational ensemble of the molecule. All-all atoms of the molecule with no type differentiation, NP – nonpolar, HBA – hydrogen bond acceptor, HBD – hydrogen bond donor, ARO – aromatic, and HS – any type of non-hydrogen atom (hydrogen suppressed; also-known-as heavy atoms).

** vsurf_XM are scalar values for particular types of molecular interaction fields generated by a probe molecule interacting with the molecule of interest. X denotes a particular iso-surface at value M in 3D space around the molecule of interest. (more information at G. Cruciani et al., , J. Molec. Struct, 2000)

CW -> Capacity factors. Capacity factors represent the ratio of the hydrophilic surface over the total molecular

surface; 5

4.4.Descriptor selection:

MOE, Volsurf and 4D-FP were used to generate the different sets of descriptors applied for the different models (separately and also in combination).

GFA: Genetic function approximation

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

29/5

Descriptor: Chemical ratio :5:29 ~ 1:6

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper

Decorated-nanotubes with 1 nm of diameter, which decorators will have close molecular structure and fall within the range of parameters (descriptors) of the training set

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

29 Carbon-based

List: Decorator-nanotube complexesShape: Fiber

Coating: CC(NC(=O)OC3C1C=CC=CC1C2C=CC=CC23)C(=O)Nc4ccccc4
CCOC(=O)c4ccc(NC(=O)C(C)NC(=O)OC3C1C=CC=CC1C2C=CC=CC23)cc
4
CC(N)C(=O)Nc1ccccc1

CC(N)C(=O)NC1CCCCC1
CCCCNC(=O)C(C)NC(=O)c1cccc(N(=O)=O)c1
CCCCNC(=O)C(C)NC(=O)c1ccc(Cl)cc1
CCCCNC(=O)C(C)NS(=O)(=O)c1cccc(N(=O)=O)c1
CCCCN(CCCC)C(=O)C(C)NC(=O)c1ccc(Cl)cc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1cccc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1ccc(C)cc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1cccc(N(=O)=O)c1
CC(NC(=O)c1cccc1)C(=O)NC2CCCCC2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)NC2CCCCC2
CCCCN(CCCC)C(=O)C(C)NC(=O)c1ccc(Cl)cc1
CC(NS(=O)(=O)c1cccc1)C(=O)NC2CCCCC2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)NC2CCCCC2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)Nc2cccc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2cccc2
CC(NS(=O)(=O)c1cccc1)C(=O)Nc2cccc2
Cc2ccc(S(=O)(=O)NC(C)C(=O)Nc1cccc1)cc2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)Nc2cccc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)NCc2cccc2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)NCc2cccc2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)N2CCCC2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)N2CCCC2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)N2CCCC2
CCOC(=O)c2ccc(NC(=O)C(C)NC(=O)c1ccc(Cl)cc1)cc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2cccc(C(F)(F)F)c2
CC(NS(=O)(=O)c1cccc1)C(=O)Nc2cccc(C(F)(F)F)c2

Size (nm): Diameter: 40 ± 10

Length: 250 ± 120

Other info: Coating compounds were reported by SMILES notation.

The number of walls of the nanotube will have little, if any, impact on the conformational behaviour of the surface attached decorator groups. The diameter(s) of the carbon nanotubes were not reported by Zhou, then two different diameters of 1 nm or 1.3 nm of diameter each 6.5 nm in length were defined.

The nanotube–decorator complex was geometry optimized using the molecular dynamics simulation (MDS) package GROMACS (version 4.5.2 for Linux) with the ffmx force field (a derivative of the GROMOS87 force field).

6.6.Pre-processing of data before modelling:

There is not an splitting data for testing, only the leave-one-out cross-validation (LOO) error was computed to minimize the complexity of the model.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.857$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

LOO was applied

 $Q^2_{cv} = 0.759$ **7. External validation - OECD Principle 4****7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Decorator-nanotube complexes

Shape: FiberCoating: CC(NC(=O)OC3C1C=CC=CC1C2C=CC=CC23)C(=O)Nc4ccccc4CCOC(=O)c4ccc(NC(=O)C(C)NC(=O)OC3C1C=CC=CC1C2C=CC=CC23)cc4CC(N)C(=O)Nc1ccccc1CC(N)C(=O)NCc1ccccc1CCCCNC(=O)C(C)NC(=O)c1cccc(N(=O)=O)c1CCCCNC(=O)C(C)NC(=O)c1ccc(Cl)cc1CCCCNC(=O)C(C)NS(=O)(=O)c1cccc(N(=O)=O)c1CCCCN(CCCC)C(=O)C(C)NC(=O)c1ccc(Cl)cc1CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1ccccc1CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1ccc(C)cc1CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1cccc(N(=O)=O)c1

```

CC(NC(=O)c1cccc1)C(=O)NC2CCCCC2
CC(NC(=O)c1ccc(N(=O)=O)c1)C(=O)NC2CCCCC2
CCCCN(CCCC)C(=O)C(C)NC(=O)c1ccc(Cl)cc1
CC(NS(=O)(=O)c1cccc1)C(=O)NC2CCCCC2
CC(NS(=O)(=O)c1ccc(N(=O)=O)c1)C(=O)NC2CCCCC2
CC(NC(=O)c1ccc(N(=O)=O)c1)C(=O)Nc2cccc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2cccc2
CC(NS(=O)(=O)c1cccc1)C(=O)Nc2cccc2
Cc2ccc(S(=O)(=O)NC(C)C(=O)Nc1cccc1)cc2
CC(NS(=O)(=O)c1ccc(N(=O)=O)c1)C(=O)Nc2cccc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2cccc2
CC(NS(=O)(=O)c1ccc(N(=O)=O)c1)C(=O)Nc2cccc2
CC(NC(=O)c1ccc(N(=O)=O)c1)C(=O)N2CCCC2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)N2CCCC2
CC(NS(=O)(=O)c1ccc(N(=O)=O)c1)C(=O)N2CCCC2
CCOC(=O)c2ccc(NC(=O)C(C)NC(=O)c1ccc(Cl)cc1)cc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2ccc(C(F)(F)F)c2
CC(NS(=O)(=O)c1cccc1)C(=O)Nc2ccc(C(F)(F)F)c2

```

Size(nm): Diameter: 40 ± 10

Length: 250 ± 120

Other properties:

Coating compounds were reported by SMILES notation.

The number of walls of the nanotube will have little, if any, impact on the conformational behaviour of the surface attached decorator groups. The diameter(s) of the carbon nanotubes were not reported by Zhou, then two different diameters of 1 nm or 1.3 nm of diameter each 6.5 nm in length were defined.

The nanotube–decorator complex was geometry optimized using the molecular dynamics simulation (MDS) package GROMACS (version 4.5.2 for Linux) with the ffgmx force field (a derivative of the GROMOS87 force field).

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The main objective of this paper was to explore the form, quality, and “complementarity” of QSAR models developed using a variety of trial descriptor sets applied to different endpoints, rather than present a new model. However, with their study they generated a great quantity of models, and we chose one of them to classify it, since all of them were developed under the same methodology.

From 80 decorator-nanotubes complexes (DNC), the 29 most toxic were selected based on cumulative index over all six endpoints measures in Zhou et al., 2008

Without an External validation, we cannot say if we are save of an overfitting case.

Future mechanistic interpretation of the descriptors is suggested by the authors, because , as they admit, the topic was briefly discussed.

LOO: Leave-One-Out Cross-Validation

GFA: Genetic function approximation

MLR: Multiple Linear Regression

Q^2_{cv} : leave-one-out croos-validation correlation coefficient

R^2 : correlation coefficient

9.2. Bibliography:

Zhou, H., Mu, Q., Gao, N., Liu, A., Xing, Y., Gao, S., ... Yan, B. (2008). A nano-combinatorial library strategy for the discovery of nanotubes with reduced protein-binding, cytotoxicity, and immune response. Nano Letters, 8(3), 859–865.
<http://doi.org/10.1021/nl0730155>

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, THP-1 (human monocytic cell line) differenciate into Macrophages, QSAR, The final descriptors of the selected model:

- * - eAll-NP227
- * - eAll-NP404
- * - eAll-NP127
- * - eAll-All83

** - vsruf_CW7


* $e(X \cdot Y_M)$ is the Mth eigenvalue representation of the pair interaction distances (Angstroms) between all atom/group types X and Y averaged over the conformational ensemble of the molecule. All-all atoms of the molecule with no type differentiation, NP – nonpolar, HBA – hydrogen bond acceptor, HBD – hydrogen bond donor, ARO – aromatic, and HS – any type of non-hydrogen atom (hydrogen suppressed; also-known-as heavy atoms).

** vsurf_XM are scalar values for particular types of molecular interaction fields generated by a probe molecule interacting with the molecule of interest. X denotes a particular iso-surface at value M in 3D space around the molecule of interest. (more information at G. Cruciani et al., J. Molec. Struct, 2000)

CW -> Capacity factors. Capacity factors represent the ratio of the hydrophilic surface over the total molecular

surface, MLR: Multiple Linear Regression

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal oxide NPs toxicity classification by SVM
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal oxide NPs toxicity classification by SVM

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yoram Cohen

yoram@ucla.edu

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Liu, R., Zhang, H. Y., Ji, Z. X., Rallo, R., Xia, T., Chang, C. H., ...
Cohen, Y. (2013). Development of structure-activity relationship
for metal oxide nanoparticles. *Nanoscale*, 5(12), 5644–5653.

<http://doi.org/10.1039/c3nr01533e>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human bronchial epithelial cells (BEAS-2B)

and

Rat alveolar macrophage cells (RAW264.7)

3.2. Endpoint:

In vitro - Cytotoxicity - measured by the the curve of dose-response and consensus Self-Organizing Map clustering on SPS and HTS assay

3.3.Comment on endpoint:

Toxicological responses of twenty-four metal oxide NPs (over a concentration range of 0.39–100 mg L⁻¹) on RAW 264.7 and BEAS-2B cell lines, using both single parameter screening (SPS) assays (MTS, ATP and LDH) and multi-parameter high-throughput screening (HTS) assays (Mito, Fluo4, JC1, and PI over exposure time of 1–24 h)

Toxicity class definition derived based on both dose–response analysis and consensus Self-Organizing Map clustering.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM: Support Vector Machine

4.3.Descriptors in the model:

- E_c : Computed Valence band energy

- $(Z^2)/r$: Ionic index, where Z and r are the charge number and ionic radius of metal cation in the NP crystals, respectively.; 2

4.4.Descriptor selection:

Theoretical framework

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

24/2

Descriptor: Chemical ratio :2:24 ~ 1:12

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

The applicability domain was analyzed by a probabilistic approach:

The probability density (i.e., $p(x)$ where x is the descriptor set identifying the NP) represents the NP population to which the original NP dataset belongs. Based on the estimated probability density ($p(x)$), the nano-SAR applicability domain can be identified as a high density region in the descriptor space that covers an acceptable high percentage (typically 60–90%) of the population represented by the NP dataset.

Only applicable to domains up to three dimensions.

They also suggest to apply Williams plot.

Data set contains information about metal oxides nanoparticles from 10 nm up to 70 nm (and also two NPs with 140nm and 190nm

The specific data for descriptor can be checked at Table S1 in supplementary material of the publication. In summary, the range of values in the final descriptor are:

Ec: (-5.174 eV) - (-1.515eV)

Z^2/r : 0.0667 - 0.6154 pm⁻²)

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

24 Metal Oxide

List: Al₂O₃

CuO

CeO₂

Co₃O₄

CoO

Cr₂O₃

Fe₂O₃

Fe₃O₄
 Gd₂O₃
 HfO₂
 In₂O₃
 La₂O₃
 Mn₂O₃
 NiO
 Ni₂O₃
 Sb₂O₃
 SiO₂
 SnO₂
 R-TiO₂
 WO₃
 Y₂O₃
 Yb₂O₃
 ZnO
 ZrO₂

Shape: NA

Coating: NA

Size (nm): 10-100

Other info: exceptions:

Cr₂O₃ : 193±90.0 and Ni₂O₃ : 140.6±52.5

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

Acc_{0.632} = 93.74 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

10-round Y-randomization

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA Metal Oxide

List

Al₂O₃

CuO

CeO₂

Co₃O₄

CoO

Cr₂O₃

Fe₂O₃

Fe₃O₄

Gd₂O₃

HfO₂

In₂O₃

La₂O₃

Mn₂O₃

NiO
 Ni₂O₃
 Sb₂O₃
 SiO₂
 SnO₂
 R-TiO₂
 WO₃
 Y₂O₃
 Yb₂O₃
 ZnO
 ZrO₂
Shape:NA
Coating:NA
Size(nm): 10-100

Other properties:

exceptions:

Cr₂O₃ : 193±90.0 and Ni₂O₃ : 140.6±52.5

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Data was normalized depending on source of data (i.e for HTS data was applied SSMD, standard mean difference).

In order to arrive at a statistically reliable NP toxicity class definition, the normalized data were first analyzed via SOM based consensus clustering and by log-logistic dose–response curve.

SVM was selected after be compared with 5 models: NBC, LIR, LDA, LGR and qLGR

Decision boundaries for the developed nano-SAR, at different acceptance levels of false negatives (FN, misclassification of a toxic NP as non-toxic) and false positives (FP, incorrect prediction of a non-toxic NP as toxic), were constructed using class probabilities. See section 4.6 Nano-SAR decision boundary.

0.632 estimator is a weighted average of re-substitution classification accuracy (Acc_resub) and bootstrapping classification accuracy (Acc_boot):

$$\text{Acc}_{0.632} = 0.632 \cdot \text{Acc}_{\text{boot}} + 0.368 \cdot \text{Acc}_{\text{resub}}$$

There is a Mechanistic Interpretation based on external previous studies.

Acc_0.632: accuracy value from 0.632 estimator.

BEAS-2B: Transformed bronchial epithelial cells

RAW264.7: Rat alveolar macrophage cells

NBC: naïve Bayesian classifier

LIR: linear regression

LDA: linear discriminate analysis

LGR: logistic regressi

9.2. Bibliography:

(already reported in this table)

Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., Liu, R., ... Nel, A. E. (2012). Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. ACS Nano, 6(5), 4349–4368. 10.1021/nn3010087

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Human bronchial epithelial cells (BEAS-2B)

and

Rat alveolar macrophage cells (RAW264.7), QSAR, - E_c : Computed Valence band energy

- $(Z^2)/r$: Ionic index, where Z and r are the charge number and ionic radius of metal cation in the NP crystals, respectively.,SVM: Support Vector Machine

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Bioactivity response (active/inactive) classification of coated iron oxide
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Bioactivity response (active/inactive) classification of coated iron oxide NPs by NBC

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yoram Cohen

yoram@ucla.edu

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Liu, R., Rallo, R., Weissleder, R., Tassa, C., Shaw, S., & Cohen, Y. (2013). Nano-SAR development for bioactivity of nanoparticles with considerations of decision boundaries. *Small*, 9(9-10), 1842–1852.

<http://doi.org/10.1002/smll.201201903>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)

Murine RAW 264.7 leukemic monocyte/macrophage cell

3.2. Endpoint:

In vitro - Cytotoxicity - measured as biological response by H4

3.3.Comment on endpoint:

The dataset provided measurements of biological response for four cell types (Endothelial cells (human aorta), Vascular smooth muscle cells (human coronary artery), Hepatocytes (human HepG2 cells), and Murine RAW 264.7 leukemic monocyte/macrophage cell), exposed to the NPs at four concentrations (0.01, 0.03, 0.1, and 0.3 mg/mL Fe), determined based on four different assays (Apo: apoptosis, Mito: mitochondrial potential, Red: reducing equivalents, and ATP: ATP content). With a certainty above 95%, NP induced response that is above that of the control. In the present work, SNR(Signal-to-Noise ratio) > 1.645 was identified as a hit for a given NP. A 5% chance of miss-identifying a non-hit as "hit" for a given NP, would be equivalent to a miss-identification of 3.2 out of the 64 measurements in its HTS profile. Therefore, even if 5% uncertainty would be acceptable it would be more practical to set the threshold to or above the next higher integer value, i.e., N hit ≥ 4.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

NBC: Naïve Bayesian Classifier

4.3.Descriptors in the model:

- R1: spin-lattice Relaxivity
- zeta potential (surface charge); 2

4.4.Descriptor selection:

The 4 initial descriptors: primary potential, zeta potential, and spin lattice (R1) and spin-spin (R2) relaxivities were tested in range of different building models through different combinations of them.

Finally the most suitable descriptors(by the accuracy of the model) were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

44/2

Descriptor: Chemical ratio :2:44 ~ 1:22

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Probabilistic approach:

Using the NBC model one can quantify the probabilities of a NP $x = [x_1, x_2]$; x_1 - R1 relaxivity) and x_2 - zeta potential) belonging to the bioactive class (T) or inactive class (N) as

$P(T|x) = p(x|T)P(T)/p(x)$ and

$P(N|x) = p(x|N)P(N)/p(x)$, respectively.

In this approach, $P(T)$ and $P(N)$ are (prior) probabilities of the bioactive class and inactive class, respectively, and $p(x|T)$ and $p(x|N)$ are the probability density functions of active and inactive NPs, respectively. The probability density function for the NP dataset is calculated as

$$p(x) = p(x|T)P(T) + p(x|N)P(N).$$

A decision boundary for False negatives and False positives was proposed to minimize the misclassification.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

44 Metal

Metal Oxide

List: Fe₂O₃

Fe₃O₄

Shape: NA

Coating: Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA
 Cross-linked dextran :: NH₂
 Cross-linked dextran :: Alexa Fluor 488
 Cross-linked dextran :: Alexa Fluor 750
 Cross-linked dextran :: FITC, R-COOH
 Cross-linked dextran :: biotin
 Cross-linked dextran :: FITC, COOH
 Cross-linked dextran :: Cy3.5
 Cross-linked dextran :: Cy5.5, protamine
 Cross-linked dextran :: Cy5.5, tat
 Cross-linked dextran :: Cy5.5
 Cross-linked dextran :: Cy5
 Cross-linked dextran :: Cy7
 Cross-linked dextran :: FITC
 Cross-linked dextran :: FITC, Glutamic acid
 Cross-linked dextran :: glycine
 Cross-linked dextran :: rhodamine, protamine
 Cross-linked dextran :: FITC, succinimidyl iodoacetate
 Cross-linked dextran :: Tat peptide
 Cross-linked dextran :: VT680
 Cross-linked dextran :: VT680, protamine
 Dextran :: NA
 Sucrose :: NA
 PVA :: COOH
 PVA :: Ethylene diamine
 PVA :: Ethylene diamine, VT680
 PVA :: protamine, rhodamine
 PVA :: L-arg8-COOH
 PVA :: COOH
 PVA :: AminoSPARK™680
 PVA :: PEG Ethylene diamine, AminoSPARK™680
 PVA :: Ethylene diamine, AminoSPARK™680
 PVA, PEG :: AngioSPARK™680- IVM
 PVA :: 15-mer peptide
 PVA :: L-arg7-COOH
 PVA :: Ethylene diamine, VT750
 PVA, PEG :: Ethylene diamine, VT750
 PVA :: D-arg7-COOH
 PVA, PEG :: NA
 Arabino-galactan :: NA
 Carboxymethyldextran :: NA
 Amphiphilic polymer – PEG :: NH₂
 Amphiphilic polymer :: COOH
Size (nm): 20-74

Other info: Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a

Bruker Minispec MQ20 NMR

6.6.Pre-processing of data before modelling:

The NPs were randomly partitioned into five mutually exclusive subsets, with four subsets used for training and one for validation (i.e., 5-fold cross-validation). This 5-fold cross-validation was repeated 1000 times (which was found to be sufficiently large with respect to the sample size of 44) and the average performance (over the 1000 cross-validation instances) was used as criterion for model assessment.

6.7.Statistics for goodness-of-fit:

Accuracy: 78.1 %

The Accuracy results for the rest of building models can be revised at publication's supplementary material Table S2

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Applied 5-fold cross validation 1000 times.

20 rounds of Y-randomization yielded an average classification of 46.25% (not chance correlation)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal

Metal Oxide

ListFe₂O₃Fe₃O₄Shape:NACoating:Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750

Cross-linked dextran :: FITC, R-COOH

Cross-linked dextran :: biotin

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: Cy3.5

Cross-linked dextran :: Cy5.5, protamine

Cross-linked dextran :: Cy5.5, tat

Cross-linked dextran :: Cy5.5

Cross-linked dextran :: Cy5

Cross-linked dextran :: Cy7

Cross-linked dextran :: FITC

Cross-linked dextran :: FITC, Glutamic acid

Cross-linked dextran :: glycine

Cross-linked dextran :: rhodamine, protamine

Cross-linked dextran :: FITC, succinimidyl iodoacetate

Cross-linked dextran :: Tat peptide

Cross-linked dextran :: VT680

Cross-linked dextran :: VT680, protamine

Dextran :: NA

Sucrose :: NA

PVA :: COOH

PVA :: Ethylene diamine

PVA :: Ethylene diamine, VT680

PVA :: protamine, rhodamine

PVA :: L-arg8-COOH

PVA :: COOH

PVA :: AminoSPARK™680

PVA :: PEG Ethylene diamine, AminoSPARK™680

PVA :: Ethylene diamine, AminoSPARK™680

PVA, PEG :: AngioSPARK™680- IVM

PVA :: 15-mer peptide

PVA :: L-arg7-COOH

PVA :: Ethylene diamine, VT750

PVA, PEG :: Ethylene diamine, VT750

PVA :: D-arg7-COOH

PVA, PEG :: NA

Arabino-galactan :: NA

Carboxymethyldextran :: NA

Amphiphilic polymer – PEG :: NH₂

Amphiphilic polymer :: COOH

Size(nm): 20-74

Other properties:

Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Consensus clustering by self-organizing maps (SOM) was also applied to obtain the previous classification of the data in order to define the end point. The accuracy result obtained by NBC (81.6%) was better than the obtained by H4 classification (78.1%). But, the acceptability of a particular nano-SAR endpoint (i.e., class definition) should not be dictated solely based on nano-SAR accuracy but also determined on the intended nano-SAR use. The H4 class definition was deemed to be a reasonable endpoint choice for a nano-SAR that reduces the level of false negative labelling.

Naïve Bayesian classifier (NBC), logistic regression (LGR), linear discriminate analysis (LDA) and nearest Neighbours (NN) were applied and compared.

Finally NBC was selected as the best suitable building model.

Mechanistic Interpretation of the final descriptors is briefly explained and referenced to previous studies.

NP: nanoparticle

NBC: naïve Bayesian classifier

LGR: logistic regression

LDA: linear discriminate analysis

NN: nearest Neighbours

9.2.Bibliography:

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21), 7387–7392. <http://doi.org/10.1073/pnas.0802878105>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)

Murine RAW 264.7 leukemic monocyte/macrophage cell, QSAR, - R1: spin-lattice Relaxivity
- zeta potential (surface charge),NBC: Naïve Bayesian Classifier

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictive model of TiO₂ NPs damage on membrane cell by SMILES-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictive model of TiO₂ NPs damage on membrane cell by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software

package:

Toropova, A. P., & Toropov, A. A. (2013). Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO₂ nanoparticles. Chemosphere, 93(10), 2650–2655.

Toropova, A. P., Toropov, A. A., Benfenat

<http://doi.org/10.1016/j.chemosphere.2013.09.089>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human Lung epithelial cell

3.2. Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release

[units/L]

3.3.Comment on endpoint:

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release units was expressed in [units/L]

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

Split 1:

A3, A4, A9, B2, B3, C7, D2, D3, D5, D9, E7, E8

Split 2:

A3, A4, A9, B2, B3, B5, C5, C8, C9, D2, D3, D5, D9, E7, E8

Split 3:

A3, A4, A9, B1, B2, B3, B5, B9, C4, C5, C7, C8, C9, D2, D3, D5, D9, E7, E8

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).

-----Descriptor ----- Code -----

- Engineered Size	A
- Size in water	B
- Size in PBS	C
- Concentration	D
- Zeta potential	E

; 0

4.4.Descriptor selection:

Normalization of the data, discriminated classification (A1-2-3...,B-1-2...,C...) and applied optimal descriptors and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :Split1 :
 12:10
 Split2 :
 15:13
 Split3 :
 19:13

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

The domain of applicability for models based on the optimal descriptors can be defined according to prevalence of physicochemical features: one can expect satisfactory prediction for a TiO₂ nanoparticle if physicochemical features of this nanoparticle take place in the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Metal Oxide

List: TiO₂

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

Three various splits into the training and test sets are examined in the present study. These splits obey the following principles:

(i) they are random;

(ii) the ranges of the endpoint for the training and test sets are similar

Three splits built up according to the above-mentioned principles are examined in the present study

6.7.Statistics for goodness-of-fit:

split1 :

$$r^2 = 0.9893$$

$$RMSE = 0.025$$

split2 :

$$r^2 = 0.9639$$

$$RMSE = 0.049$$

split3 :

$$r^2 = 0.9792$$

$$RMSE = 0.049$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

split1 :

$$LOO (q^2) = 0.9845$$

- Y-randomization:

(Rp²) = 0.9362

split2 :

LOO (q²) = 0.9495

- Y-randomization:

(Rp²) = 0.8772

split3 :

LOO (q²) = 0.9718

- Y-randomization:

(Rp²) = 0.9477

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

split1 : 9

split2 : 6

split3 : 6 Metal Oxide

List

TiO₂

Shape:NA

Coating:NA

Size(nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each

nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

split1 :

$$r^2 = 0.8679$$

$$RMSE = 0.115$$

split2 :

$$r^2 = 0.9748$$

$$RMSE = 0.054$$

split3 :

$$r^2 = 0.9925$$

$$RMSE = 0.025$$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The obtained model is a linear regression, the interesting fact is how are selected and managed the descriptors to generate one variable form them for the linear regression.

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the

majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

The same work with an extension of the data set was developed in a posterior work:

Toropova, A. P., Toropov, A. A., Benfenati, E., Puzyn, T., Leszczynska, D., & Leszczynski, J. (2014). Optimal descriptor as a translator of eclectic information into the prediction of membrane damage: The case of a group of ZnO and TiO₂ nanoparticles. *Ecotoxicology and Environmental Safety*, 108, 203–209. <http://doi.org/10.1016/j.ecoenv.2014.07.005>

CORAL: CORrelation And Logic

r^2 : correlation coefficient

RMSE: root-mean-square-error

q^2 : leave-one-out cross-validation correlation coefficient

LOO: Leave-One-Out cross-validation

R_p^2 : Parameter computed from correlations coefficients of Y-scam

9.2.Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Human Lung epithelial cell, QSAR, Split 1:

A3, A4, A9, B2, B3, C7, D2, D3, D5, D9, E7, E8

Split 2:

A3, A4, A9, B2, B3, B5, C5, C8, C9, D2, D3, D5, D9, E7, E8

Split 3:

A3, A4, A9, B1, B2, B3, B5, B9, C4, C5, C7, C8, C9, D2, D3, D5, D9, E7, E8

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).


-----Descriptor ----- Code -----

- Engineered Size A
- Size in water B
- Size in PBS C
- Concentration D
- Zeta potential E

,Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Toxic Effect (Active/Inactive) prediction on embryonic zebrafish due
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Toxic Effect (Active/Inactive) prediction on embryonic zebrafish due wide type of nanomaterials by kNN under ABMiner software

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Xiong Liu

xliu09@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software

package:

Liu, X., Tang, K., Harper, S., Harper, B., Steevens, J. A., & Xu, R. (2013). Predictive modelling of nanomaterial exposure effects in biological systems. International Journal of Nanomedicine, 8 Suppl 1(Supplement 1 Nanoinformatics), 31–43.

<http://doi.org/10.2147/IJN.S40742>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Embryonic zebrafish

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as 24 hours post fertilization (hpf) mortality (M)

3.3. Comment on endpoint:

The defined biological effect (Mortality, Delayed development, Spontaneous movement, Malformations, etc) is defined as the ratio of the number of zebrafish embryos having the effect over the total number of embryos tested. All endpoints were modeled but the most relevant result was classified in this table. For specific details see (in the publication) Table 1 and Table 2 to check all the endpoints on the modelling results

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

IBK (kNN: K-nearest neighbour predictor)

ABMiner tool was used to perform the model. Different prediction algorithms were compared:

IBK

Bagging

M5P

Kstar

4.3.Descriptors in the model:

- material type
- synthesis precursors
- purity
- primary particle size: average (nm)
- primary particle size: minimum (nm)
- primary particle size: maximum (nm)
- core shape
- core structure
- core atomic composition
- number of core atoms
- mass core atoms (ng)
- shell composition
- outermost surface functional groups
- minimum number of ligands
- surface charge: (positive, negative, or neutral)
- solubility/dispersity medium
- primary exposure route
- material zeta potential in media (mV)
- stable average agglomerate size in media (nm)
- dosage concentrations used (ppm); 20

4.4.Descriptor selection:

The descriptors were obtained from NBI:

Nanomaterial-Biological Interactions Knowledgebase [homepage on the Internet]. Corvallis, OR: Nanomaterial-Biological Interactions Knowledgebase. Available from: <http://oregonstate.edu/nbi>. Accessed November 25, 2012.

There was exposed an attribute weights classification for the 20 descriptors, but it was not applied to reduce the number of final descriptors (attributes).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/20

Descriptor: Chemical ratio :20:82 ~ 1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Metal
 Metal oxide
 Dendrimer
 Polymeric
List: NA
Shape: NA
Coating: NA
Size (nm): NA
Other info: NA

6.6. Pre-processing of data before modelling:

656 data points obtained from 82 nanomaterials tested at eight different dosage concentrations.

6.7. Statistics for goodness-of-fit:

$r^2 = 0.837$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

k-fold was applied by ABMiner (not more specifications)

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No
 Chemical Name: not applicable
 SMILES: not applicable
 Formula: not applicable
 INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA
 NP size: NA
 NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MMetal

Metal oxide

Dendrimer

Polymeric

List

NA

Shape: NACoating: NASize(nm): NAOther properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

ABMiner software applied to develop the models. Variety of models, for different endpoints were developed, but only displayed the best ones that were presented.

They implement a model base with a model query interface where you can vary the input parameters in order to predict your results for unsynthesized nanomaterials at:

http://neiminer.i-a-i.com/nei_models (denied access)

There is a lack of validation information(also values) and data set information of employed nanomaterials.

EZ: Embryonic zebrafish

r^2 : Pearson correlation between actual and predicted score

- IBK is a K-nearest neighbour predictor.
- Bagging is a meta-learning algorithm for generating multiple versions of a predictor and using these to get an aggregated

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Organism, Embryonic zebrafish, QSAR, - material type

- synthesis precursors
- purity
- primary particle size: average (nm)
- primary particle size: minimum (nm)
- primary particle size: maximum (nm)
- core shape
- core structure
- core atomic composition
- number of core atoms
- mass core atoms (ng)
- shell composition
- outermost surface functional groups
- minimum number of ligands
- surface charge: (positive, negative, or neutral)
- solubility/dispersity medium
- primary exposure route
- material zeta potential in media (mV)
- stable average agglomerate size in media (nm)
- dosage concentrations used (ppm),IBK (kNN: K-nearest neighbour predictor)

ABMiner tool was used to perform the model. Different prediction algorithms were compared:


IBK

Bagging

M5P

Kstar

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Toxic Effect (Active/Inactive) prediction on embryonic zebrafish due
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Toxic Effect (Active/Inactive) prediction on embryonic zebrafish due wide type of nanomaterials and Weighted EZ Metric score as Endpoint by kNN under ABMiner software

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Xiong Liu

xliu09@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software

package:

Liu, X., Tang, K., Harper, S., Harper, B., Steevens, J. A., & Xu, R. (2013). Predictive modelling of nanomaterial exposure effects in biological systems. International Journal of Nanomedicine, 8 Suppl 1(Supplement 1 Nanoinformatics), 31–43.

<http://doi.org/10.2147/IJN.S40742>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Embryonic zebrafish

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as (M)Weighted EZ Metric score

3.3. Comment on endpoint:

The defined biological effect (Mortality, Delayed development, Spontaneous movement, Malformations, etc) is defined as the ratio of the number of zebrafish embryos having the effect over the total number of embryos tested. Weightened endpoints were added in one final endpoint (Weightened EZ Metric score)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

IBK (kNN: K-nearest neighbour predictor)

ABMiner tool was used to perform the model. Different prediction algorithms were compared:

IBK

Bagging

M5P

Kstar

4.3.Descriptors in the model:

- material type
- synthesis precursors
- purity
- primary particle size: average (nm)
- primary particle size: minimum (nm)
- primary particle size: maximum (nm)
- core shape
- core structure
- core atomic composition
- number of core atoms
- mass core atoms (ng)
- shell composition
- outermost surface functional groups
- minimum number of ligands
- surface charge: (positive, negative, or neutral)
- solubility/dispersity medium
- primary exposure route
- material zeta potential in media (mV)
- stable average agglomerate size in media (nm)
- dosage concentrations used (ppm); 20

4.4.Descriptor selection:

The descriptors were obtained from NBI:

Nanomaterial-Biological Interactions Knowledgebase [homepage on the Internet]. Corvallis, OR: Nanomaterial-Biological Interactions Knowledgebase. Available from: <http://oregonstate.edu/nbi>. Accessed November 25, 2012.

There was exposed an attribute weights classification for the 20 descriptors, but it was not applied to reduce the number of final descriptors (attributes).

4.5. Algorithm and descriptor generation:

No information available

4.6. Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

0/20

Descriptor: Chemical ratio :20:82 ~ 1:4

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Metal
 Metal oxide
 Dendrimer
 Polymeric
List: NA
Shape: NA
Coating: NA
Size (nm): NA
Other info: NA

6.6. Pre-processing of data before modelling:

656 data points obtained from 82 nanomaterials tested at eight different dosage concentrations.

6.7. Statistics for goodness-of-fit:

$r^2 = 0.792$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

k-fold is applied by ABMiner (not more specifications)

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No
 Chemical Name: not applicable
 SMILES: not applicable
 Formula: not applicable
 INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA
 NP size: NA
 NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MMetal

Metal oxide

Dendrimer

Polymeric

List

NA

Shape:NACoating:NASize(nm): NAOther properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

ABMiner software applied to develop the models. Variety of models, for different endpoints were developed, but only displayed the best ones that were presented.

They implement a model base with a model query interface where you can vary the input parameters in order to predict your results for unsynthesized nanomaterials at:

http://neiminer.i-a-i.com/nei_models

There is a lack of validation information(also values) and data set information of employed nanomaterials.

EZ: Embryonic zebrafish

r^2 : Pearson correlation between actual and predicted score

- IBK is a K-nearest neighbour predictor.
- Bagging is a meta-learning algorithm for generating multiple versions of a predictor and using these to get an aggregated

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Organism, Embryonic zebrafish, QSAR, - material type

- synthesis precursors
- purity
- primary particle size: average (nm)
- primary particle size: minimum (nm)
- primary particle size: maximum (nm)
- core shape
- core structure
- core atomic composition
- number of core atoms
- mass core atoms (ng)
- shell composition
- outermost surface functional groups
- minimum number of ligands
- surface charge: (positive, negative, or neutral)
- solubility/dispersity medium
- primary exposure route
- material zeta potential in media (mV)
- stable average agglomerate size in media (nm)
- dosage concentrations used (ppm),IBK (kNN: K-nearest neighbour predictor)

ABMiner tool was used to perform the model. Different prediction algorithms were compared:


IBK

Bagging

M5P

Kstar

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model for nanoparticles uptake of PaCa2 cells by SMILES-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model for nanoparticles uptake of PaCa2 cells by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software

package:

Toropov, A. A., Toropova, A. P., Puzyn, T., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2013). QSAR as a random event: Modelling of nanoparticles uptake in PaCa2 cancer cells. Chemosphere, 92(1), 31–37.

<http://doi.org/10.1016/j.chemosphere.2013.03.012>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3. Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

SMILES was used as input to the CORAL software.

For each split ten the most significant promoters which increases and decreases the endpoint are represented in Table S2 (Supplementary material of the publication)

(10 x 2 = 20 attributes were presented for each split); 0

4.4.Descriptor selection:

SMILES-based optimal descriptors and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The domain of applicability for CORAL models can be defined as nanoparticles (with the same core) which do not contain SMILES attributes absent in the sub-training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No
 Chemical Name: not applicable
 SMILES: not applicable
 Formula: not applicable
 INChI: not applicable
 MOL file: not applicable
 Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes
 NP size: Yes
 NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide
List: (Fe₂O₃)_n(Fe₃O₄)_m
Shape: NA
Coating: Trifluoroacetic anhydride
 Chlorodifluoroacetic anhydride
 Pentafluoropropanoic anhydride
 4 3,3-Dimethyldihydrofuran-2,5-dione
 Furan-2,5-dione
 3-Methylfuran-2,5-dione
 7 3,4-Dimethylfuran-2,5-dione
 Hexanoic anhydride
 3-Methyldihydrofuran-2,5-dione
 5,5'-Carbonylbis(2-benzofuran-1,3-dione)
 5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride

3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine

Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with

ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

The data for these 109 nanoparticles were randomly split into the sub-training, calibration, test, and validation sets. The roles of these sets are different:

- sub-training set (training set in the table) is the “developer” of the model since correlation weights of compounds from the set are used to build up the model
- calibration set is the “critic” of the model since data from this set are used to check whether model is working for compounds which are absent in the sub-training set
- the test set is “estimator” of the model in cases of various threshold values
- validation set (external set in the table) is used for the final estimation of the model with threshold value which gives the best statistical quality for the test set.

These splits are random and various.

Five splits built up according to the above-mentioned principles are examined in the present study

6.7.Statistics for goodness-of-fit:

split1 :

$$r^2 = 0.6913$$

$$MAE = 0.186$$

split2 :

$$r^2 = 0.6972$$

$$MAE = 0.185$$

split3 :

$$r^2 = 0.7287$$

$$MAE = 0.173$$

split4 :

$$r^2 = 0.7557$$

$$MAE = 0.173$$

split5 :

$$r^2 = 0.6504$$

$$MAE = 0.215$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

split1 : 15

split2 : 17

split3 : 20

split4 : 19

split5 : 18 Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine

4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid

2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

split1 :

$r^2 = 0.9341$

MAE = 0.129

split2 :

$r^2 = 0.8043$

MAE = 0.0.143

split3 :

$r^2 = 0.8723$

MAE = 0.148

split4 :

$r^2 = 0.8232$

MAE = 0.112

split5 :

$r^2 = 0.8429$

MAE = 0.153

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The obtained model is a linear regression, the interesting fact is how are selected and managed the descriptors to generate one variable form them for the linear regression.

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

SMILES: Simplified Molecular Input Line Entry Specification

r^2 : correlation coefficient

MAE : mean average error

9.2.Bibliography:

(already reported in this table)

Fourches, D. et al., 2010. Quantitative nanostructure-activity relationship modelling. ACS nano, 4(10), pp.5703–12 (Case Study 2)

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, SMILES was used as input to the CORAL software.

For each split ten the most significant promoters which increases and decreases the endpoint are represented in Table S2 (Supplementary material of the publication)

(10 x 2 = 20 attributes were presented for each split),Linear regression model based on SMILES-based optimal descriptors by the software CORAL.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression-tree-based analysis of rodent pulmonary toxicity by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure.
(PMNs - RT case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Elisabeth A. Casman

casman@andrew.cmu.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Gernand, J. M., & Casman, E. A. (2014). A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal, 34(3), 583–597.

(PMNs - RT case)

<http://doi.org/10.1111/risa.12109>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Rodent's lung exposed via Instillation or Aspiration to CNTs.

3.2. Endpoint:

In vivo - Rodent lung inflammation - Immune response measured by polymorphonuclear neutrophils

(PMNs)

3.3.Comment on endpoint:

The endpoints in the paper reflect several dimensions of immune response and cell membrane damage and death.

These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

Converted all toxicity results to fold of control format (ratio of the desired measure over a control measure), a form that many of the studies already reported.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

RT: Regression Trees

4.3.Descriptors in the model:

- Total mass dose ($\mu\text{g/kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject

- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period

- Dose Fe ($\mu\text{g/kg}$): The total dose received by the animal subject of iron impurities present in the CNT particulate

- Dose Cr ($\mu\text{g/kg}$): The total dose received by the animal subject of chromium impurities present in the CNT particulate

- Dose Co ($\mu\text{g/kg}$): The total dose received by the animal subject of copper impurities present in the CNT particulate; 5

4.4.Descriptor selection:

Initial descriptors selected from the source data (from experimental conditions to nanoparticle properties).

Final descriptors were obtained by the results of the model building algorithm.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

52500/5

Descriptor: Chemical ratio :5:52,500

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The model cannot extrapolate based on trends, however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

52500 Carbon-based

List: CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape: Fiber

Coating: Uncoated

Size (nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other info: The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints

permitted greater contrast to be made between the effects of the different input variables.

6.6.Pre-processing of data before modelling:

10-fold cross validation applied to RT models

6.7.Statistics for goodness-of-fit:

$R^2 = 0.89$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross-validation applied to RT (not more specifications)

From 2 to 6 more columns (descriptors) were added with random normal distributed data between 0 and 1, in order to check the discriminatory performance of the models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape:Fiber

Coating:Uncoated

Size(nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other properties:

The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Coefficients and model performance statistics for stepwise linear regression models were performed and can be revised in Table 1 of the publication's Supplementary Material. Those results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Mechanistic Interpretation was widely explained.

Lack of validation techniques which could give more reliability to the model

RT: Regression Trees

R²: correlation coefficient

MSE: Mean square error

PMNs: polymorphonuclear neutrophils

9.2.Bibliography:

Publication reference indexes list in the paper:

7, 15, 16, 17, 20, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Ellinger-Ziegelbauer et al., 2009

Nygaard 2009

Pauluhn 2010

Muller et al.2005

Ma-Hock et al.2009

Warheit et al.2004

Shvedova et al.2005

Shvedova et al.2007

Shvedova et al.2008

Muller et al.2008

Elgrabli et al.2008

Mercer et al.2008

Inoue et al.2008

Ryman-Rasmussen et al., 2008

Porter et al.2010

Park et al.2011

Teeguarden et al.2011

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Organism, Rodent's lung exposed via Instillation or Aspiration to CNTs., QSAR, - Total mass dose ($\mu\text{g/kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject


- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period

- Dose Fe ($\mu\text{g/kg}$): The total dose received by the animal subject of iron impurities present in the CNT particulate

- Dose Cr ($\mu\text{g/kg}$): The total dose received by the animal subject of chromium impurities present in the CNT particulate

- Dose Co ($\mu\text{g/kg}$): The total dose received by the animal subject of copper impurities present in the CNT particulate, RT: Regression Trees

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression-tree-based analysis of rodent pulmonary toxicity by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure.
(PMNs - RF case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Elisabeth A. Casman

casman@andrew.cmu.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Gernand, J. M., & Casman, E. A. (2014). A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal, 34(3), 583–597.

(PMNs - RF case)

<http://doi.org/10.1111/risa.12109>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Rodent's lung exposed via Instillation or Aspiration to CNTs.

3.2. Endpoint:

In vivo - Rodent lung inflammation - Immune response measured by polymorphonuclear neutrophils

(PMNs)

3.3. Comment on endpoint:

The endpoints in the paper reflect several dimensions of immune response and cell membrane damage and death.

These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

Converted all toxicity results to fold of control format (ratio of the desired measure over a control measure), a form that many of the studies already reported.

3.4. Endpoint units:

See 3.2

3.5. Dependent variable:

See 3.2

3.6. Experimental protocol:

No information available

3.7. Endpoint data quality and variability:

No information available

4. Defining the algorithm - OECD Principle 2

4.1. Type of model:

QSAR

4.2. Explicit algorithm:

RF: Random Forest

4.3. Descriptors in the model:

- Total mass dose ($\mu\text{g/kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject
- Dose Co ($\mu\text{g/kg}$): The total dose received by the animal subject of copper impurities present in the CNT particulate
- Dose Co 24h avg. ($\mu\text{g/kg}$): The average daily dose received by the animal subject of copper impurities present in the CNT particulate
- Mass concentration (mg/m^3): The mass concentration of CNTs in the air of the animal subject inhalation chamber (inhalation exposures only)
- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period; 5

4.4. Descriptor selection:

Initial descriptors selected from the source data (from experimental conditions to nanoparticle properties).

Final descriptors were obtained by the results of the model building algorithm.

4.5. Algorithm and descriptor generation:

No information available

4.6. Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

52500/5

Descriptor: Chemical ratio :5:52,500

5. Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The model cannot extrapolate based on trends, however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

52500 Carbon-based

List: CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape: Fiber

Coating: Uncoated

Size (nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other info: The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints

permitted greater contrast to be made between the effects of the different input variables.

6.6.Pre-processing of data before modelling:

10-fold cross validation applied to RT models

6.7.Statistics for goodness-of-fit:

$R^2 = 0.83$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

From 2 to 6 more columns (descriptors) were added with random normal distributed data between 0 and 1, in order to check the discriminatory performance of the models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape:Fiber

Coating:Uncoated

Size(nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other properties:

The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Some source studies were withheld to be used as test data and the results were expressed in terms of MSE (mean square error)

For specific details see (in the publication) Table IV

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Coefficients and model performance statistics for stepwise linear regression models were performed and can be revised in Table 1 of the publication's Supplementary Material. Those results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Mechanistic Interpretation was widely explained.

Lack of validation techniques which could give more reliability to the model

RF: Random Forest

R^2 : correlation coefficient

MSE: Mean square error

PMNs: polymorphonuclear neutrophils

9.2. Bibliography:

Publication reference indexes list in the paper:

7, 15, 16, 17, 20, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Ellinger-Ziegelbauer et al., 2009

Nygaard 2009

Pauluhn 2010

Muller et al. 2005

Ma-Hock et al. 2009

Warheit et al. 2004

Shvedova et al. 2005

Shvedova et al. 2007

Shvedova et al. 2008

Muller et al. 2008

Elgrabli et al. 2008

Mercer et al. 2008

Inoue et al. 2008

Ryman-Rasmussen et al., 2008

Porter et al. 2010

Park et al. 2011

Teeguarden et al. 2011

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Organism, Rodent's lung exposed via Instillation or Aspiration to CNTs., QSAR, - Total mass dose ($\mu\text{g/kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject

- Dose Co ($\mu\text{g/kg}$): The total dose received by the animal subject of copper impurities present in the CNT particulate


- Dose Co 24h avg. ($\mu\text{g/kg}$): The average daily dose received by the animal subject of copper impurities present in the CNT particulate

- Mass concentration (mg/m^3): The mass concentration of CNTs in the air of the animal subject inhalation chamber (inhalation exposures only)

- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period, RF:

Random Forest

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression-tree-based analysis of rodent pulmonary toxicity by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure.
(MAC - RT case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Elisabeth A. Casman

casman@andrew.cmu.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Gernand, J. M., & Casman, E. A. (2014). A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal, 34(3), 583–597.

(MAC - RT case)

<http://doi.org/10.1111/risa.12109>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Rodent's lung exposed via Instillation or Aspiration to CNTs.

3.2. Endpoint:

In vivo - Rodent lung inflammation - Immune response measured by macrophages (MAC)

3.3.Comment on endpoint:

The endpoints in the paper reflect several dimensions of immune response and cell membrane damage and death.

These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

Converted all toxicity results to fold of control format (ratio of the desired measure over a control measure), a form that many of the studies already reported.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RT: Regression Trees

4.3.Descriptors in the model:

- Mass concentration (mg/m^3): The mass concentration of CNTs in the air of the animal subject inhalation chamber (inhalation exposures only)
- Total surface area dose (m^2/kg): The highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject
- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period; 3

4.4.Descriptor selection:

Initial descriptors selected from the source data (from experimental conditions to nanoparticle properties).

Final descriptors were obtained by the results of the model building algorithm.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

44000/3

Descriptor: Chemical ratio :3:44,000

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

The model cannot extrapolate based on trends, however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

44000 Carbon-based

List: CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape: Fiber

Coating: Uncoated

Size (nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other info: The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

6.6. Pre-processing of data before modelling:

10-fold cross validation applied to RT models

6.7.Statistics for goodness-of-fit:

$R^2 = 0.62$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross-validation applied

From 2 to 6 more columns (descriptors) were added with random normal distributed data between 0 and 1, in order to check the discriminatory performance of the models.

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape:Fiber

Coating:Uncoated

Size(nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other properties:

The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

Coefficients and model performance statistics for stepwise linear regression models were performed and can be revised in Table 1 of the publication's Supplementary Material. Those results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Mechanistic Interpretation was widely explained.

Lack of validation techniques which could give more reliability to the model

RT: Regression Trees

R²: correlation coefficient

MSE: Mean square error

MAC: macrophages

9.2.Bibliography:

Publication reference indexes list in the paper:

7, 15 , 16, 17, 20, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Ellinger-Ziegelbauer et al., 2009

Nygaard 2009

Pauluhn 2010

Muller et al.2005

Ma-Hock et al.2009

Warheit et al.2004

Shvedova et al.2005

Shvedova et al.2007

Shvedova et al.2008

Muller et al.2008

Elgrabli et al.2008

Mercer et al.2008

Inoue et al.2008

Ryman-Rasmussen et al., 2008

Porter et al.2010

Park et al.2011

Teeguarden et al.2011

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Organism, Rodent's lung exposed via Instillation or Aspiration to CNTs., QSAR, - Mass concentration (mg/m³): The mass concentration of CNTs in the air of the animal subject inhalation chamber (inhalation exposures only)

- Total surface area dose (m²/kg): The highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject

- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period,RT:
Regression Trees

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression-tree-based analysis of rodent pulmonary toxicity by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure.
(MAC - RF case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Elisabeth A. Casman

casman@andrew.cmu.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Gernand, J. M., & Casman, E. A. (2014). A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal, 34(3), 583–597.

(MAC - RF case)

<http://doi.org/10.1111/risa.12109>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Rodent's lung exposed via Instillation or Aspiration to CNTs.

3.2. Endpoint:

In vivo - Rodent lung inflammation - Immune response measured by macrophages (MAC)

3.3.Comment on endpoint:

The endpoints in the paper reflect several dimensions of immune response and cell membrane damage and death.

These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

Converted all toxicity results to fold of control format (ratio of the desired measure over a control measure), a form that many of the studies already reported.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RF: Random Forest

4.3.Descriptors in the model:

- Mass concentration (mg/m³): The mass concentration of CNTs in the air of the animal subject inhalation chamber (inhalation exposures only)

- Dose Co (µg/kg): The total dose received by the animal subject of copper impurities present in the CNT particulate

- Dose Co 24h avg. (µg/kg): The average daily dose received by the animal subject of copper impurities present in the CNT particulate; 3

4.4.Descriptor selection:

Initial descriptors selected from the source data (from experimental conditions to nanoparticle properties).

Final descriptors were obtained by the results of the model building algorithm.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

44000/3

Descriptor: Chemical ratio :3:44,000

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

The model cannot extrapolate based on trends, however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

44000 Carbon-based

List: CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape: Fiber

Coating: Uncoated

Size (nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other info: The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

6.6. Pre-processing of data before modelling:

10-fold cross validation applied to RT models

6.7.Statistics for goodness-of-fit:

$R^2 = 0.84$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

From 2 to 6 more columns (descriptors) were added with random normal distributed data between 0 and 1, in order to check the discriminatory performance of the models.

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape:Fiber

Coating:Uncoated

Size(nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other properties:

The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Some source studies were withheld to be used as test data and the results were expressed in terms of MSE (mean square error)

For specific details see (in the publication) Table IV

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Coefficients and model performance statistics for stepwise linear regression models were performed and can be revised in Table 1 of the publication's Supplementary Material. Those results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Mechanistic Interpretation was widely explained.

Lack of validation techniques which could give more reliability to the model

RF: Random Forest

R^2 : correlation coefficient

MSE: Mean square error

MAC: macrophages

9.2.Bibliography:

Publication reference indexes list in the paper:

7, 15 , 16, 17, 20, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Ellinger-Ziegelbauer et al., 2009

Nygaard 2009

Pauluhn 2010

Muller et al.2005

Ma-Hock et al.2009

Warheit et al.2004

Shvedova et al.2005

Shvedova et al.2007

Shvedova et al.2008

Muller et al.2008

Elgrabli et al.2008

Mercer et al.2008

Inoue et al.2008

Ryman-Rasmussen et al., 2008

Porter et al.2010

Park et al.2011

Teeguarden et al.2011

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Organism, Rodent's lung exposed via Instillation or Aspiration to CNTs., QSAR, - Mass concentration (mg/m³): The mass concentration of CNTs in the air of the animal subject inhalation chamber (inhalation exposures only)

- Dose Co (µg/kg): The total dose received by the animal subject of copper impurities present in the CNT particulate

- Dose Co 24h avg. (µg/kg): The average daily dose received by the animal subject of copper impurities present in the CNT particulate,RF: Random Forest

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression-tree-based analysis of rodent pulmonary toxicity by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure.
(LDH -RT case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Elisabeth A. Casman

casman@andrew.cmu.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Gernand, J. M., & Casman, E. A. (2014). A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal, 34(3), 583–597.

(LDH - RT case)

<http://doi.org/10.1111/risa.12109>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Rodent's lung exposed via Instillation or Aspiration to CNTs.

3.2. Endpoint:

In vivo - Rodent lung inflammation - Cell membrane damage measured by lactate dehydrogenase

(LDH) release

3.3.Comment on endpoint:

The endpoints in the paper reflect several dimensions of immune response and cell membrane damage and death.

These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

Converted all toxicity results to fold of control format (ratio of the desired measure over a control measure), a form that many of the studies already reported.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

RT: Regression Trees

4.3.Descriptors in the model:

- Total mass dose ($\mu\text{g/kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject
- Total surface area dose (m^2/kg): The highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject
- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period
- Purity (%): The fraction by percent mass of the amount of the CNT sample composed of carbon atoms
- Dose Fe ($\mu\text{g/kg}$): The total dose received by the animal subject of iron impurities present in the CNT particulate
- Dose Fe 24h avg. ($\mu\text{g/kg}$): The average daily dose received by the animal subject of iron impurities
- Min length (nm): The minimum reported length of the free individual CNT fibers either measured or stated by manufacturer's; 7

4.4.Descriptor selection:

Initial descriptors selected from the source data (from experimental conditions to nanoparticle properties).

Final descriptors were obtained by the results of the model building algorithm.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

58400/7

Descriptor: Chemical ratio :7:58,400

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

The model cannot extrapolate based on trends, however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

58400 Carbon-based

List: CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape: Fiber

Coating: Uncoated

Size (nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other info: The data was obtained from 17 different studies, which were

selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

6.6.Pre-processing of data before modelling:

10-fold cross validation applied to RT models

6.7.Statistics for goodness-of-fit:

$R^2 = 0.84$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross-validation applied to RT (not more specifications)

From 2 to 6 more columns (descriptors) were added with random normal distributed data between 0 and 1, in order to check the discriminatory performance of the models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

CNT: Carbon nanotubes
(Single/Multiwalled nanotubes)
(with the presence of metal impurities)

Shape:Fiber

Coating:Uncoated

Size(nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other properties:

The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Coefficients and model performance statistics for stepwise linear regression models were performed and can be revised in Table 1 of the publication's Supplementary Material. Those results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Mechanistic Interpretation was widely explained.

Lack of validation techniques which could give more reliability to the model

RT: Regression Trees

R^2 : correlation coefficient

MSE: Mean square error

LDH: lactate dehydrogenase

9.2. Bibliography:

Publication reference indexes list in the paper:

7, 15, 16, 17, 20, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Ellinger-Ziegelbauer et al., 2009

Nygaard 2009

Pauluhn 2010

Muller et al. 2005

Ma-Hock et al. 2009

Warheit et al. 2004

Shvedova et al. 2005

Shvedova et al. 2007

Shvedova et al. 2008

Muller et al. 2008

Elgrabli et al. 2008

Mercer et al. 2008

Inoue et al. 2008

Ryman-Rasmussen et al., 2008

Porter et al. 2010

Park et al. 2011

Teeguarden et al. 2011

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Organism, Rodent's lung exposed via Instillation or Aspiration to CNTs., QSAR, - Total mass dose ($\mu\text{g/kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject


- Total surface area dose (m^2/kg): The highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject

- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period

- Purity (%): The fraction by percent mass of the amount of the CNT sample composed of carbon atoms

- Dose Fe ($\mu\text{g/kg}$): The total dose received by the animal subject of iron impurities present in the CNT particulate
- Dose Fe 24h avg. ($\mu\text{g/kg}$): The average daily dose received by the animal subject of iron impurities
- Min length (nm): The minimum reported length of the free individual CNT fibers either measured or stated by manufacturer's, RT: Regression Trees

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression-tree-based analysis of rodent pulmonary toxicity by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure.
(LDH - RF case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Elisabeth A. Casman

casman@andrew.cmu.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Gernand, J. M., & Casman, E. A. (2014). A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal, 34(3), 583–597.

(LDH - RF case)

<http://doi.org/10.1111/risa.12109>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Rodent's lung exposed via Instillation or Aspiration to CNTs.

3.2. Endpoint:

In vivo - Rodent lung inflammation - Cell membrane damage measured by lactate dehydrogenase

(LDH) release

3.3.Comment on endpoint:

The endpoints in the paper reflect several dimensions of immune response and cell membrane damage and death.

These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

Converted all toxicity results to fold of control format (ratio of the desired measure over a control measure), a form that many of the studies already reported.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

RF: Random Forest

4.3.Descriptors in the model:

- Total mass dose ($\mu\text{g/kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject
- Total surface area dose (m^2/kg): The highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject
- Dose Co ($\mu\text{g/kg}$): The total dose received by the animal subject of copper impurities present in the CNT particulate
- Exposure period (hr): first to last exposure period. The time period in hours between the first hour of exposure and the last hour of exposure by the animal subject to CNTs
- 24h avg. of surface area dose (m^2/kg): The average daily surface area dose of CNTs received over the course of the experiment (first to last exposure period) by the animal subject; 5

4.4.Descriptor selection:

Initial descriptors selected from the source data (from experimental conditions to nanoparticle properties).

Final descriptors were obtained by the results of the model building algorithm.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

58400/5

Descriptor: Chemical ratio :5:58,400

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The model cannot extrapolate based on trends, however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

58400 Carbon-based

List: CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape: Fiber

Coating: Uncoated

Size (nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other info: The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints

permitted greater contrast to be made between the effects of the different input variables.

6.6.Pre-processing of data before modelling:

10-fold cross validation applied to RT models

6.7.Statistics for goodness-of-fit:

$R^2 = 0.89$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross-validation applied to RT (not more specifications)

From 2 to 6 more columns (descriptors) were added with random normal distributed data between 0 and 1, in order to check the discriminatory performance of the models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape:Fiber

Coating:Uncoated

Size(nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other properties:

The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Some source studies were withheld to be used as test data and the results were expressed in terms of MSE (mean square error)

For specific details see (in the publication) Table IV

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Coefficients and model performance statistics for stepwise linear regression models were performed and can be revised in Table 1 of the publication's Supplementary Material. Those results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Mechanistic Interpretation was widely explained.

Lack of validation techniques which could give more reliability to the model

RF: Random Forest

R^2 : correlation coefficient

MSE: Mean square error

LDH: lactate dehydrogenase

9.2. Bibliography:

Publication reference indexes list in the paper:

7, 15, 16, 17, 20, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Ellinger-Ziegelbauer et al., 2009

Nygaard 2009

Pauluhn 2010

Muller et al. 2005

Ma-Hock et al. 2009

Warheit et al. 2004

Shvedova et al. 2005

Shvedova et al. 2007

Shvedova et al. 2008

Muller et al. 2008

Elgrabli et al. 2008

Mercer et al. 2008

Inoue et al. 2008

Ryman-Rasmussen et al., 2008

Porter et al. 2010

Park et al. 2011

Teeguarden et al. 2011

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Organism, Rodent's lung exposed via Instillation or Aspiration to CNTs., QSAR, - Total mass dose ($\mu\text{g}/\text{kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject

- Total surface area dose (m^2/kg): The highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject


- Dose Co ($\mu\text{g}/\text{kg}$): The total dose received by the animal subject of copper impurities present in the CNT particulate

- Exposure period (hr): first to last exposure period. The time period in hours between the first hour of exposure and the last hour of exposure by the animal subject to CNTs

- 24h avg. of surface area dose (m^2/kg): The average daily surface area dose of CNTs received over the course of the experiment (first to last exposure period) by the animal subject, RF: Random

Forest

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression-tree-based analysis of rodent pulmonary toxicity by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure.
(TP - RT case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Elisabeth A. Casman

casman@andrew.cmu.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Gernand, J. M., & Casman, E. A. (2014). A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal, 34(3), 583–597.

(TP case)

<http://doi.org/10.1111/risa.12109>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Rodent's lung exposed via Instillation or Aspiration to CNTs.

3.2. Endpoint:

In vivo - Rodent lung inflammation - Cell death measured by total protein (TP)

3.3.Comment on endpoint:

The endpoints in the paper reflect several dimensions of immune response and cell membrane damage and death.

These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

Converted all toxicity results to fold of control format (ratio of the desired measure over a control measure), a form that many of the studies already reported.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RT: Regression Trees

4.3.Descriptors in the model:

- 24h avg. of surface area dose (m^2/kg): The average daily surface area dose of CNTs received over the course of the experiment (first to last exposure period) by the animal subject
- Total mass dose ($\mu\text{g}/\text{kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject
- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period
- Length median (nm): The median length of the free individual CNT fibers either measured or stated by manufacturer's specification
- Dose Co ($\mu\text{g}/\text{kg}$): The total dose received by the animal subject of copper impurities present in the CNT particulate
- Dose Co 24h avg. ($\mu\text{g}/\text{kg}$): The average daily dose received by the animal subject of copper impurities present in the CNT particulate
- Exposure hours (h): The number of hours that the animal subject was exposed to the CNTs
- Configuration (SW/MW): A categorical variable indicating whether the CNTs are multi walled (MWCNTs) or single walled (SWCNTs)
- Mean animal mass (g): The mean mass of the animal subjects in a given experiment
- Specific surface area (m^2/g): Specific surface area as measured by the N2-BET (Nitrogen, Brunauer–Emmett–Teller) gas adsorption method
- Purity (%): The fraction by percent mass of the amount of the CNT sample composed of carbon atoms
- Dose Fe ($\mu\text{g}/\text{kg}$): The total dose received by the animal subject of iron impurities present in the CNT particulate
- Dose Al ($\mu\text{g}/\text{kg}$): The total dose received by the animal subject of aluminum impurities present in the CNT particulate; 13

4.4.Descriptor selection:

Initial descriptors selected from the source data (from experimental conditions to nanoparticle properties).

Final descriptors were obtained by the results of the model building algorithm.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

52400/13

Descriptor: Chemical ratio :13:52,400

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The model cannot extrapolate based on trends, however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data (Carbon nanotubes)

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

52400 Carbon-based

List: CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)
(with the presence of metal impurities)

Shape: Fiber

Coating: Uncoated

Size (nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other info: The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

6.6.Pre-processing of data before modelling:

10-fold cross validation applied to RT models

6.7.Statistics for goodness-of-fit:

$R^2 = 0.92$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross-validation applied to RT (not more specifications)

From 2 to 6 more columns (descriptors) were added with random normal distributed data between 0 and 1, in order to check the discriminatory performance of the models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape:Fiber

Coating:Uncoated

Size(nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other properties:

The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Coefficients and model performance statistics for stepwise linear regression models were performed and can be revised in Table 1 of the publication's Supplementary Material. Those results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Mechanistic Interpretation was widely explained.

Lack of validation techniques which could give more reliability to the model

RT: Regression Trees

R²: correlation coefficient

MSE: Mean square error

TP: total protein

9.2.Bibliography:

Publication reference indexes list in the paper:

7, 15, 16, 17, 20, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Ellinger-Ziegelbauer et al., 2009

Nygaard 2009

Pauluhn 2010

Muller et al.2005

Ma-Hock et al.2009

Warheit et al.2004

Shvedova et al.2005

Shvedova et al.2007

Shvedova et al.2008

Muller et al.2008

Elgrabli et al.2008

Mercer et al.2008

Inoue et al.2008

Ryman-Rasmussen et al., 2008

Porter et al.2010

Park et al.2011

Teeguarden et al.2011

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Organism, Rodent's lung exposed via Instillation or Aspiration to CNTs., QSAR, - 24h avg. of surface area dose (m^2/kg): The average daily surface area dose of CNTs received over the course of the experiment (first to last exposure period) by the animal subject

- Total mass dose ($\mu\text{g}/\text{kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject

- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period

- Length median (nm): The median length of the free individual CNT fibers either measured or stated by manufacturer's specification

- Dose Co ($\mu\text{g}/\text{kg}$): The total dose received by the animal subject of copper impurities present in the CNT particulate

- Dose Co 24h avg. ($\mu\text{g}/\text{kg}$): The average daily dose received by the animal subject of copper impurities present in the CNT particulate

- Exposure hours (h): The number of hours that the animal subject was exposed to the CNTs

- Configuration (SW/MW): A categorical variable indicating whether the CNTs are multi walled (MWCNTs) or single walled (SWCNTs)

- Mean animal mass (g): The mean mass of the animal subjects in a given experiment


- Specific surface area (m^2/g): Specific surface area as measured by the N₂-BET (Nitrogen, Brunauer–Emmett–Teller) gas adsorption method

- Purity (%): The fraction by percent mass of the amount of the CNT sample composed of carbon atoms

- Dose Fe ($\mu\text{g}/\text{kg}$): The total dose received by the animal subject of iron impurities present in the CNT particulate

- Dose Al ($\mu\text{g}/\text{kg}$): The total dose received by the animal subject of aluminum impurities present in the CNT particulate, RT: Regression Trees

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Regression-tree-based analysis of rodent pulmonary toxicity by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Regression-tree-based analysis of rodent pulmonary toxicity by nanotube exposure.
(TP - RF case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Elisabeth A. Casman

casman@andrew.cmu.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Gernand, J. M., & Casman, E. A. (2014). A meta-analysis of carbon nanotube pulmonary toxicity studies-how physical dimensions and impurities affect the toxicity of carbon nanotubes. Risk Anal, 34(3), 583–597.

(TP - RF case)

<http://doi.org/10.1111/risa.12109>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Rodent's lung exposed via Instillation or Aspiration to CNTs.

3.2. Endpoint:

In vivo - Rodent lung inflammation - Cell death measured by total protein (TP)

3.3.Comment on endpoint:

The endpoints in the paper reflect several dimensions of immune response and cell membrane damage and death.

These indicators were all measured in bronchoalveolar lavage (BAL) fluid extracted from the lungs of the mice or rats, and were reported as a counts per subject or fold of control measurements (the average indicator count or concentration in animal test subjects divided by the average count or concentration in control animals).

Converted all toxicity results to fold of control format (ratio of the desired measure over a control measure), a form that many of the studies already reported.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RF: Random Forest

4.3.Descriptors in the model:

- 24h avg. of surface area dose (m^2/kg): The average daily surface area dose of CNTs received over the course of the experiment (first to last exposure period) by the animal subject
- Total surface area dose (m^2/kg): The highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject
- Total mass dose ($\mu\text{g}/\text{kg}$): The total mass dose of CNTs received over the course of the experiment by the animal subject
- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period
- Dose Co 24h avg. ($\mu\text{g}/\text{kg}$): The average daily dose received by the animal subject of copper impurities present in the CNT particulate; 5

4.4.Descriptor selection:

Initial descriptors selected from the source data (from experimental conditions to nanoparticle properties).

Final descriptors were obtained by the results of the model building algorithm.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

52400/5

Descriptor: Chemical ratio :5:52,400

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

The model cannot extrapolate based on trends, however, and can only be used in this manner for combinations of inputs that lie within the limits of the training data (Carbon nanotubes)

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

52400 Carbon-based

List: CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape: Fiber

Coating: Uncoated

Size (nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other info: The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different

input variables.

6.6.Pre-processing of data before modelling:

10-fold cross validation applied to RT models

6.7.Statistics for goodness-of-fit:

$R^2 = 0.95$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross-validation applied to RT (not more specifications)

From 2 to 6 more columns (descriptors) were added with random normal distributed data between 0 and 1, in order to check the discriminatory performance of the models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

CNT: Carbon nanotubes

(Single/Multiwalled nanotubes)

(with the presence of metal impurities)

Shape:Fiber

Coating:Uncoated

Size(nm): Median Length: 550 - 100,000

Median Diameter: 0.8 - 49

Other properties:

The data was obtained from 17 different studies, which were selected under screening criteria: To be included, studies had to report at least minimal CNT characterization and quantitative toxicity output measures, at least one of which also occurred in another published study. Not included studies whose endpoints were the presence or absence of gross pathologies because limiting the data set to studies with continuous toxicity endpoints permitted greater contrast to be made between the effects of the different input variables.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Some source studies were withheld to be used as test data and the results were expressed in terms of MSE (mean square error)

For specific details see (in the publication) Table IV

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Coefficients and model performance statistics for stepwise linear regression models were performed and can be revised in Table 1 of the publication's Supplementary Material. Those results provide another perspective on input variable importance, however the amount of data excluded from these models reduces the confidence as compared to the RT and RF models.

Mechanistic Interpretation was widely explained.

Lack of validation techniques which could give more reliability to the model

RF: Random Forest

R^2 : correlation coefficient

MSE: Mean square error

TP: total protein

9.2. Bibliography:

Publication reference indexes list in the paper:

7, 15, 16, 17, 20, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Ellinger-Ziegelbauer et al., 2009

Nygaard 2009

Pauluhn 2010

Muller et al. 2005

Ma-Hock et al. 2009

Warheit et al. 2004

Shvedova et al. 2005

Shvedova et al. 2007

Shvedova et al. 2008

Muller et al. 2008

Elgrabli et al. 2008

Mercer et al. 2008

Inoue et al. 2008

Ryman-Rasmussen et al., 2008

Porter et al. 2010

Park et al. 2011

Teeguarden et al. 2011

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Organism, Rodent's lung exposed via Instillation or Aspiration to CNTs., QSAR, - 24h avg. of surface area dose (m^2/kg): The average daily surface area dose of CNTs received over the course of the experiment (first to last exposure period) by the animal subject


- Total surface area dose (m^2/kg): The highest peak hourly surface area dose of CNTs received over the course of the experiment by the animal subject

- Total mass dose ($\mu g/kg$): The total mass dose of CNTs received over the course of the experiment by the animal subject

- Post exposure (days): The number of days between the final exposure to CNTs and the sacrifice and measurement of the toxicity status of the subject, also referred to as recovery period

- Dose Co 24h avg. ($\mu g/kg$): The average daily dose received by the animal subject of copper impurities present in the CNT particulate, RF: Random Forest

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Simultaneous prediction of ecotoxic effects of nanoparticles under
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions by a perturbational approach and LDA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A. Speck-Planche

M.N.D.S. Cordeiro

alejspivanovich@gmail.com

ncordeir@fc.up.pt

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Kleandrova, V. V, Luan, F., González-Díaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., & Cordeiro, M. N. D. S. (2014).

Computational ecotoxicology: Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions.

Env

<http://doi.org/10.1016/j.envint.2014.08.009>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Vibrio fischeri

Saccharomyces cerevisiae S288C
Desmodesmus subspicatus
Chlorella sp.
Chlorella vulgaris
Scenedesmus sp.
Pseudokirchneriella subcapitata
Brassica napus
Cucumis sativus
Raphanus sativus
Lolium perenne
Daphnia magna (neonates)
Thamnocephalus

3.2.Endpoint:

In vivo - Ecotoxicological endpoint - measured as binary classification into ecotoxic ("1") or non-ecotoxic ("-1") class which was obtained by different ecotoxic units (EC50, IC50, TC50 and LC50)

3.3.Comment on endpoint:

The nanoparticles/cases were considered as non-ecotoxic [$Tox_i(c_j) = 1$] when they exhibited high values of measures of ecotoxicity; otherwise, they were selected as ecotoxic [$Tox_i(c_j) = -1$]. $Tox_i(c_j)$ is a categorical variable that is used to classify nanoparticles as non-ecotoxic or ecotoxic, and the assignments for all the cases were realized by taking into account certain arbitrary (but rigorous) cutoff values of ecotoxicity. For specific details see (in the publication) Table 1.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

LDA applied to the perturbation approach obtained equation.

by software STATISTICA 6.0

The final model will be a consensus prediction, after apply the desired NP to the 85 NPs used as reference NP.

4.3.Descriptors in the model:

- $Toxi(c_j)_{rf}$ -> Binary (classification) variable reflecting the ecotoxicity of the nanoparticle used as reference.
- $\Delta\Delta V(mt)$ -> Perturbation term which accounts for the changes in the molar volume between the new (output or final state) nanoparticle and that used as reference, also depending on the measures of ecotoxicity.
- $\Delta\Delta V(ao)$ -> Perturbation term which accounts for the changes in the molar volume between the new (output or final state) nanoparticle and that used as reference, also depending on the assay organisms.
- $\Delta\Delta E(ao)$ -> Perturbation term which accounts for the changes in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the assay

organisms.

- $\Delta\Delta E(cp)$ -> Perturbation term which accounts for the changes in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the condition under which the size of each nanoparticle was measured.

- $\Delta\Delta E(te)$ -> Perturbation term which accounts for the changes in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the time during which the assays were carried out.

- $\Delta\Delta P(ps)$ -> Perturbation term which accounts for the changes in the polarizability between the new (output or final state) nanoparticle and that used as reference, also depending on the shapes of the nanoparticles.

- $\Delta\Delta L(cp)$ -> Perturbation term which accounts for the changes in the size between the new (output or final state) nanoparticle and that used as reference, also depending on the condition under which the size of each nanoparticle was measured.

; 8

4.4.Descriptor selection:

If the original descriptors (V, E, P, and L) are used, they will not be able to discriminate the ecotoxic effect of a defined nanoparticle by varying the experimental condition c_j . In this sense, the moving average approach (MAA) was applied.

A perturbation approach was applied (Gonzalez-Diaz et al., 2013) to the original dataset of 85 nanoparticles/cases. To do so, 5520 pairs were randomly chosen from the 85 different nanoparticles, being in each pair one nanoparticle taken as the initial state (reference) and the other one to be predicted (new, final or output state)

see equations 1, 2 and 3 in the publication

Building model was applied using step-wise procedure (which will affect on the descriptors that will be available in the final model).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

4133/8

Descriptor: Chemical ratio :8:85 ~ 1:10

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal and metal oxide NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No
Chemical Name: not applicable
SMILES: not applicable
Formula: not applicable
INChI: not applicable
MOL file: not applicable
Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes
NP size: Yes
NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

4133 Metal

Metal Oxide

List: TiO₂

ZnO

CuO

Al₂O₃

NiO

Al

Cu

Ni

Ag

Fe₃O₄

CeO₂

Zn

La₂O₃

Yb₂O₃

Gd₂O₃

Fe₂O₃

Co

Fe

Shape: Spherical

Pyramidal

Irregular

Elliptical

Needle

Strip

Coating: NA

Size (nm): 3-100

Other info: Original data comprised in 85 cases of nanoparticles, which were retrieved from the literature:

See supplementary material (mmc1.xls) for more details.

6.6.Pre-processing of data before modelling:

5520 pairs were randomly chosen from the 85 different nanoparticles, being in each pair one nanoparticle taken as the initial state (reference) and the other one to be predicted (new, final or output state).

The training set was used to search for the QSAR-perturbation model, containing 4133 cases, with 1949 of them considered as non-ecotoxic and 2184 ecotoxic. The prediction (validation) set was employed to demonstrate the predictive power of the model. This set was formed by 1387 cases, 648 non-ecotoxic and 739 ecotoxic cases.

Since the validation data that they have used are pairs from the already used NPs as part of other pairs, we have classified them as test set and the remaining 3 totally independent NPs of Niquel used to demonstrate the applicability of the model

6.7.Statistics for goodness-of-fit:

Training set:

- Sensitivity = 99.28 %
- Specificity = 98.90 %
- ROC_AUC = 0.9996

Test set:

- Sensitivity = 99.23 %
- Specificity = 99.19 %
- ROC_AUC = 0.9997

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

3 Metal

Metal Oxide

List

TiO₂

ZnO

CuO

Al₂O₃

NiO

Al

Cu

Ni

Ag

Fe₃O₄

CeO₂

Zn

La₂O₃

Yb₂O₃

Gd₂O₃

Fe₂O₃

Co

Fe

Shape: Spherical

Pyramidal

Irregular

Elliptical

Needle

Strip

Coating: NA

Size(nm): 3-100

Other properties:

Original data comprised in 85 cases of nanoparticles, which were retrieved from the literature:

See supplementary material (mmc1.xls) for more details.

7.6. Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Niquel-based NPs:

30nm

- Accuracy = 55.29 %

60 nm

- Accuracy = 65.8 %

100 nm

- Accuracy = 78.82 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

New kind of model was presented, that differs from the normal QSARs since they take into account the experimental results that could affect on the physical measured properties for the model.

One should notice here that they are focusing their study on uncoated nanoparticles, because it is important to evaluate and predict first the real ecotoxic effects of nanoparticles in their bare forms (uncoated) to gather a baseline reference. Then, future studies can be devoted to the influence of other factors such as the presence of coating agents, and light irradiation.

Huge external data was expected to check the reliability of the model.

Mechanistic Interpretation briefly explained.

The same methodology and procedure with different set of NPs was applied by the same group in a posterior work:

Horev-Azaria, L., Baldi, G., Beno, D., Bonacchi, D., Golla-Schindler, U., Kirkpatrick, J. C., ... Korenstein, R. (2013). Predictive Toxicology of cobalt ferrite nanoparticles: Comparative in-vitro study of different cellular models using methods of knowledge discovery from data. *Particle and Fibre Toxicology*, 10(1). <http://doi.org/10.1186/1743-8977-10-32>

NP: Nanoparticle

EC50: Effective concentration of the nanoparticle which inhibits at 50% the growth of the assay organism.

IC50: Concentration of the nanoparticle which inhibits the root elongation of the assay organism (plants) at 50%.

TC50: Concentr

9.2.Bibliography:

Original data comprised in 85 cases of nanoparticles, which were retrieved from the literature:

(Bar-Ilan et al., 2009; García et al., 2011; Gong et al., 2011; Griffitt et al., 2008; Heinlaan et al., 2008; Hund-Rinke and Simon, 2006; Kasemets et al., 2009; Lin and Xing, 2007; Ma et al., 2010, 2011; Marsalek et al., 2012; Sadiq et al., 2011; Zhu et al., 2008, 2009, 2010, 2012)

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Organism, *Vibrio fischeri*

Saccharomyces cerevisiae S288C

Desmodesmus subspicatus

Chlorella sp.

Chlorella vulgaris

Scenedesmus sp.

Pseudokirchneriella subcapitata

Brassica napus

Cucumis sativus

Raphanus sativus

Lolium perenne

Daphnia magna (neonates)

Thamnocephalus, QSAR, - Toxi(cj)rf -> Binary (classification) variable reflecting the ecotoxicity of the nanoparticle used as reference.

- $\Delta\Delta V(mt)$ -> Perturbation term which accounts for the changes in the molar volume between the new (output or final state) nanoparticle and that used as reference, also depending on the measures of ecotoxicity.

- $\Delta\Delta V(ao)$ -> Perturbation term which accounts for the changes in the molar volume between the new (output or final state) nanoparticle and that used as reference, also depending on the assay organisms.

- $\Delta\Delta E(ao)$ -> Perturbation term which accounts for the changes in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the assay organisms.

- $\Delta\Delta E(cp)$ -> Perturbation term which accounts for the changes in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the condition under which the size of each nanoparticle was measured.

- $\Delta\Delta E(te)$ -> Perturbation term which accounts for the changes in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the time during which the assays were carried out.

- $\Delta\Delta P(ps)$ -> Perturbation term which accounts for the changes in the polarizability between the new (output or final state) nanoparticle and that used as reference, also depending on the shapes of the nanoparticles.


- $\Delta\Delta L(cp)$ -> Perturbation term which accounts for the changes in the size between the new (output or final state) nanoparticle and that used as reference, also depending on the condition under which the size of each nanoparticle was measured.

,LDA applied to the perturbation approach obtained equation.

by software STATISTICA 6.0

The final model will be a consensus prediction, after apply the desired NP to the 85 NPs used as reference NP.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Extended version of : Simultaneous prediction of ecotoxic effects of
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Extended version of : Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions by a perturbational approach and LDA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A. Speck-Planche

M.N.D.S. Cordeiro

alejspivanovich@gmail.com

ncordeir@fc.up.pt

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Kleandrova, V. V, Luan, F., González-Díaz, H., Ruso, J. M.,
Speck-Planche, A., & Cordeiro, M. N. D. S. (2014). Computational
tool for risk assessment of nanomaterials: Novel QSTR-
perturbation model for simultaneous prediction of ecotoxicity and
cytotoxicity

<http://doi.org/10.1021/es503861x>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Saccharomyces cerevisiae
Vibrio fischeri
Tetrahymena thermophila
Danio rerio (embryos)
Desmodesmus subspicatus
Chlorella sp.
Chlorella vulgaris
Scenedesmus sp.
Pseudokirchneriella subcapitata
Brassica napus
Cucumis sativus
Raphanus sativus
Lolium perenne

3.2.Endpoint:

In vivo and In vitro - Toxic Effect - measured as binary classification into toxic ("1") or non-toxic ("-1") class which was obtained by different ecotoxic units (CC50, EC50, IC50, TC50 and LC50)

3.3.Comment on endpoint:

Classes were related with the toxic effect of a NP i in a defined experimental condition c_j [$TE_i(c_j)$]. Thus, a case was chosen as nontoxic [$TE_i(c_j) = 1$] when it exhibited a high value of measured toxicity; otherwise, the compound was considered as toxic [$TE_i(c_j) = -1$]. All these assignments were realized according to cutoff values, which are represented in Table 1 in the publication

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

LDA applied to the perturbation approach obtained equation.

by software STATISTICA 6.0

The final model will be a consensus prediction, after apply the desired NP to the 229 NPs used as reference NP.

4.3.Descriptors in the model:

- $TE_i(c_j)_{rf}$ -> binary (classification) variable expressing the toxic effect of the nanoparticle used as reference

- $\Delta\Delta V(me)$ -> perturbation term that characterizes the variations in the molar volume between the new (output or final state) nanoparticle and the other used as reference, also depending on the measures of the toxic effects

- $\Delta\Delta E(bt)$ -> perturbation term that describes the changes in the electronegativity between the new (output or final state) nanoparticle and the other used as reference, also depending on the biological targets

- $\Delta\Delta E(dm)$ -> perturbation term that describes the variations in the electronegativity between the

new (output or final state) nanoparticle and the other used as reference, also depending on the conditions under which the sizes of the nanoparticles were measured

- $\Delta\Delta P(bt)$ -> perturbation term that accounts for the variations in the polarizability between the new (output or final state) nanoparticle and the other used as reference, also depending on the biological targets

- $\Delta\Delta P(ns)$ -> perturbation term that characterizes the changes in the polarizability between the new (output or final state) nanoparticle and the other used as reference, also depending on the shapes of the nanoparticles

- $\Delta\Delta L(ta)$ -> perturbation term that accounts for the changes in the size between the new (output or final state) nanoparticle and the other used as reference, also depending on the intervals of time during which the biological targets were exposed to the nanoparticles

- $\Delta G\mu_3(Hyd)$ -> perturbation spectral moment of order 3, weighted by the hydrophobicity, and characterizing the differences between the chemical structure of the coating agent used in the new (output or final state) nanoparticle, and the coating agent used for the reference nanoparticle

- $\Delta G\mu_5(PSA)$ -> perturbation spectral moment of order 5, weighted by the polar surface area, and describing the differences between the chemical structure of the coating agent used in the new (output or final state) nanoparticle, and the coating agent used for the reference nanoparticle

; 9

4.4.Descriptor selection:

If the original descriptors (V, E, P, and L) are used, they will not be able to discriminate the ecotoxic effect of a defined nanoparticle by varying the experimental condition c_j . In this sense, the moving average approach (MAA) was applied.

A perturbation approach was applied (Gonzalez-Diaz et al., 2013) to the original dataset of 229 nanoparticles/cases. To do so, 36488 pairs were randomly chosen from the 229 different nanoparticles, being in each pair one nanoparticle taken as the initial state (reference) and the other one to be predicted (new, final or output state)

(see equations 1, 2, 3, and 4 in the publication)

Molecular descriptors (coating related) were calculated from order 1 to 6, being weighted by physicochemical properties (PP) such as hydrophobicity and polar surface area. The $\mu_k(PP)$ descriptors were calculated with the software MODELSLAB.

Building model was applied using forward wise procedure (which will affect on the descriptors that will be available in the final model).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

27347/9

Descriptor: Chemical ratio :9:229 ~ 1:25

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal and metal oxide NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

27347 Metal

Metal Oxide

List: SiO₂

Al

CoFe₂O₄

Al₂O₄

Al₂O₃

TiO₂

ZnO

MoO₃

Cu

Fe₃O₄

NiO

Ag

Cr₂O₃

Fe₂O₃

Ca₅(PO₄)₃(OH)

Y₂O₃

ZnFe₂O₄

CuO

Au

Mn₂O₃

Ni

Co

Si

Ge

CdTe

CeO₂

Zn

La₂O₃

Yb₂O₃

Gd₂O₃

Pt

Fe

Shape: Spherical

pseudo-spherical

Pyramidal

Irregular

Elliptical

Needle

slice-shaped

Strip

rod

polyhedral

Coating: CTAB

propylammonium fragment

undecylazide fragment

N,N,N-trimethyl-3(1-propene) ammonium fragment

sodium citrate

11-mercaptoundecanoic acid

PVP

thioglycolic acid

PVA

N -acetylcysteine

PEG-Si(OMe)₃

potato starch

Size (nm): 1.6-123

Other info: Original data comprised in 229 (combination of core NPs with the different coatings) cases of nanoparticles, which were retrieved from the literature.

See supplementary material (es503861x_si_001.xls) for more details.

6.6.Pre-processing of data before modelling:

Final data set, which comprises 36488 cases (NP–NP pairs), was randomly split into two series: training and prediction (validation or test) sets. The training set was employed to generate the QSTR-perturbation model, being formed by 27347 cases, 17560 of them assigned as nontoxic and 9787 toxic. The prediction (validation or test) set was used to assess the predictive power of the model. This set encompassed 9141 cases, 5880 nontoxic and 3261 toxic. It should be specifically detailed here that all the cases belonging to the prediction set were never used in the training set.

Since the validation data that they have used are pairs from the already used NPs as part of other pairs, we have classified them as test set and the remaining 3 totally independent NPs of Niquel used to demonstrate the applicability of the model

6.7.Statistics for goodness-of-fit:

Training set:

- Accuracy = 98.35 %
- Sensitivity = 98.45 %
- Specificity = 98.16 %
- ROC_AUC = 0.999

Test set:

- Accuracy = 98.95 %
- Sensitivity = 98.34 %
- Specificity = 98.73 %
- ROC_AUC = 0.999

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

9 Metal

Metal Oxide

List

SiO₂

Al

CoFe₂O₄

Al₂O₃

Al₂O₃

TiO₂

ZnO

MoO₃

Cu

Fe₃O₄

NiO

Ag

Cr₂O₃

Fe₂O₃

Ca₅(PO₄)₃(OH)

Y₂O₃

ZnFe₂O₄

CuO

Au

Mn₂O₃

Ni

Co

Si

Ge

CdTe

CeO₂

Zn

La₂O₃

Yb₂O₃

Gd₂O₃

Pt

Fe

Shape:Spherical

pseudo-spherical

Pyramidal

Irregular

Elliptical

Needle

slice-shaped

Strip

rod
 polyhedral
Coating: CTAB
 propylammonium fragment
 undecylazide fragment
 N,N,N-trimethyl-3(1-propene) ammonium fragment
 sodium citrate
 11-mercaptoundecanoic acid
 PVP
 thioglycolic acid
 PVA
 N -acetylcysteine
 PEG-Si(OMe)₃
 potato starch
Size(nm): 1.6-123
Other properties:

Original data comprised in 229 (combination of core NPs with the different coatings) cases of nanoparticles, which were retrieved from the literature.

See supplementary material (es503861x_si_001.xls) for more details.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Ag-43.4 nm - CCT = 100 %
 Ag-62.6 nm - CCT = 73.80 %
 Ag-46.3 nm - CCT = 89.96 %
 NiFe₂O₄-97 nm - CCT = 98.69 %
 Fe₂O₃-30 nm - CCT = 79.04-86.03%
 (at different exposure times)

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

An extension methodology from the previous reported paper in the table (Kleandrova, V. V, Luan, F., González-Díaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., & Cordeiro, M. N. D. S. (2014).

Computational ecotoxicology: Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions. *Environment International*, 73, 288–294.)

Here, they increase the number of data, including several number of new NPs, some of them with coatings, also there are an increase of organisms.

The most interesting difference is the new descriptor related with the coatings and the increase in the amount of input data and the external used data.

Mechanistic Interpretation briefly explained.

NP: Nanoparticle

CC50: cytotoxic concentration of the nanoparticle leading to 50% reduction in cell viability assays

EC50: effective concentration of the nanoparticle that inhibits at 50% the growth of the living system.

IC50: concentration of the n

9.2. Bibliography:

Original data comprised in 229 cases of nanoparticles, which were retrieved from the literature: (references 23 to 77 in the publication)

Ha, S. W.; Sikorski, J. A.; Weitzmann, M. N.; Beck, G. R., Jr. Bio-active engineered 50 nm silica nanoparticles with bone anabolic activity: therapeutic index, effective concentration, and cytotoxicity profile in vitro *Toxicol. In Vitro* 2014, 28 (3) 354– 364

Hussain, S. M.; Hess, K. L.; Gearhart, J. M.; Geiss, K. T.; Schlager, J. J. In vitro toxicity of nanoparticles in BRL 3A rat liver cells *Toxicol. In Vitro* 2005, 19 (7) 975– 983

Salunkhe, A. B.; Khot, V. M.; Thorat, N. D.; Phadatare, M. R.; Sathish, C. I.; Dhawale, D. S.; Pawar, S. H. Polyvinyl alcohol functionalized cobalt ferrite nanoparticles for biomedical applications *Appl. Surf. Sci.* 2013, 264 (1) 598– 604

Radziun, E.; Dudkiewicz Wilczynska, J.; Ksiazek, I.; Nowak, K.; Anuszewska, E. L.; Kunicki, A.; Olszyna, A.; Zabkowski, T. Assessment of the cytotoxicity of aluminium oxide nanoparticles on selected mammalian cells *Toxicol. In Vitro* 2011, 25 (8) 1694– 1700

Gonzalez, L.; Thomassen, L. C.; Plas, G.; Rabolli, V.; Napierska, D.; Decordier, I.; Roelants, M.; Hoet, P. H.; Kirschhock, C. E.; Martens, J. A.; Lison, D.; Kirsch-Volders, M. Exploring the aneugenic and clastogenic potential in the nanosize range: A549 human lung carcinoma cells and amorphous monodisperse silica nanoparticles as models *Nanotoxicology* 2010, 4 (4) 382– 395

Jeng, H. A.; Swanson, J. Toxicity of metal oxide nanoparticles in mammalian cells *J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng.* 2006, 41 (12) 2699– 2711

- Li, C. H.; Shen, C. C.; Cheng, Y. W.; Huang, S. H.; Wu, C. C.; Kao, C. C.; Liao, J. W.; Kang, J. J. Organ biodistribution, clearance, and genotoxicity of orally administered zinc oxide nanoparticles in mice *Nanotoxicology* 2012, 6 (7) 746– 756
- Uboldi, C.; Giudetti, G.; Broggi, F.; Gilliland, D.; Ponti, J.; Rossi, F. Amorphous silica nanoparticles do not induce cytotoxicity, cell transformation or genotoxicity in Balb/3T3 mouse fibroblasts *Mutat. Res.* 2012, 745 (1–2) 11– 20
- Lin, W.; Huang, Y. W.; Zhou, X. D.; Ma, Y. In vitro toxicity of silica nanoparticles in human lung cancer cells *Toxicol. Appl. Pharmacol.* 2006, 217 (3) 252– 259
- Chusuei, C. C.; Wu, C. H.; Mallavarapu, S.; Hou, F. Y.; Hsu, C. M.; Winiarz, J. G.; Aronstam, R. S.; Huang, Y. W. Cytotoxicity in the age of nano: the role of fourth period transition metal oxide nanoparticle physicochemical properties *Chem. Biol. Interact.* 2013, 206 (2) 319– 326
- Song, L.; Connolly, M.; Fernandez-Cruz, M. L.; Vijver, M. G.; Fernandez, M.; Conde, E.; de Snoo, G. R.; Peijnenburg, W. J.; Navas, J. M. Species-specific toxicity of copper nanoparticles among mammalian and piscine cell lines *Nanotoxicology* 2014, 8 (4) 383– 393
- Horev-Azaria, L.; Baldi, G.; Beno, D.; Bonacchi, D.; Golla-Schindler, U.; Kirkpatrick, J. C.; Kolle, S.; Landsiedel, R.; Maimon, O.; Marche, P. N.; Ponti, J.; Romano, R.; Rossi, F.; Sommer, D.; Uboldi, C.; Unger, R. E.; Villiers, C.; Korenstein, R. Predictive toxicology of cobalt ferrite nanoparticles: comparative in-vitro study of different cellular models using methods of knowledge discovery from data *Part. Fibre Toxicol* 2013, 10, 32
- Carlson, C.; Hussain, S. M.; Schrand, A. M.; Braydich-Stolle, L. K.; Hess, K. L.; Jones, R. L.; Schlager, J. J. Unique cellular interaction of silver nanoparticles: size-dependent generation of reactive oxygen species *J. Phys. Chem. B* 2008, 112 (43) 13608– 13619
- Ahamed, M.; Ali, D.; Alhadlaq, H. A.; Akhtar, M. J. Nickel oxide nanoparticles exert cytotoxicity via oxidative stress and induce apoptotic response in human liver cells (HepG2) *Chemosphere* 2013, 93 (10) 2514– 2522
- Motskin, M.; Wright, D. M.; Muller, K.; Kyle, N.; Gard, T. G.; Porter, A. E.; Skepper, J. N. Hydroxyapatite nano and microparticles: correlation of particle properties with cytotoxicity and biostability *Biomaterials* 2009, 30 (19) 3307– 3317
- Selvaraj, V.; Bodapati, S.; Murray, E.; Rice, K. M.; Winston, N.; Shokuhfar, T.; Zhao, Y.; Blough, E. Cytotoxicity and genotoxicity caused by yttrium oxide nanoparticles in HEK293 cells *Int. J. Nanomedicine* 2014, 9 (1) 1379– 1391
- Saqib, Q.; Al-Khedhairy, A. A.; Ahmad, J.; Siddiqui, M. A.; Dwivedi, S.; Khan, S. T.; Musarrat, J. Zinc ferrite nanoparticles activate IL-1b, NFkB1, CCL21 and NOS2 signaling to induce mitochondrial dependent intrinsic apoptotic pathway in WISH cells *Toxicol. Appl. Pharmacol.* 2013, 273 (2) 289– 297
- Ahamed, M.; Siddiqui, M. A.; Akhtar, M. J.; Ahmad, I.; Pant, A. B.; Alhadlaq, H. A. Genotoxic potential of copper oxide nanoparticles in human lung epithelial cells *Biochem. Biophys. Res. Commun.* 2010, 396 (2) 578– 583

- Coradeghini, R.; Gioria, S.; Garcia, C. P.; Nativo, P.; Franchini, F.; Gilliland, D.; Ponti, J.; Rossi, F. Size-dependent toxicity and cell interaction mechanisms of gold nanoparticles on mouse fibroblasts *Toxicol. Lett.* 2013, 217 (3) 205– 216
- Grosse, S.; Evje, L.; Syversen, T. Silver nanoparticle-induced cytotoxicity in rat brain endothelial cell culture *Toxicol. In Vitro* 2013, 27 (1) 305– 313
- Fraga, S.; Faria, H.; Soares, M. E.; Duarte, J. A.; Soares, L.; Pereira, E.; Costa-Pereira, C.; Teixeira, J. P.; de Lourdes Bastos, M.; Carmo, H. Influence of the surface coating on the cytotoxicity, genotoxicity and uptake of gold nanoparticles in human HepG2 cells *J. Appl. Toxicol.* 2013, 33 (10) 1111– 1119
- Xu, Z.; Liu, C.; Wei, J.; Sun, J. Effects of four types of hydroxyapatite nanoparticles with different nanocrystal morphologies and sizes on apoptosis in rat osteoblasts *J. Appl. Toxicol.* 2012, 32 (6) 429– 435
- Wang, Z.; Li, N.; Zhao, J.; White, J. C.; Qu, P.; Xing, B. CuO nanoparticle interaction with human epithelial cells: cellular uptake, location, export, and genotoxicity *Chem. Res. Toxicol.* 2012, 25 (7) 1512– 1521
- Ahamed, M. Toxic response of nickel nanoparticles in human lung epithelial A549 cells *Toxicol. In Vitro* 2011, 25 (4) 930– 936
- Sabbioni, E.; Fortaner, S.; Farina, M.; Del Torchio, R.; Olivato, I.; Petrarca, C.; Bernardini, G.; Mariani-Costantini, R.; Perconti, S.; Di Giampaolo, L.; Gornati, R.; Di Gioacchino, M. Cytotoxicity and morphological transforming potential of cobalt nanoparticles, microparticles and ions in Balb/3T3 mouse fibroblasts: an in vitro model *Nanotoxicology* 2014, 8 (4) 455– 464
- Bhattacharjee, S.; Rietjens, I. M.; Singh, M. P.; Atkins, T. M.; Purkait, T. K.; Xu, Z.; Regli, S.; Shukaliak, A.; Clark, R. J.; Mitchell, B. S.; Alink, G. M.; Marcelis, A. T.; Fink, M. J.; Veinot, J. G.; Kauzlarich, S. M.; Zuilhof, H. Cytotoxicity of surface-functionalized silicon and germanium nanoparticles: the dominant role of surface charges *Nanoscale* 2013, 5 (11) 4870– 4883
- Tarantola, M.; Pietuch, A.; Schneider, D.; Rother, J.; Sunnick, E.; Rosman, C.; Pierrat, S.; Sonnichsen, C.; Wegener, J.; Janshoff, A. Toxicity of gold-nanoparticles: synergistic effects of shape and surface functionalization on micromotility of epithelial cells *Nanotoxicology* 2011, 5 (2) 254– 268
- Chueh, P. J.; Liang, R. Y.; Lee, Y. H.; Zeng, Z. M.; Chuang, S. M. Differential cytotoxic effects of gold nanoparticles in different mammalian cell lines *J. Hazard. Mater.* 2014, 264, 303– 312
- Kasemets, K.; Ivask, A.; Dubourguier, H. C.; Kahru, A. Toxicity of nanoparticles of ZnO, CuO and TiO₂ to yeast *Saccharomyces cerevisiae* *Toxicol. In Vitro* 2009, 23 (6) 1116– 1122
- Heinlaan, M.; Ivask, A.; Blinova, I.; Dubourguier, H. C.; Kahru, A. Toxicity of nanosized and bulk ZnO, CuO and TiO₂ to bacteria *Vibrio fischeri* and crustaceans *Daphnia magna* and *Thamnocephalus platyurus* *Chemosphere* 2008, 71 (7) 1308– 1316
- Mortimer, M.; Kasemets, K.; Kahru, A. Toxicity of ZnO and CuO nanoparticles to ciliated

- protozoa *Tetrahymena thermophila* *Toxicology* 2010, 269 (2–3) 182– 189
- Hund-Rinke, K.; Simon, M. Ecotoxic effect of photocatalytic active nanoparticles (TiO₂) on algae and daphnids *Environ. Sci. Pollut. Res. Int.* 2006, 13 (4) 225– 232
- Sadiq, I. M.; Pakrashi, S.; Chandrasekaran, N.; Mukherjee, A. Studies on toxicity of aluminum oxide (Al₂O₃) nanoparticles to microalgae species: *Scenedesmus* sp. and *Chlorella* sp *J. Nanopart. Res.* 2011, 13 (8) 3287– 3299
- Gong, N.; Shao, K.; Feng, W.; Lin, Z.; Liang, C.; Sun, Y. Biototoxicity of nickel oxide nanoparticles and bio-remediation by microalgae *Chlorella vulgaris* *Chemosphere* 2011, 83 (4) 510– 516
- Griffitt, R. J.; Luo, J.; Gao, J.; Bonzongo, J. C.; Barber, D. S. Effects of particle composition and species on toxicity of metallic nanomaterials in aquatic organisms *Environ. Toxicol. Chem.* 2008, 27 (9) 1972– 1978
- Sadiq, I. M.; Dalai, S.; Chandrasekaran, N.; Mukherjee, A. Ecotoxicity study of titania (TiO₂) NPs on two microalgae species: *Scenedesmus* sp. and *Chlorella* sp *Ecotoxicol. Environ. Saf.* 2011, 74 (5) 1180– 1187
- Tuominen, M.; Schultz, E.; Sillanpää, M. Toxicity and stability of silver nanoparticles to the green alga *Pseudokirchneriella subcapitata* in boreal freshwater samples and growth media *Nanomater. Environ.* 2013, 1, 48– 57
- García, A.; Espinosa, R.; Delgado, L.; Casals, E.; González, E.; Puentes, V.; Barata, C.; Font, X.; Sánchez, A. Acute toxicity of cerium oxide, titanium oxide and iron oxide nanoparticles using standardized tests *Desalination* 2011, 269 (1–3) 136– 141
- Franklin, N. M.; Rogers, N. J.; Apte, S. C.; Batley, G. E.; Gadd, G. E.; Casey, P. S. Comparative toxicity of nanoparticulate ZnO, bulk ZnO, and ZnCl₂ to a freshwater microalga (*Pseudokirchneriella subcapitata*): The importance of particle solubility *Environ. Sci. Technol.* 2007, 41 (24) 8484– 8490
- Han, X.; Lai, L.; Tian, F.; Jiang, F. L.; Xiao, Q.; Li, Y.; Yu, Q.; Li, D.; Wang, J.; Zhang, Q.; Zhu, B.; Li, R.; Liu, Y. Toxicity of CdTe quantum dots on yeast *Saccharomyces cerevisiae* *Small* 2012, 8 (17) 2680– 2689
- Ma, Y.; Kuang, L.; He, X.; Bai, W.; Ding, Y.; Zhang, Z.; Zhao, Y.; Chai, Z. Effects of rare earth oxide nanoparticles on root elongation of plants *Chemosphere* 2010, 78 (3) 273– 279
- Lin, D.; Xing, B. Phytotoxicity of nanoparticles: inhibition of seed germination and root growth *Environ. Pollut.* 2007, 150 (2) 243– 250
- Ma, Y.; He, X.; Zhang, P.; Zhang, Z.; Guo, Z.; Tai, R.; Xu, Z.; Zhang, L.; Ding, Y.; Zhao, Y.; Chai, Z. Phytotoxicity and biotransformation of La₂O₃ nanoparticles in a terrestrial plant cucumber (*Cucumis sativus*) *Nanotoxicology* 2011, 5 (4) 743– 753
- Zhu, X.; Zhu, L.; Chen, Y.; Tian, S. Acute toxicities of six manufactured nanomaterial suspensions to *Daphnia magna* *J. Nanopart. Res.* 2009, 11 (1) 67– 75
- Zhu, X.; Chang, Y.; Chen, Y. Toxicity and bioaccumulation of TiO₂ nanoparticle aggregates in *Daphnia magna* *Chemosphere* 2010, 78 (3) 209– 215
- Pakrashi, S.; Dalai, S.; Humayun, A.; Chakravarty, S.; Chandrasekaran, N.; Mukherjee,

A. Ceriodaphnia dubia as a Potential Bio-Indicator for Assessing Acute Aluminum Oxide Nanoparticle Toxicity in Fresh Water Environment PLoS One 2013, 8 (9) e74003

Wang, H.; Wick, R. L.; Xing, B. Toxicity of nanoparticulate and bulk ZnO, Al₂O₃ and TiO₂ to the nematode Caenorhabditis elegans Environ. Pollut. 2009, 157 (4) 1171–1177

Asharani, P. V.; Lianwu, Y.; Gong, Z.; Valiyaveetil, S. Comparison of the toxicity of silver, gold and platinum nanoparticles in developing zebrafish embryos Nanotoxicology 2011, 5 (1) 43– 54

Bar-Ilan, O.; Albrecht, R. M.; Fako, V. E.; Furgeson, D. Y. Toxicity assessments of multisized gold and silver nanoparticles in zebrafish embryos Small 2009, 5 (16) 1897–1910

Zhu, X.; Zhu, L.; Duan, Z.; Qi, R.; Li, Y.; Lang, Y. Comparative toxicity of several metal oxide nanoparticle aqueous suspensions to Zebrafish (Danio rerio) early developmental stage J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng. 2008, 43 (3) 278–284

Zhang, W.; Lin, K.; Miao, Y.; Dong, Q.; Huang, C.; Wang, H.; Guo, M.; Cui, X. Toxicity assessment of zebrafish following exposure to CdTe QDs J. Hazard. Mater. 2012, 213–214, 413– 420

Silva, T.; Pokhrel, L. R.; Dubey, B.; Tolaymat, T. M.; Maier, K. J.; Liu, X. Particle size, surface charge and concentration dependent ecotoxicity of three organo-coated silver nanoparticles: comparison between general linear model-predicted and observed toxicity Sci. Total Environ. 2014, 468–469, 968– 976

Marsalek, B.; Jancula, D.; Marsalkova, E.; Mashlan, M.; Safarova, K.; Tucek, J.; Zboril, R. Multimodal action and selective toxicity of zerovalent iron nanoparticles against cyanobacteria Environ. Sci. Technol. 2012, 46 (4) 2316– 2323

Nations, S.; Wages, M.; Canas, J. E.; Maul, J.; Theodorakis, C.; Cobb, G. P. Acute effects of Fe(2)O(3), TiO(2), ZnO and CuO nanomaterials on Xenopus laevis Chemosphere 2011, 83 (8) 1053– 1061

Artells, E.; Issartel, J.; Auffan, M.; Borschneck, D.; Thill, A.; Tella, M.; Brousset, L.; Rose, J.; Bottero, J. Y.; Thiery, A. Exposure to cerium dioxide nanoparticles differently affect swimming performance and survival in two daphnid species PLoS One 2013, 8 (8) e71260

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cells and Organism
, Saccharomyces cerevisiae

Vibrio fischeri

Tetrahymena thermophila

Danio rerio (embryos)

Desmodesmus subspicatus

Chlorella sp.

Chlorella vulgaris

Scenedesmus sp.

Pseudokirchneriella subcapitata

Brassica napus

Cucumis sativus

Raphanus sativus

Lolium perenne

, QSAR, - TE_i (c_j)_{rf} -> binary (classification) variable expressing the toxic effect of the nanoparticle used as reference

- $\Delta\Delta V(\text{me})$ -> perturbation term that characterizes the variations in the molar volume between the new (output or final state) nanoparticle and the other used as reference, also depending on the measures of the toxic effects

- $\Delta\Delta E(\text{bt})$ -> perturbation term that describes the changes in the electronegativity between the new (output or final state) nanoparticle and the other used as reference, also depending on the biological targets

- $\Delta\Delta E(\text{dm})$ -> perturbation term that describes the variations in the electronegativity between the new (output or final state) nanoparticle and the other used as reference, also depending on the conditions under which the sizes of the nanoparticles were measured

- $\Delta\Delta P(\text{bt})$ -> perturbation term that accounts for the variations in the polarizability between the new (output or final state) nanoparticle and the other used as reference, also depending on the biological targets

- $\Delta\Delta P(\text{ns})$ -> perturbation term that characterizes the changes in the polarizability between the new (output or final state) nanoparticle and the other used as reference, also depending on the shapes of the nanoparticles

- $\Delta\Delta L(\text{ta})$ -> perturbation term that accounts for the changes in the size between the new (output or final state) nanoparticle and the other used as reference, also depending on the intervals of time during which the biological targets were exposed to the nanoparticles

- $\Delta G\mu_3(\text{Hyd})$ -> perturbation spectral moment of order 3, weighted by the hydrophobicity, and characterizing the differences between the chemical structure of the coating agent used in the new (output or final state) nanoparticle, and the coating agent used for the reference nanoparticle


- $\Delta G\mu_5(\text{PSA})$ -> perturbation spectral moment of order 5, weighted by the polar surface area, and describing the differences between the chemical structure of the coating agent used in the new (output or final state) nanoparticle, and the coating agent used for the reference nanoparticle

,LDA applied to the perturbation approach obtained equation.

by software STATISTICA 6.0

The final model will be a consensus prediction, after apply the desired NP to the 229 NPs used as reference NP.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells byMold2 and
	Printing Date:30/03/2017

1.QSAR identifier

1.1.QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells byMold2 and kNN

1.2.Other related models:

NA

1.3.Software coding the model:

NA

2.General information

2.1.Date of QMRF:

30/03/2017

2.2.QMRF author(s) and contact details:

LEITAT

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

G. Melagraki

A. Afantitis

melagraki@novamechanics.com

afantitis@novamechanics.com

2.6.Date of model development and/or publication:

2014

2.7.Reference(s) to main scientific papers and/or software package:

Melagraki, G., & Afantitis, A. (2014). Enalos InSilicoNano platform: An online decision support tool for the design and virtual screening of nanoparticles. RSC Advances, 4(92), 50713–50725.

<http://doi.org/10.1039/c4ra07756c>

2.8.Availability of information about the model:

No information available

2.9.Availability of another QMRF for exactly the same model:

No information available

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2.Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3.Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

kNN: k-Nearest Neighbour

by KNMINE software

4.3.Descriptors in the model:

- Geary topological structure autocorrelation length-7 weighted by atomic van der Waals volumes (D461)
- Geary topological structure autocorrelation length-5 weighted by atomic Sanderson electronegativities (D467)
- Number of total quaternary C-sp3 (D599)
- Number of group secondary amines (aliphatic) (D649)
- Number of group donor atoms for H-bonds (with N and O) (D712)
- Number of group CH3R and CH4 (D714)
- Number of group phenol or enol or carboxylOH(D753)
- Number of group Al2-NH (D758)
- Hydrophilic factor index (D775).; 9

4.4.Descriptor selection:

Mold2 software generates 777 descriptors, and number of them were removed as some of the descriptors do not have any discrimination power (no variation).

Correlation – based feature subset selection (CfsSubset) variable selection combined with BestFirst evaluator were chosen to evaluate the most critical parameters.

CfsSubset algorithm evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them.

BestFirst evaluator searches the space of attribute subsets by greedy hillclimbing augmented with a backtracking facility.

A forward search has been chosen for this work

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

89/9

Descriptor: Chemical ratio :9:80 ~1:9

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

The distance of a test compound to its nearest neighbour in the training set is compared to a predefined threshold (APD) and the prediction is considered unreliable when the distance is higher than that. APD was calculated based on the following formula:

$$APD = \langle d \rangle + Z \cdot \sigma$$

$\langle d \rangle$: average of Euclidean distance.

Sigma: the Standard Deviation of the average.

Z: is a constant set to 0.5.

The calculated threshold was: $APD_t = 2.153$

All compounds in the test set had values in the range of 0.019–1.06 except for one which slightly falls outside with a value of 2.29. The predictions for all compounds that fell inside the domain of applicability of the model can be considered reliable.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

89 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_mShape: NACoating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione

4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione

Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride

3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione

Dibenz(c,e)oxepin-5,7-dione

6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone

Lauric anhydride

1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid

5-Methyl-2-benzofuran-1,3-dione

4-Nitro-2-benzofuran-1,3-dione

1H-isochromene-1,3(4H)-dione

Dihydro-2H-pyran-2,6(3H)-dione

4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione

2H-3,1-benzoxazine-2,4(1H)-dione

1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione

4-Methyldihydro-2Hpyran-2,6(3H)-dione

4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione

2,5-Dioxotetrahydrofuran-3,4-diyl diacetate

4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione

Hexahydro-2-benzofuran-1,3-dione

5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.02,6]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.13,7]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.03,7]nonan-3-amine
 Tricyclo[3.3.1.13,7]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid

2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(=O)O
 Amino(4-chlorophenyl)acetic acid
NC(C(=O)O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfonyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

Among the 109 compounds originally included in the dataset 89 constituted the training set and 20 the test set. Only compounds included in the training set were used to develop the QNAR model whereas compounds included in the test set were not involved in the model development.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.848$

see "Scheme 1" to a summary of other statistics.

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

10 fold cross validation:

$R^2_{L100} = 0.74$

Y-randomization eliminate the possibility of chance correlation.

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20 Metal Oxide

List

$(Fe_2O_3)_n(Fe_3O_4)_m$

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine

4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid

2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2_{cvext} = 0.82$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is very crucial that the developed model does not remain within the developers' group but is widely disseminated to the community so that it could immediately serve as an important source of information as it was initially designed to be. The proposed model was made publicly available online through Enalos InSilicoNano platform. Enalos InSilicoNano platform is a webservice that can host several validated

and predictive models that can be utilized in the NPs design process

R^2 : correlation coefficient

R^2_{L100} : correlation coefficient for 10-fold cross-validation

KNIME: Konstanz Information Miner (software)

CfsSubset: Correlation - based feature subset selection

kNN: k nearest Neighbour

APD: a predefined threshold

9.2. Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - Geary topological structure autocorrelation length-7 weighted by atomic van der Waals volumes (D461)

- Geary topological structure autocorrelation length-5 weighted by atomic Sanderson electronegativities (D467)

- Number of total quaternary C-sp³ (D599)

- Number of group secondary amines (aliphatic) (D649)

- Number of group donor atoms for H-bonds (with N and O) (D712)

- Number of group CH₃R and CH₄ (D714)


- Number of group phenol or enol or carboxylOH (D753)

- Number of group Al₂-NH (D758)

- Hydrophilic factor index (D775), kNN: k-Nearest Neighbour

by KNMINE software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: The “liquid drop” approach applied to develop predictive classification
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles by Random Forest
(E. Coli case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Jerzy Leszczynski

jerzy@icnanotox.org

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Sizochenko, N., Rasulev, B., Gajewicz, A., Kuz'Min, V., Puzyn, T., & Leszczynski, J. (2014). From basic physics to mechanisms of toxicity: The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparti
<http://doi.org/10.1039/c4nr03487b>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/EC50)

3.3.Comment on endpoint:

Originally measured in vitro effective concentration EC50 toxicity data (mol L⁻¹) were expressed as logarithms of the inverse molar concentration (log(1/EC50)) response variables, which represents the cytotoxicity that reduces bacteria viability up to 50%

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RF: Random Forest

4.3.Descriptors in the model:

(% -> relative contribution of certain descriptor to toxicity)

- S1 – unbounded two-atomic fragments [Me]...[Me], which were encoded based on SiRMS-derived descriptors, encoding the distance where the potential reaches minimum at van der Waals interactions (7%)

- r_w –Wigner–Seitz radius (22%)

- ρ – mass density (2%)

- CPP – cation polarizing power (30%)

- S2 – SiRMS-derived electronegativity aligned descriptor of oxides molecules – in a sense of the acid–base property of oxides. This parameter increases with the number of oxygen atoms in a molecule (3%)

- S3 – tri-atomic fragments[Me]–[O]–[Me], which were encoded by SiRMS-derived descriptors, encoding electronegativity (29%);

- SV – proportion of surface molecules to molecules in volume (7%).

; 7

4.4.Descriptor selection:

At the initial preparatory step, a number of descriptors were generated. By SiRMS, LDM and MLB

Then, non-significant, constant descriptors and descriptors with high cross-correlation ($r > 0.90$) were eliminated (one of the two descriptors with cross-correlation).

Finally, the RF algorithm will select the most relevant descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

13/7

Descriptor: Chemical ratio :7:13 ~ 1:2

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal and metal oxide NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

13 Metal Oxide

List: Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): Size range : 15 - 150

Aggregation size range: 180 - 2029

Other info: Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

6.6.Pre-processing of data before modelling:

The splitting of the dataset to training and test sets (for both HaCaT cells and E. coli sets) was the same for both cases and fulfill two conditions:

- (1) metal oxides from each activity group should be presented in both training and test sets;
- (2) metal oxides presented in the test set should cover all types of oxides (MeO, Me₂O₃, MeO₂), similarly to the training set.

The splitting of data into training and test sets is displayed in publication's Table 1.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.93$

RMSE = 0.13

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

3 M Metal Oxide

ListAl₂O₃Bi₂O₃

CoO

Cr₂O₃Fe₂O₃In₂O₃La₂O₃

NiO

Sb₂O₃SiO₂SnO₂TiO₂V₂O₃WO₃Y₂O₃

ZnO

ZrO₂Shape: NACoating: NASize(nm): Size range : 15 - 150

Aggregation size range: 180 - 2029

Other properties:

Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.78$

RMSE = 0.32

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1. Comments:

The developed nano-QSAR models reveal the differences in the mechanisms of toxicity of metal oxide nanoparticles to bacteria and a human keratinocyte cell line, which belong to prokaryotic and eukaryotic systems, respectively.

There is not a test for robustness, but, a cleaning process of the high correlated descriptors were developed before the model was built.

Mechanistic interpretation of the different descriptors was explained and a comparison between both cases was presented.

NPs: Nanoparticles

RF: Random Forest

R²: Correlation coefficient

RMSE: Root-mean-square-error

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

9.2. Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ... Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, 6(3), 175–178.

and

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, (% -> relative contribution of certain descriptor to toxicity)

- S1 – unbounded two-atomic fragments [Me]...[Me], which were encoded based on SiRMS-derived descriptors, encoding the distance where the potential reaches minimum at van der Waals interactions (7%)

- r_w –Wigner–Seitz radius (22%)


- ρ – mass density (2%)

- CPP – cation polarizing power (30%)

- S2 – SiRMS-derived electronegativity aligned descriptor of oxides molecules – in a sense of the acid–base property of oxides. This parameter increases with the number of oxygen atoms in a molecule (3%)
- S3 – tri-atomic fragments[Me]–[O]–[Me], which were encoded by SiRMS-derived descriptors, encoding electronegativity (29%);
- SV – proportion of surface molecules to molecules in volume (7%).

,RF: Random Forest

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: The “liquid drop” approach applied to develop predictive classification
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles by Random Forest
(HaCaT case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Jerzy Leszczynski

jerzy@icnanotox.org

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Sizochenko, N., Rasulev, B., Gajewicz, A., Kuz'Min, V., Puzyn, T., & Leszczynski, J. (2014). From basic physics to mechanisms of toxicity: The “liquid drop” approach applied to develop predictive classification models for toxicity of metal oxide nanoparti
<http://doi.org/10.1039/c4nr03487b>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human keratinocyte cell line (HaCaT)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/EC50)

3.3.Comment on endpoint:

Originally measured in vitro effective concentration EC50 toxicity data (mol L⁻¹) were expressed as logarithms of the inverse molar concentration (log(1/EC50)) response variables, which represents the cytotoxicity that reduces bacteria viability up to 50%

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RF: Random Forest

4.3.Descriptors in the model:

(% -> relative contribution of certain descriptor to toxicity)

- S1 – unbounded two-atomic fragments [Me]...[Me], which were encoded based on SiRMS-derived descriptors, describing the distance where potential reaches minimum at van der Waals interactions (43%)

- r_w –Wigner–Seitz radius of oxide's molecule (24%)

- p – mass density (6%)

- CI – covalent index of the metal ion (10%)

- S2 – SiRMS-derived number of oxygen's atoms in a molecule, which was described by their electronegativity (15%)

- AP – aggregation parameter (2%); 6

4.4.Descriptor selection:

At the initial preparatory step, a number of descriptors were generated. By SiRMS, LDM and MLB

Then, non-significant, constant descriptors and descriptors with high cross-correlation ($r > 0.90$) were eliminated (one of the two descriptors with cross-correlation).

Finally, the RF algorithm will select the most relevant descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

14/6

Descriptor: Chemical ratio :6:14 ~ 1:2

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal and metal oxide NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

14 Metal Oxide

List: Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): Size range : 15 - 150

Aggregation size range: 180 - 2029

Other info: Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

6.6.Pre-processing of data before modelling:

The splitting of the dataset to training and test sets (for both HaCaT cells and E. coli sets) was the same for both cases and fulfill two conditions:

- (1) metal oxides from each activity group should be presented in both training and test sets;
- (2) metal oxides presented in the test set should cover all types of oxides (MeO, Me₂O₃, MeO₂), similarly to the training set.

The splitting of data into training and test sets is displayed in publication's Table 1.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.96$

RMSE = 0.10

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

3 M Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape:NA

Coating:NA

Size(nm): Size range : 15 - 150

Aggregation size range: 180 - 2029

Other properties:

Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.92$

RMSE = 0.12

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The developed nano-QSAR models reveal the differences in the mechanisms of toxicity of metal oxide nanoparticles to bacteria and a

human keratinocyte cell line, which belong to prokaryotic and eukaryotic systems, respectively.

There is not a test for robustness, but, a cleaning process of the high correlated descriptors were developed before the model was built.

Mechanistic interpretation of the different descriptors was explained and a comparison between both cases was presented.

NPs: Nanoparticles

RF: Random Forest

R²: Correlation coefficient

RMSE: Root-mean-square-error

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

9.2.Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ...

Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, 6(3), 175–178.

and

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Human keratinocyte cell line (HaCaT), QSAR, (% -> relative contribution of certain descriptor to toxicity)

- S1 – unbounded two-atomic fragments [Me]...[Me], which were encoded based on SiRMS-derived descriptors, describing the distance where potential reaches minimum at van der Waals interactions (43%)

- r_w –Wigner–Seitz radius of oxide's molecule (24%)


- ρ – mass density (6%)

- CI – covalent index of the metal ion (10%)

- S2 – SiRMS-derived number of oxygen's atoms in a molecule, which was described by their electronegativity (15%)

- AP – aggregation parameter (2%).,RF: Random Forest

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction of the Induction of apoptosis by metal oxide nanoparticles
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction of the Induction of apoptosis by metal oxide nanoparticles by MLR
(SMA -MLR- case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

dave.winkler@csiro.au

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software package:

Modelling Biological Activities of Nanoparticles

V. Chandana Epa, Frank R. Burden, Carlos Tassa, Ralph Weissleder, Stanley Shaw, and David A. Winkler

Nano Letters 2012 12 (11), 5808-5812

(SMA -MLR- case)

Winkler, D. A., Burden, F. R., Yan, B., Weissled

<http://doi.org/10.1021/nl303144k>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Vascular smooth muscle cells (human coronary artery)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as biological activity by Dose-response curve of smooth muscle cell apoptosis (SMA)

3.3.Comment on endpoint:

Biological activity was defined as the arithmetic mean of in vitro tests on four different cell lines,

- Endothelial cells (human aorta)
- Vascular smooth muscle cells (human coronary artery)
- Hepatocytes (human HepG2 cells)
- Murine RAW 264.7 leukemic monocyte/macrophage cell

Using four doses, and four different assays of cellular physiology.

The four assays measured

- (i) ATP content,
- (ii) Reducing equivalents,
- (iii) Caspase-mediated apoptosis,
- (iv) Mitochondrial membrane potential.

Only the apoptosis assay (smooth muscle cell apoptosis, SMA) exhibited dose-response relationship.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLREM: Multiple Linear Regression with Expectation Maximization

4.3.Descriptors in the model:

- I_Fe2O3: presence (1) or absence (0) of core material Fe2O3
 - I_dextran: presence (1) or absence (0) of coating material dextran
 - I_charge: presence (1) or absence (0) of surface charge
- ; 3

4.4.Descriptor selection:

Only the smooth muscle cell apoptosis assay generated statistical significant models. They initially investigated the dependence of the apoptosis response on the relaxivities (R1 and R2) and the zeta potential (available for 32 of the nanoparticles).

Developed models using three indicator variables for these three properties: nature of the nanoparticle core (+1 for Fe2O3 and 0 for Fe3O4); type of coating (+1 for dextran and 0 for other coatings); and nature of surface functionality (encoded as +1 (basic), -1 (acidic), or 0 (neutral))

Chosen those ones which represent better statistics of the building algorithms.

Optimal descriptor set that had been selected

by the MLREM protocol were applied for bot model building algorithms

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

26/3

Descriptor: Chemical ratio :3:26 ~1:9

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal and metal oxide NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

26 Metal

Metal Oxide

List: Fe₂O₃Fe₃O₄Shape: NACoating: Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488
 Cross-linked dextran :: Alexa Fluor 750
 Cross-linked dextran :: FITC, R-COOH
 Cross-linked dextran :: biotin
 Cross-linked dextran :: FITC, COOH
 Cross-linked dextran :: Cy3.5
 Cross-linked dextran :: Cy5.5, protamine
 Cross-linked dextran :: Cy5.5, tat
 Cross-linked dextran :: Cy5.5
 Cross-linked dextran :: Cy5
 Cross-linked dextran :: Cy7
 Cross-linked dextran :: FITC
 Cross-linked dextran :: FITC, Glutamic acid
 Cross-linked dextran :: glycine
 Cross-linked dextran :: rhodamine, protamine
 Cross-linked dextran :: FITC, succinimidyl iodoacetate
 Cross-linked dextran :: Tat peptide
 Cross-linked dextran :: VT680
 Cross-linked dextran :: VT680, protamine
 Dextran :: NA
 Sucrose :: NA
 PVA :: COOH
 PVA :: Ethylene diamine
 PVA :: Ethylene diamine, VT680
 PVA :: protamine, rhodamine
 PVA :: L-arg8-COOH
 PVA :: COOH
 PVA :: AminoSPARK™680
 PVA :: PEG Ethylene diamine, AminoSPARK™680
 PVA :: Ethylene diamine, AminoSPARK™680
 PVA, PEG :: AngioSPARK™680- IVM
 PVA :: 15-mer peptide
 PVA :: L-arg7-COOH
 PVA :: Ethylene diamine, VT750
 PVA, PEG :: Ethylene diamine, VT750
 PVA :: D-arg7-COOH
 PVA, PEG :: NA
 Arabino-galactan :: NA
 Carboxymethyldextran :: NA
 Amphiphilic polymer – PEG :: NH₂
 Amphiphilic polymer :: COOH
Size (nm): 20-74

Other info: Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

6.6.Pre-processing of data before modelling:

Only the smooth muscle cell apoptosis assay generated statistical significant models. They initially investigated the dependence of the apoptosis response on the relaxivities (R1 and R2) and the zeta potential (available for 32 of the nanoparticles).

Generated models either using all of the data in the model, or splitting the data into a training set of 26, and a test set of six nanoparticles using a k-means k-means clustering method

6.7.Statistics for goodness-of-fit:

$r^2_{\text{train}} = 0.81$

SEE = 3.6

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

6 MMetal

Metal Oxide

List

Fe₂O₃

Fe₃O₄

Shape:NA

Coating:Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750

Cross-linked dextran :: FITC, R-COOH

Cross-linked dextran :: biotin

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: Cy3.5

Cross-linked dextran :: Cy5.5, protamine

Cross-linked dextran :: Cy5.5, tat

Cross-linked dextran :: Cy5.5

Cross-linked dextran :: Cy5

Cross-linked dextran :: Cy7

Cross-linked dextran :: FITC

Cross-linked dextran :: FITC, Glutamic acid

Cross-linked dextran :: glycine

Cross-linked dextran :: rhodamine, protamine

Cross-linked dextran :: FITC, succinimidyl iodoacetate

Cross-linked dextran :: Tat peptide

Cross-linked dextran :: VT680

Cross-linked dextran :: VT680, protamine

Dextran :: NA

Sucrose :: NA

PVA :: COOH

PVA :: Ethylene diamine

PVA :: Ethylene diamine, VT680

PVA :: protamine, rhodamine

PVA :: L-arg8-COOH

PVA :: COOH

PVA :: AminoSPARK™680

PVA :: PEG Ethylene diamine, AminoSPARK™680

PVA :: Ethylene diamine, AminoSPARK™680

PVA, PEG :: AngioSPARK™680- IVM

PVA :: 15-mer peptide

PVA :: L-arg7-COOH

PVA :: Ethylene diamine, VT750

PVA, PEG :: Ethylene diamine, VT750

PVA :: D-arg7-COOH

PVA, PEG :: NA

Arabino-galactan :: NA

Carboxymethyldextran :: NA

Amphiphilic polymer – PEG :: NH₂

Amphiphilic polymer :: COOH

Size(nm): 20-74

Other properties:

Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2_{\text{test}} = 0.86$

SEP = 3.3

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

NA

SMA: smooth muscle cell apoptosis

MLREM: Multiple Linear Regression with expectation

maximization

r^2 : Correlation coefficient

SEE: Standard Error Estimation

SEP: standard error of external prediction

9.2.Bibliography:

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. Proceedings of the National Academy of Sciences of the United States of America, 105(21), 7387–7392. <http://doi.org/10.1073/pnas.0802878105>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Vascular smooth muscle cells (human coronary artery), QSAR, - I_Fe2O3: presence (1) or absence (0) of core material Fe2O3

- I_dextran: presence (1) or absence (0) of coating material dextran

- I_charge: presence (1) or absence (0) of surface charge

,MLREM: Multiple Linear Regression with Expectation Maximization

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction of the Induction of apoptosis by metal oxide nanoparticles
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction of the Induction of apoptosis by metal oxide nanoparticles by BRANN
(SMA -BRANN- case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

dave.winkler@csiro.au

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software package:

Modelling Biological Activities of Nanoparticles

V. Chandana Epa, Frank R. Burden, Carlos Tassa, Ralph Weissleder, Stanley Shaw, and David A. Winkler

Nano Letters 2012 12 (11), 5808-5812

(SMA -BRANN- case)

Winkler, D. A., Burden, F. R., Yan, B., Weissl

<http://doi.org/10.1021/nl303144k>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Vascular smooth muscle cells (human coronary artery)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as biological activity by Dose-response curve of smooth muscle cell apoptosis (SMA)

3.3.Comment on endpoint:

Biological activity was defined as the arithmetic mean of in vitro tests on four different cell lines,

- Endothelial cells (human aorta)
- Vascular smooth muscle cells (human coronary artery)
- Hepatocytes (human HepG2 cells)
- Murine RAW 264.7 leukemic monocyte/macrophage cell

Using four doses, and four different assays of cellular physiology.

The four assays measured

- (i) ATP content,
- (ii) Reducing equivalents,
- (iii) Caspase-mediated apoptosis,
- (iv) Mitochondrial membrane potential.

Only the apoptosis assay (smooth muscle cell apoptosis, SMA) exhibited dose-response relationship.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

BRANN: Bayesian Regularization Artificial Neural Network

4.3.Descriptors in the model:

- I_Fe2O3: presence (1) or absence (0) of core material Fe2O3
- I_dextran: presence (1) or absence (0) of coating material dextran
- I_charge: presence (1) or absence (0) of surface charge; 3

4.4.Descriptor selection:

Only the smooth muscle cell apoptosis assay generated statistical significant models. They initially investigated the dependence of the apoptosis response on the relaxivities (R1 and R2) and the zeta potential (available for 32 of the nanoparticles).

Developed models using three indicator variables for these three properties: nature of the nanoparticle core (+1 for Fe2O3 and 0 for Fe3O4); type of coating (+1 for dextran and 0 for other coatings); and nature of surface functionality (encoded as +1 (basic), -1 (acidic), or 0 (neutral))

Chosen those ones which represent better statistics of the building algorithms.

Optimal descriptor set that had been selected by the MLREM protocol were applied for both model building algorithms

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

26/3

Descriptor: Chemical ratio :3:26 ~1:9

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal and metal oxide NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

26 Metal

Metal Oxide

List: Fe₂O₃

Fe₃O₄

Shape: NA

Coating: Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750
 Cross-linked dextran :: FITC, R-COOH
 Cross-linked dextran :: biotin
 Cross-linked dextran :: FITC, COOH
 Cross-linked dextran :: Cy3.5
 Cross-linked dextran :: Cy5.5, protamine
 Cross-linked dextran :: Cy5.5, tat
 Cross-linked dextran :: Cy5.5
 Cross-linked dextran :: Cy5
 Cross-linked dextran :: Cy7
 Cross-linked dextran :: FITC
 Cross-linked dextran :: FITC, Glutamic acid
 Cross-linked dextran :: glycine
 Cross-linked dextran :: rhodamine, protamine
 Cross-linked dextran :: FITC, succinimidyl iodoacetate
 Cross-linked dextran :: Tat peptide
 Cross-linked dextran :: VT680
 Cross-linked dextran :: VT680, protamine
 Dextran :: NA
 Sucrose :: NA
 PVA :: COOH
 PVA :: Ethylene diamine
 PVA :: Ethylene diamine, VT680
 PVA :: protamine, rhodamine
 PVA :: L-arg8-COOH
 PVA :: COOH
 PVA :: AminoSPARK™680
 PVA :: PEG Ethylene diamine, AminoSPARK™680
 PVA :: Ethylene diamine, AminoSPARK™680
 PVA, PEG :: AngioSPARK™680- IVM
 PVA :: 15-mer peptide
 PVA :: L-arg7-COOH
 PVA :: Ethylene diamine, VT750
 PVA, PEG :: Ethylene diamine, VT750
 PVA :: D-arg7-COOH
 PVA, PEG :: NA
 Arabino-galactan :: NA
 Carboxymethyldextran :: NA
 Amphiphilic polymer – PEG :: NH2
 Amphiphilic polymer :: COOH
Size (nm): 20-74
Other info: Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

6.6.Pre-processing of data before modelling:

Only the smooth muscle cell apoptosis assay generated statistical significant models. They initially investigated the dependence of the apoptosis response on the relaxivities (R1 and R2) and the zeta potential (available for 32 of the nanoparticles).

Generated models either using all of the data in the model, or splitting the data into a training set of 26, and a test set of six nanoparticles using a k-means k-means clustering method

6.7.Statistics for goodness-of-fit:

$r^2_{\text{train}} = 0.80$

SEE = 2.8

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

6 MMetal

Metal Oxide

List

Fe2O3

Fe3O4

Shape:NA

Coating:Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750

Cross-linked dextran :: FITC, R-COOH

Cross-linked dextran :: biotin

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: Cy3.5

Cross-linked dextran :: Cy5.5, protamine

Cross-linked dextran :: Cy5.5, tat

Cross-linked dextran :: Cy5.5

Cross-linked dextran :: Cy5

Cross-linked dextran :: Cy7

Cross-linked dextran :: FITC

Cross-linked dextran :: FITC, Glutamic acid

Cross-linked dextran :: glycine

Cross-linked dextran :: rhodamine, protamine

Cross-linked dextran :: FITC, succinimidyl iodoacetate

Cross-linked dextran :: Tat peptide

Cross-linked dextran :: VT680

Cross-linked dextran :: VT680, protamine

Dextran :: NA

Sucrose :: NA

PVA :: COOH

PVA :: Ethylene diamine

PVA :: Ethylene diamine, VT680

PVA :: protamine, rhodamine

PVA :: L-arg8-COOH

PVA :: COOH

PVA :: AminoSPARK™680

PVA :: PEG Ethylene diamine, AminoSPARK™680

PVA :: Ethylene diamine, AminoSPARK™680

PVA, PEG :: AngioSPARK™680- IVM

PVA :: 15-mer peptide

PVA :: L-arg7-COOH

PVA :: Ethylene diamine, VT750

PVA, PEG :: Ethylene diamine, VT750

PVA :: D-arg7-COOH

PVA, PEG :: NA

Arabino-galactan :: NA

Carboxymethyldextran :: NA

Amphiphilic polymer – PEG :: NH₂

Amphiphilic polymer :: COOH

Size(nm): 20-74

Other properties:

Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2_{\text{test}} = 0.90$

SEP = 2.9

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

NA

SMA: smooth muscle cell apoptosis

BRANN: Bayesian regularized Neural Network

r^2 : Correlation coefficient

SEE: Standard Error Estimation

SEP: standard error of external prediction

9.2.Bibliography:

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. Proceedings of the National Academy of Sciences of the United States of America, 105(21), 7387–7392. <http://doi.org/10.1073/pnas.0802878105>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Cell, Vascular smooth muscle cells (human coronary artery), QSAR, - I_Fe2O3: presence (1) or absence (0) of core material Fe2O3

- I_dextran: presence (1) or absence (0) of coating material dextran

- I_charge: presence (1) or absence (0) of surface charge, BRANN: Bayesian Regularization Artificial Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cellular uptake by CLIO NPs in several types of cells by BRANN
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cellular uptake by CLIO NPs in several types of cells by BRANN
(HUVEC case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

dave.winkler@csiro.au

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software package:

Modelling Biological Activities of Nanoparticles

V. Chandana Epa, Frank R. Burden, Carlos Tassa, Ralph Weissleder, Stanley Shaw, and David A. Winkler

Nano Letters 2012 12 (11), 5808-5812

(HUVEC case)

Winkler, D. A., Burden, F. R., Yan, B., Weissleder,

<http://doi.org/10.1021/nl303144k>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human umbilical vein endothelial cells (HUVEC)

3.2.Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3.Comment on endpoint:

Measured by well fluorescein isothiocyanate (FITC) concentrations. Experimental data were used as its log10 transform.

In assessing whether both of the above assays contain useful biological information, the z-scored data were used where $Z^{NP} = (\mu_{NP} - \mu_{PBS}) / \sigma_{PBS}$, where μ and σ are the mean and standard deviation of assay replicates, respectively, and the NP and PBS subscripts represent assays in the presence of PBS buffer controls. Assays where most of the dose-response curves fell within a z-score of ± 2 were considered to demonstrate negligible effect

Of the five cell lines tested:

- Human umbilical vein endothelial cells (HUVEC)
- Primary resting human macrophages (RestMph)
- Granulocyte macrophage colony stimulating factor-stimulated human macrophages (GMCSF_Mph)
- Human macrophage-like cell line (U937)
- Human pancreatic ductal adenocarcinoma cells (PaCa2)

Only the pancreatic cancer (PaCa2) and human umbilical vein endothelial cell (HUVEC) lines showed significant variation in uptake of surface modified nanoparticles.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

BRANN: Bayesian Regularization Artificial Neural Network

4.3.Descriptors in the model:

- nR10: Ring complexity (constitutional); Number of 10 membered rings (e.g. naphthalene)
- ASP: Molecular shape (geometrical); Molecular asphericity
- DISPM: Molecular shape (geometrical); d COMMA2 value/weighted by atomic masses. The moment of inertia divided by the mass.
- QZZm: Molecular shape (geometrical); Qzz COMMA2 value/weighted by atomic masses. The second order mass moment divided by the mass.
- nRCONHR: Hydrogen bonding capacity (functional group counts); Number of secondary amides (aliphatic)
- nArOCON: Hydrogen bonding capacity (functional group counts); Number of (thio-) carbamates (aromatic)
- C-005: Hydrophobicity (atom-centred fragments); CH3X:

; 7

4.4.Descriptor selection:

Initial descriptors (691) generated by software DRAGON 5.0 for Windows (only 2D descriptors)

Chosen those ones which represent better statistics of the building algorithms.

After the evaluation of a first set of models, and in order to enforce the interpretability of the applied descriptors, final model was initialized from 124 chemically interpretable descriptors.

Optimal descriptor set that had been selected by the MLREM protocol were applied for the model building algorithm

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

87/7

Descriptor: Chemical ratio :7:87 ~ 1:12

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of Cross-Linked Iron Oxide (CLIO-NH₂) NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

87 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione

4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione

Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride

3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione

Dibenz(c,e)oxepin-5,7-dione

6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone

Lauric anhydride

1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid

5-Methyl-2-benzofuran-1,3-dione

4-Nitro-2-benzofuran-1,3-dione

1H-isochromene-1,3(4H)-dione

Dihydro-2H-pyran-2,6(3H)-dione

4,4'-Ethane-1,2-diylmorpholine-2,6-dione

2H-3,1-benzoxazine-2,4(1H)-dione

1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione

4-Methyldihydro-2Hpyran-2,6(3H)-dione

4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione

2,5-Dioxotetrahydrofuran-3,4-diyl diacetate

4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione

Hexahydro-2-benzofuran-1,3-dione

5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione

Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid

2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

Experimental data were used as its log₁₀ transform, and clustering was used to divide it into a test set of 21 molecules (i.e., 20% of the data) and a training set of 87 molecules (80% of the data)

6.7.Statistics for goodness-of-fit:

$r^2_{\text{train}} = 0.55$

SEE = 0.38

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

21 Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic dianhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.02,6]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine

2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione

2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2_{\text{test}} = 0.72$

SEP = 0.30

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

NA

BRANN: Bayesian regularized Neural Network

r^2 : Correlation coefficient

SEE: Standard Error Estimation

SEP: standard error of external prediction

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-

specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Human umbilical vein endothelial cells (HUVEC), QSAR, - nR10: Ring complexity (constitutional); Number of 10 membered rings (e.g. naphthalene)

- ASP: Molecular shape (geometrical); Molecular asphericity

- DISPm: Molecular shape (geometrical); d COMMA2 value/weighted by atomic masses. The moment of inertia divided by the mass.

- QZZm: Molecular shape (geometrical); Qzz COMMA2 value/weighted by atomic masses. The second order mass moment divided by the mass.


- nRCONHR: Hydrogen bonding capacity (functional group counts); Number of secondary amides (aliphatic)

- nArOCON: Hydrogen bonding capacity (functional group counts); Number of (thio-) carbamates (aromatic)

- C-005: Hydrophobicity (atom-centred fragments); CH3X:

,BRANN: Bayesian Regularization Artificial Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cellular uptake by CLIO NPs in several types of cells by MLREM
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cellular uptake by CLIO NPs in several types of cells by MLREM
(PaCa2 case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

dave.winkler@csiro.au

2.6. Date of model development and/or publication:

2012

2.7. Reference(s) to main scientific papers and/or software package:

Modelling Biological Activities of Nanoparticles

V. Chandana Epa, Frank R. Burden, Carlos Tassa, Ralph Weissleder, Stanley Shaw, and David A. Winkler

Nano Letters 2012 12 (11), 5808-5812

(PaCa2 case)

Winkler, D. A., Burden, F. R., Yan, B., Weissleder,

<http://doi.org/10.1021/nl303144k>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2.Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3.Comment on endpoint:

Measured by well fluorescein isothiocyanate (FITC) concentrations. Experimental data were used as its log10 transform.

In assessing whether both of the above assays contain useful biological information, the z-scored data were used where $Z^{NP} = (\mu_{NP} - \mu_{PBS}) / \sigma_{PBS}$, where μ and σ are the mean and standard deviation of assay replicates, respectively, and the NP and PBS subscripts represent assays in the presence of PBS buffer controls. Assays where most of the dose-response curves fell within a z-score of ± 2 were considered to demonstrate negligible effect

Of the five cell lines tested:

- Human umbilical vein endothelial cells (HUVEC)
- Primary resting human macrophages (RestMph)
- Granulocyte macrophage colony stimulating factor-stimulated human macrophages (GMCSF_Mph)
- Human macrophage-like cell line (U937)
- Human pancreatic ductal adenocarcinoma cells (PaCa2)

Only the pancreatic cancer (PaCa2) and human umbilical vein endothelial cell (HUVEC) lines showed significant variation in uptake of surface modified nanoparticles.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLREM: Multiple Linear Regression with Expectation Maximization

4.3.Descriptors in the model:

- nCIR: Branching? (constitutional); Number of circuits
- nN: Hydrogen bonding capacity (constitutional); Number of N atoms
- SPAM: (geometrical); Average molecular span R
- QYYp: Molecular shape and polarizability (geometrical); Qyy COMMA2 value/weighted by atomic polarizabilities. The second order mass moment divided by the mass.
- nCs: Branching (functional group counts); Number of total secondary C(sp3)
- nArOH: Hydrogen bonding capacity (functional group counts); Number of aromatic hydroxyls
- H-053: (atom-centred fragments); H attached to CO(sp3) with 2X attached to next C
- O-058: Hydrogen bonding capacity (atom-centred fragments) =O; 8

4.4.Descriptor selection:

Initial descriptors (691) generated by software DRAGON 5.0 for Windows (only 2D descriptors)

Chosen those ones which represent better statistics of the building algorithms.

After the evaluation of a first set of models, and in order to enforce the interpretability of the applied descriptors, final model was initialized from 124 chemically interpretable descriptors.

Optimal descriptor set was obtained within the model building algorithm

4.5.Algorithm and descriptor generation:

No information available

4.6. Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

87/8

Descriptor: Chemical ratio :8:87 ~ 1:11

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of Cross-Linked Iron Oxide (CLIO-NH₂) NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

87 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride
 4 3,3-Dimethyldihydrofuran-2,5-dione
 Furan-2,5-dione
 3-Methylfuran-2,5-dione
 7 3,4-Dimethylfuran-2,5-dione
 Hexanoic anhydride
 3-Methyldihydrofuran-2,5-dione
 5,5'-Carbonylbis(2-benzofuran-1,3-dione)
 5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride

5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid

NCCCCC(N)C(O)=O

Amino(4-chlorophenyl)acetic acid

NC(C(O)=O)c1ccc(Cl)cc1

2-Aminopropanoic acid

2-Amino-5-carbamimidamidopentanoic acid

2-Aminobutanedioic acid

2,5-Diamino-5-oxopentanoic acid

2-Aminopentanedioic acid

2-Amino-3-(1Himidazol-4-yl)propanoic acid

2-Amino-4-(methylsulfanyl)butanoic acid

2-Amino-3-phenylpropanoic acid

Dihydrofuran-2,5-dione

Acetic anhydride

3-Methylidenedihydrofuran-2,5-dione

1,4-Dioxane-2,6-dione

2-Benzofuran-1,3-dione

(2,5-Dioxotetrahydrofuran-3-yl)acetic acid

4,7-Difluoro-2-benzofuran-1,3-dione

{Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

Experimental data were used as its log₁₀ transform, and clustering was used to divide it into a test set of 21 molecules (i.e., 20% of the data) and a training set of 87 molecules (80% of the data)

6.7.Statistics for goodness-of-fit:

$r^2_{\text{train}} = 0.64$

SEE = 0.26

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

21 Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape:NA

Coating:Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine

Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2_{\text{test}} = 0.62$

SEP = 0.32

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

NA

MLREM: Multiple Linear Regression with Expectation Maximization

r^2 : Correlation coefficient

SEE: Standard Error Estimation

SEP: standard error of external prediction

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - nCIR: Branching? (constitutional); Number of circuits

- nN: Hydrogen bonding capacity (constitutional); Number of N atoms
- SPAM: (geometrical); Average molecular span R
- QYYp: Molecular shape and polarizability (geometrical); Qyy COMMA2 value/weighted by atomic polarizabilities. The second order mass moment divided by the mass.
- nCs: Branching (functional group counts); Number of total secondary C(sp3)
- nArOH: Hydrogen bonding capacity (functional group counts); Number of aromatic hydroxyls
- H-053: (atom-centred fragments); H attached to CO(sp3) with 2X attached to next C
- O-058: Hydrogen bonding capacity (atom-centred fragments) =O, MLREM: Multiple Linear Regression with Expectation Maximization

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Ensemble Learning for predicting biological activity of diverse
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Ensemble Learning for predicting biological activity of diverse nanomaterials
(cell apoptosis -Regression- case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunwar P. Singh

kpsingh_52@yahoo.com

kunwarpsingh@gmail.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modelling for predicting biological activity of diverse nanomaterials. RSC Advances, 4(26), 13215–13230.

(cell apoptosis -Regression- case)

<http://doi.org/10.1039/c4ra01274g>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)

Murine RAW 264.7 leukemic monocyte/macrophage cell

3.2.Endpoint:

In vitro - Cytotoxicity - measured as biological activity by Dose-response curve of smooth muscle cell apoptosis (SMA)

3.3.Comment on endpoint:

Biological activity was defined as the arithmetic mean of in vitro tests on four different cell lines, Using four doses, and four different assays of cellular physiology :

- (i) ATP content,
- (ii) Reducing equivalents,
- (iii) Caspase-mediated apoptosis,
- (iv) Mitochondrial membrane potential.

Only the apoptosis assay (smooth muscle cell apoptosis, SMA) exhibited dose-response relationship.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

In all cases two Ensemble learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a regression model.

They are inherently

4.3.Descriptors in the model:

- Size of the NP
- R1: spin-lattice Relaxivity
- R2: spin-spin Relaxivity
- Zeta potential, surface charge. (ZP); 4

4.4.Descriptor selection:

Descriptors were analyzed for the existence of a constant or near constant values and the descriptors with low variation were excluded from the original pool of descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/4

Descriptor: Chemical ratio :4:31 ~ 1:8

5.Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

The ranges of individual descriptors used for the model building.

According to this method, a NP with descriptor values within the range of those of the training set NPs is considered as being inside the AD of the model.

For specific details see (in the publication) Table 5 to check the range values of the descriptors for each case study.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Metal

Metal Oxide

List: Fe₂O₃

Fe₃O₄

Shape: NA

Coating: Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750
 Cross-linked dextran :: FITC, R-COOH
 Cross-linked dextran :: biotin
 Cross-linked dextran :: FITC, COOH
 Cross-linked dextran :: Cy3.5
 Cross-linked dextran :: Cy5.5, protamine
 Cross-linked dextran :: Cy5.5, tat
 Cross-linked dextran :: Cy5.5
 Cross-linked dextran :: Cy5
 Cross-linked dextran :: Cy7
 Cross-linked dextran :: FITC
 Cross-linked dextran :: FITC, Glutamic acid
 Cross-linked dextran :: glycine
 Cross-linked dextran :: rhodamine, protamine
 Cross-linked dextran :: FITC, succinimidyl iodoacetate
 Cross-linked dextran :: Tat peptide
 Cross-linked dextran :: VT680
 Cross-linked dextran :: VT680, protamine
 Dextran :: NA
 Sucrose :: NA
 PVA :: COOH
 PVA :: Ethylene diamine
 PVA :: Ethylene diamine, VT680
 PVA :: protamine, rhodamine
 PVA :: L-arg8-COOH
 PVA :: COOH
 PVA :: AminoSPARK™680
 PVA :: PEG Ethylene diamine, AminoSPARK™680
 PVA :: Ethylene diamine, AminoSPARK™680
 PVA, PEG :: AngioSPARK™680- IVM
 PVA :: 15-mer peptide
 PVA :: L-arg7-COOH
 PVA :: Ethylene diamine, VT750
 PVA, PEG :: Ethylene diamine, VT750
 PVA :: D-arg7-COOH
 PVA, PEG :: NA
 Arabino-galactan :: NA
 Carboxymethyldextran :: NA
 Amphiphilic polymer – PEG :: NH2
 Amphiphilic polymer :: COOH
Size (nm): 20-74
Other info: Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

6.6.Pre-processing of data before modelling:

For internal validation, a V-fold cross validation (CV) method was adopted

For external validation, a separate validation (test) sub-set of the data was used which was kept out during the training process. For classification and regression modelling, data were split into training (80%) and test (20%) subsets using random distribution approach.

6.7.Statistics for goodness-of-fit:

- DTB:

$$R^2 = 0.950$$

- DTF:

$$R^2 = 0.868$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

- DTB:

$$CV-RMSE = 3.83$$

- DTF:

$$CV-RMSE = 4.03$$

5-fold CV validation.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20% of data MMetal

Metal Oxide

List

Fe₂O₃

Fe₃O₄

Shape:NA

Coating:Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750

Cross-linked dextran :: FITC, R-COOH

Cross-linked dextran :: biotin

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: Cy3.5

Cross-linked dextran :: Cy5.5, protamine

Cross-linked dextran :: Cy5.5, tat

Cross-linked dextran :: Cy5.5

Cross-linked dextran :: Cy5

Cross-linked dextran :: Cy7

Cross-linked dextran :: FITC

Cross-linked dextran :: FITC, Glutamic acid

Cross-linked dextran :: glycine

Cross-linked dextran :: rhodamine, protamine

Cross-linked dextran :: FITC, succinimidyl iodoacetate

Cross-linked dextran :: Tat peptide

Cross-linked dextran :: VT680

Cross-linked dextran :: VT680, protamine

Dextran :: NA

Sucrose :: NA

PVA :: COOH

PVA :: Ethylene diamine

PVA :: Ethylene diamine, VT680

PVA :: protamine, rhodamine

PVA :: L-arg8-COOH

PVA :: COOH

PVA :: AminoSPARK™680

PVA :: PEG Ethylene diamine, AminoSPARK™680

PVA :: Ethylene diamine, AminoSPARK™680

PVA, PEG :: AngioSPARK™680- IVM

PVA :: 15-mer peptide

PVA :: L-arg7-COOH

PVA :: Ethylene diamine, VT750

PVA, PEG :: Ethylene diamine, VT750

PVA :: D-arg7-COOH

PVA, PEG :: NA

Arabino-galactan :: NA

Carboxymethyldextran :: NA

Amphiphilic polymer – PEG :: NH₂

Amphiphilic polymer :: COOH

Size(nm): 20-74

Other properties:

Nanoparticle size and zeta potential were measured by using a Zetasizer 1000

(Malvern Instruments); relaxivities were determined by using a Bruker

Minispec MQ20 NMR

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

- DTB:

$$R^2 = 0.906$$

- DTF:

$$R^2 = 0.817$$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The statistical parameters given might bring confusion, since for example Q_F^2 should be equal to R^2 if we look at the equation. Then, those values were not used in the discussion, and also the values to be compared with the previous studies were those which come from whole complete data

EL: Ensemble learning

DTF: Decision tree forest

DTB: Decision treeboost

CV-RMSE: Cross validation root-mean-square error

R²: correlation coefficient

NP: Nanoparticle

AD: Applicability domain

9.2. Bibliography:

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21), 7387–7392. <http://doi.org/10.1073/pnas.0802878105>

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)

Murine RAW 264.7 leukemic monocyte/macrophage cell, QSAR, - Size of the NP

- R1: spin-lattice Relaxivity

- R2: spin-spin Relaxivity

- Zeta potential, surface charge. (ZP), In all cases two Ensemble learning (EL) based nano-QSAR models were applied:


Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a regression model.

They are inherently

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Ensemble Learning for predicting biological activity of diverse
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Ensemble Learning for predicting biological activity of diverse nanomaterials
(cell apoptosis -Classification- case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunwar P. Singh

kpsingh_52@yahoo.com

kunwarpsingh@gmail.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modelling for predicting biological activity of diverse nanomaterials. RSC Advances, 4(26), 13215–13230.

(cell apoptosis -Classification- case)

<http://doi.org/10.1039/c4ra01274g>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)

Murine RAW 264.7 leukemic monocyte/macrophage cell

3.2.Endpoint:

In vitro - Cytotoxicity - measured as biological activity by Dose-response curve of smooth muscle cell apoptosis (SMA)

3.3.Comment on endpoint:

Biological activity was defined as the arithmetic mean of in vitro tests on four different cell lines, Using four doses, and four different assays of cellular physiology :

- (i) ATP content,
- (ii) Reducing equivalents,
- (iii) Caspase-mediated apoptosis,
- (iv) Mitochondrial membrane potential.

Only the apoptosis assay (smooth muscle cell apoptosis, SMA) exhibited dose-response relationship.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

In all cases two Ensemble learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a classification model.

They are inherent

4.3.Descriptors in the model:

- Size of the NP
- R1: spin-lattice Relaxivity
- R2: spin-spin Relaxivity
- Zeta potential, surface charge. (ZP); 4

4.4.Descriptor selection:

Descriptors were analyzed for the existence of a constant or near constant values and the descriptors with low variation were excluded from the original pool of descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/4

Descriptor: Chemical ratio :4:31 ~ 1:8

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The ranges of individual descriptors used for the model building.

According to this method, a NP with descriptor values within the range of those of the training set NPs is considered as being inside the AD of the model.

For specific details see (in the publication) Table 5 to check the range values of the descriptors for each case study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal

Metal Oxide

List: Fe₂O₃

Fe₃O₄

Shape: NA

Coating: Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750
 Cross-linked dextran :: FITC, R-COOH
 Cross-linked dextran :: biotin
 Cross-linked dextran :: FITC, COOH
 Cross-linked dextran :: Cy3.5
 Cross-linked dextran :: Cy5.5, protamine
 Cross-linked dextran :: Cy5.5, tat
 Cross-linked dextran :: Cy5.5
 Cross-linked dextran :: Cy5
 Cross-linked dextran :: Cy7
 Cross-linked dextran :: FITC
 Cross-linked dextran :: FITC, Glutamic acid
 Cross-linked dextran :: glycine
 Cross-linked dextran :: rhodamine, protamine
 Cross-linked dextran :: FITC, succinimidyl iodoacetate
 Cross-linked dextran :: Tat peptide
 Cross-linked dextran :: VT680
 Cross-linked dextran :: VT680, protamine
 Dextran :: NA
 Sucrose :: NA
 PVA :: COOH
 PVA :: Ethylene diamine
 PVA :: Ethylene diamine, VT680
 PVA :: protamine, rhodamine
 PVA :: L-arg8-COOH
 PVA :: COOH
 PVA :: AminoSPARK™680
 PVA :: PEG Ethylene diamine, AminoSPARK™680
 PVA :: Ethylene diamine, AminoSPARK™680
 PVA, PEG :: AngioSPARK™680- IVM
 PVA :: 15-mer peptide
 PVA :: L-arg7-COOH
 PVA :: Ethylene diamine, VT750
 PVA, PEG :: Ethylene diamine, VT750
 PVA :: D-arg7-COOH
 PVA, PEG :: NA
 Arabino-galactan :: NA
 Carboxymethyldextran :: NA
 Amphiphilic polymer – PEG :: NH2
 Amphiphilic polymer :: COOH
Size (nm): 20-74
Other info: Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

6.6.Pre-processing of data before modelling:

For internal validation, a V-fold cross validation (CV) method was adopted

For external validation, a separate validation (test) sub-set of the data was used which was kept out during the training process. For classification and regression modelling, data were split into training (80%) and test (20%) subsets using random distribution approach.

6.7.Statistics for goodness-of-fit:

- DTB:

Sensitivity = 100.00 %

Specificity = 100.00 %

Accuracy = 100.00 %

MCC = 1.00

- DTF:

Sensitivity = 100.00 %

Specificity = 94.74 %

Accuracy = 97.30 %

MCC

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

5-fold CV validation.

It was considered that the bag and bootstrapping applied within the model provide an independent test without requiring a separate data.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20% of data MMetal

Metal Oxide

ListFe₂O₃Fe₃O₄Shape:NACoating:Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750

Cross-linked dextran :: FITC, R-COOH

Cross-linked dextran :: biotin

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: Cy3.5

Cross-linked dextran :: Cy5.5, protamine

Cross-linked dextran :: Cy5.5, tat

Cross-linked dextran :: Cy5.5

Cross-linked dextran :: Cy5

Cross-linked dextran :: Cy7

Cross-linked dextran :: FITC

Cross-linked dextran :: FITC, Glutamic acid

Cross-linked dextran :: glycine

Cross-linked dextran :: rhodamine, protamine

Cross-linked dextran :: FITC, succinimidyl iodoacetate

Cross-linked dextran :: Tat peptide

Cross-linked dextran :: VT680

Cross-linked dextran :: VT680, protamine

Dextran :: NA

Sucrose :: NA

PVA :: COOH

PVA :: Ethylene diamine

PVA :: Ethylene diamine, VT680

PVA :: protamine, rhodamine

PVA :: L-arg8-COOH

PVA :: COOH

PVA :: AminoSPARK™680

PVA :: PEG Ethylene diamine, AminoSPARK™680

PVA :: Ethylene diamine, AminoSPARK™680

PVA, PEG :: AngioSPARK™680- IVM

PVA :: 15-mer peptide

PVA :: L-arg7-COOH

PVA :: Ethylene diamine, VT750

PVA, PEG :: Ethylene diamine, VT750

PVA :: D-arg7-COOH

PVA, PEG :: NA

Arabino-galactan :: NA

Carboxymethyldextran :: NA

Amphiphilic polymer – PEG :: NH2

Amphiphilic polymer :: COOH

Size(nm): 20-74

Other properties:

Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

- DTB:

Sensitivity = 100.00 %

Specificity = 66.67 %

Accuracy = 71.43 %

MCC = 0.47

- DTF:

Sensitivity = 100.00 %

Specificity = 66.67 %

Accuracy = 71.43 %

MCC

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The statistical parameters given might bring confusion, since for example Q_F1^2 should be equal to R^2 if we look at the equation. Then, those values were not used in the discussion, and also the values to be compared with the previous studies were those which come from whole complete data

EL: Ensemble learning

DTF: Decision tree forest

DTB: Decision treeboost

CV-RMSE: Cross validation root-mean-square error

R^2 : correlation coefficient

NP: Nanoparticle

AD: Applicability domain

9.2. Bibliography:

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21), 7387–7392. <http://doi.org/10.1073/pnas.0802878105>

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)

Murine RAW 264.7 leukemic monocyte/macrophage cell, QSAR, - Size of the NP

- R1: spin-lattice Relaxivity

- R2: spin-spin Relaxivity

- Zeta potential, surface charge. (ZP), In all cases two Ensemble learning (EL) based nano-QSAR models were applied:


Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a classification model.

They are inherent

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Ensemble Learning for predicting biological activity of diverse
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Ensemble Learning for predicting biological activity of diverse nanomaterials
(PaCa2 -Regression- case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunwar P. Singh

kpsingh_52@yahoo.com

kunwarpsingh@gmail.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modelling for predicting biological activity of diverse nanomaterials. RSC Advances, 4(26), 13215–13230.

(PaCa2 -Regression- case)

<http://doi.org/10.1039/c4ra01274g>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3.Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

In all cases two Ensemble learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a regression model.

4.3.Descriptors in the model:

- Chi valance path descriptor of order 4 (VP-4)
- Chi valance path cluster of order 6 (VPC-6)
- Ionization potential (IP)
- Number of rotatable bonds (nRotB)
- Number of hydrogen acceptors (nHBAcc); 5

4.4.Descriptor selection:

Topological, electronic, geometrical and constitutional descriptors obtained by Chemistry Development Kit (CDK v 1.0.3)

Calculated descriptors were analyzed for the existence of a constant or near constant values and the descriptors with low variation were excluded from the original pool of descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/5

Descriptor: Chemical ratio :5:109 ~ 1 :22

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The ranges of individual descriptors used for the model building.

According to this method, a NP with descriptor values within the range of those of the training set NPs is considered as being inside the AD of the model.

For specific details see (in the publication) Table 5 to check the range values of the descriptors for each case study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine

4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid

2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

For internal validation, a V-fold cross validation (CV) method was adopted

For external validation, a separate validation (test) sub-set of the data was used which was kept out during the training process. For classification and regression modelling, data were split into training (80%) and test (20%) subsets using random distribution approach.

6.7.Statistics for goodness-of-fit:

- DTB:

$$R^2 = 0.947$$

- DTF:

$$R^2 = 0.942$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

- DTB:

$$CV-RMSE = 0.31$$

- DTF:

$$CV-RMSE = 0.31$$

5-fold CV validation.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20% of data MMetal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape:NA

Coating:Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine

Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

- DTB:

$$R^2 = 0.905$$

- DTF:

$$R^2 = 0.889$$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The statistical parameters given might bring confusion, since for example Q_F^2 should be equal to R^2 if we look at the equation. Then, those values were not used in the discussion, and also the values to be compared with the previous studies were those which come from whole complete data

EL: Ensemble learning

DTF: Decision tree forest

DTB: Decision treeboost

CV-RMSE: Cross validation root-mean-square error

R²: correlation coefficient

NP: Nanoparticle

AD: Applicability domain

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - Chi valance path descriptor of order 4 (VP-4)

- Chi valance path cluster of order 6 (VPC-6)

- Ionization potential (IP)

- Number of rotatable bonds (nRotB)


- Number of hydrogen acceptors (nHBAcc), In all cases two Ensemble learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a regression model.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Ensemble Learning for predicting biological activity of diverse
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Ensemble Learning for predicting biological activity of diverse nanomaterials
(PaCa2 -Classification- case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunwar P. Singh

kpsingh_52@yahoo.com

kunwarpsingh@gmail.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modelling for predicting biological activity of diverse nanomaterials. RSC Advances, 4(26), 13215–13230.

(PaCa2 -Classification- case)

<http://doi.org/10.1039/c4ra01274g>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3.Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

In all cases two Ensemble learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a classification model.

4.3.Descriptors in the model:

- Chi valance path descriptor of order 4 (VP-4)
- Chi valance path cluster of order 6 (VPC-6)
- Ionization potential (IP)
- Number of rotatable bonds (nRotB)
- Number of hydrogen acceptors (nHBAcc); 5

4.4.Descriptor selection:

Topological, electronic, geometrical and constitutional descriptors obtained by Chemistry Development Kit (CDK v 1.0.3)

Calculated descriptors were analyzed for the existence of a constant or near constant values and the descriptors with low variation were excluded from the original pool of descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/5

Descriptor: Chemical ratio :5:109 ~ 1 :22

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The ranges of individual descriptors used for the model building.

According to this method, a NP with descriptor values within the range of those of the training set NPs is considered as being inside the AD of the model.

For specific details see (in the publication) Table 5 to check the range values of the descriptors for each case study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine

4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid

2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

For internal validation, a V-fold cross validation (CV) method was adopted

For external validation, a separate validation (test) sub-set of the data was used which was kept out during the training process. For classification and regression modelling, data were split into training (80%) and test (20%) subsets using random distribution approach.

6.7.Statistics for goodness-of-fit:

- DTB:

Sensitivity = 100.00 %

Specificity = 97.44 %

Accuracy = 98.78 %

MCC = 0.98

- DTF:

Sensitivity = 100.00 %

Specificity = 100.00 %

Accuracy = 100.00 %

MCC

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

5-fold CV validation.

It was considered that the bag and bootstrapping applied within the model provide an independent test without requiring a separate data.

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20% of data Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic dianhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine

2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione

1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

- DTB:

Sensitivity = 88.24 %

Specificity = 100.00 %

Accuracy = 92.59 %

MCC = 0.86

- DTF:

Sensitivity = 87.50 %

Specificity = 90.91 %

Accuracy = 88.89 %

MCC

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1. Comments:

The statistical parameters given might bring confusion, since for example Q_{F1}^2 should be equal to R^2 if we look at the equation. Then, those values were not used in the discussion, and also the values to be compared with the previous studies were those which come from whole complete data

EL: Ensemble learning

DTF: Decision tree forest

DTB: Decision treeboost

CV-RMSE: Cross validation root-mean-square error

R^2 : correlation coefficient

NP: Nanoparticle

AD: Applicability domain

9.2. Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - Chi valance path descriptor of order 4 (VP-4)

- Chi valance path cluster of order 6 (VPC-6)

- Ionization potential (IP)

- Number of rotatable bonds (nRotB)


- Number of hydrogen acceptors (nHBAcc), In all cases two Ensemble learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a classification model.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Ensemble Learning for predicting biological activity of diverse
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Ensemble Learning for predicting biological activity of diverse nanomaterials
(E. Coli case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunwar P. Singh

kpsingh_52@yahoo.com

kunwarpsingh@gmail.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modelling for predicting biological activity of diverse nanomaterials. RSC Advances, 4(26), 13215–13230.

(E. Coli case)

<http://doi.org/10.1039/c4ra01274g>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as $\log(1/EC_{50})$

3.3.Comment on endpoint:

Determined the cytotoxicity of the metal oxide nanoparticles in terms of EC_{50} (concentration which cytotoxicity reduces bacteria viability up to 50%) based on the curve fitting least squares procedure. For experimental testing protocol see: Supplementary information (Section 1) from Puzyn et al. 2011 (already reported in this table).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

In all cases two Ensemble learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a regression model.

4.3.Descriptors in the model:

- Oxygen percent (OP)
- Molar reactivity (MolRef)
- Polar surface area (PSA); 3

4.4.Descriptor selection:

Topological, electronic, geometrical and constitutional descriptors obtained by Chemspider

Calculated descriptors were analyzed for the existence of a constant or near constant values and the descriptors with low variation were excluded from the original pool of descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/3

Descriptor: Chemical ratio :3:17 ~1:6

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The ranges of individual descriptors used for the model building.

According to this method, a NP with descriptor values within the range of those of the training set NPs is considered as being inside the AD of the model.

For specific details see (in the publication) Table 5 to check the range values of the descriptors for each case study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO
Cr₂O₃
La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of publication's supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package

6.6.Pre-processing of data before modelling:

For internal validation, a V-fold cross validation (CV) method was adopted

For external validation, a separate validation (test) sub-set of the data was used which was kept out during the training process. For classification and regression modelling, data were split into training (80%) and test (20%) subsets using random distribution approach.

6.7.Statistics for goodness-of-fit:

- DTB:

$$R^2 = 0.974$$

- DTF:

$$R^2 = 0.911$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

- DTB:

$$CV\text{-}RMSE = 0.16$$

- DTF:

$$CV\text{-}RMSE = 0.29$$

5-fold CV validation.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable
 INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA
 NP size: Yes
 NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20% of data Metal Oxide

List

ZnO
 CuO
 Al₂O₃
 Fe₂O₃
 SnO₂
 TiO₂
 V₂O₃
 Y₂O₃
 Bi₂O₃
 In₂O₃
 Sb₂O₃
 SiO₂
 ZrO₂
 CoO
 NiO
 Cr₂O₃
 La₂O₃

Shape: NA

Coating: NA

Size(nm): 15-90

Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of publication's supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

- DTB:

$$R^2 = 0.936$$

- DTF:

$$R^2 = 0.894$$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The statistical parameters given might bring confusion, since for example Q_{F1}^2 should be equal to R^2 if we look at the equation. Then, those values were not used in the discussion, and also the values to be compared with the previous studies were those which come from whole complete data

EL: Ensemble learning

DTF: Decision tree forest

DTB: Decision treeboost

CV-RMSE: Cross validation root-mean-square error

R^2 : correlation coefficient

NP: Nanoparticle

AD: Applicability domain

9.2.Bibliography:

(already reported in this table) Puzyn et al., 2011:

Experimental data plus previous published work:

Hu, X., Cook, S., Wang, P. & Hwang, H. M. In vitro evaluation of cytotoxicity of engineered metal oxide nanoparticles. Sci. Total Environ. 407, 3070–3072 (2009).

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - Oxygen percent (OP)

- Molar reactivity (MolRef)


- Polar surface area (PSA), In all cases two Ensemble learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a regression model.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Ensemble Learning for predicting biological activity of diverse
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Ensemble Learning for predicting biological activity of diverse nanomaterials
(Toxicity prediction of decorated nanotubes)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunwar P. Singh

kpsingh_52@yahoo.com

kunwarpsingh@gmail.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modelling for predicting biological activity of diverse nanomaterials. RSC Advances, 4(26), 13215–13230.

(mitochondrial dehydrogenases' activity of macrophages case)

<http://doi.org/10.1039/c4ra01274g>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

THP-1 (human monocytic cell line) differentiate into Macrophages

3.2. Endpoint:

In vitro - Cytotoxicity - measured as cellular viability by determining the mitochondrial dehydrogenases' activity

3.3.Comment on endpoint:

Cellular viability was determined on the basis of the activity of the mitochondrial dehydrogenases in the presence and absence of the nanotube–decorator complexes and expressed on a 0 to 100 scale (100 most “toxic”)

To evaluate the acute cytotoxicity (Cell) of the DNC library in macrophages, Zhou et al., used the WST-1 assay. Cellular viability was measured by determining the mitochondrial dehydrogenases' activity. The immune response was measured by treating macrophages with DNC for 24 h in a solution of lipopolysaccharide

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

In all cases two Ensemble Learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a regression model.

4.3.Descriptors in the model:

- The third Kier and Hall kappa molecular shape indices (Kier 3)
- Molecular distance edge between all secondary carbons (MDEC-22)
- Simple path descriptor of order 5 (SP-5),
- Weighted holistic invariant molecular descriptor (WTunity)
- X log P constitutional
- Moment of inertia along y/z-axis (MOMI-YZ); 6

4.4.Descriptor selection:

Topological, electronic, geometrical and constitutional descriptors obtained by Chemistry Development Kit (CDK v 1.0.3)

Calculated descriptors were analyzed for the existence of a constant or near constant values and the descriptors with low variation were excluded from the original pool of descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/6

Descriptor: Chemical ratio :6:29 ~ 1:5

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

The ranges of individual descriptors used for the model building.

According to this method, a NP with descriptor values within the range of those of the training set NPs is considered as being inside the AD of the model.

For specific details see (in the publication) Table 5 to check the range values of the descriptors for each case study.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Carbon-based

List: Decorator-nanotube complexes

Shape: Fiber

Coating: CC(NC(=O)OC3C1C=CC=CC1C2C=CC=CC23)C(=O)Nc4ccccc4

CCOC(=O)c4ccc(NC(=O)C(C)NC(=O)OC3C1C=CC=CC1C2C=CC=CC23)cc4

CC(N)C(=O)Nc1ccccc1

CC(N)C(=O)NCc1ccccc1

CCCCNC(=O)C(C)NC(=O)c1cccc(N(=O)=O)c1

```

CCCCNC(=O)C(C)NC(=O)c1ccc(Cl)cc1
CCCCNC(=O)C(C)NS(=O)(=O)c1cccc(N(=O)=O)c1
CCCCN(CCCC)C(=O)C(C)NC(=O)c1ccc(Cl)cc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1cccc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1ccc(C)cc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1cccc(N(=O)=O)c1
CC(NC(=O)c1cccc1)C(=O)NC2CCCCC2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)NC2CCCCC2
CCCCN(CCCC)C(=O)C(C)NC(=O)c1ccc(Cl)cc1
CC(NS(=O)(=O)c1cccc1)C(=O)NC2CCCCC2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)NC2CCCCC2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)Nc2cccc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2cccc2
CC(NS(=O)(=O)c1cccc1)C(=O)Nc2cccc2
Cc2ccc(S(=O)(=O)NC(C)C(=O)Nc1cccc1)cc2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)Nc2cccc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2cccc2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)Nc2cccc2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)N2CCCC2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)N2CCCC2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)N2CCCC2
CCOC(=O)c2ccc(NC(=O)C(C)NC(=O)c1ccc(Cl)cc1)cc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2cccc(C(F)(F)F)c2
CC(NS(=O)(=O)c1cccc1)C(=O)Nc2cccc(C(F)(F)F)c2

```

Size (nm): Diameter: 40 ± 10

Length: 250 ± 120

Other info: Coating compounds were reported by SMILES notation.

The number of walls of the nanotube will have little, if any, impact on the conformational behaviour of the surface attached decorator groups. The diameter(s) of the carbon nanotubes were not reported by Zhou, then two different diameters of 1 nm or 1.3 nm of diameter each 6.5 nm in length were defined.

The nanotube–decorator complex was geometry optimized using the molecular dynamics simulation (MDS) package GROMACS (version 4.5.2 for Linux) with the ffgmx force field (a derivative of the GROMOS87 force field).

6.6.Pre-processing of data before modelling:

For internal validation, a V-fold cross validation (CV) method was adopted

For external validation, a separate validation (test) sub-set of the data was used which was kept out during the training process. For classification and regression modelling, data were split into training (80%) and test (20%) subsets using random distribution approach.

6.7.Statistics for goodness-of-fit:

- DTB:

$$R^2 = 0.931$$

- DTF:

$$R^2 = 0.929$$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

- DTB:

CV-RMSE = 5.25

- DTF:

CV-RMSE = 4.85

5-fold CV validation.

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20% of data MCarbon-based

List

Decorator-nanotube complexes

Shape: Fiber

Coating: CC(NC(=O)OC3C1C=CC=CC1C2C=CC=CC23)C(=O)Nc4ccccc4

CCOC(=O)c4ccc(NC(=O)C(C)NC(=O)OC3C1C=CC=CC1C2C=CC=CC23)cc4

CC(N)C(=O)Nc1ccccc1

CC(N)C(=O)NCc1ccccc1

CCCCNC(=O)C(C)NC(=O)c1cccc(N(=O)=O)c1

CCCCNC(=O)C(C)NC(=O)c1ccc(Cl)cc1

```

CCCCNC(=O)C(C)NS(=O)(=O)c1cccc(N(=O)=O)c1
CCCCN(CCCC)C(=O)C(C)NC(=O)c1ccc(Cl)cc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1ccccc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1ccc(C)cc1
CCCCN(CCCC)C(=O)C(C)NS(=O)(=O)c1cccc(N(=O)=O)c1
CC(NC(=O)c1ccccc1)C(=O)NC2CCCCC2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)NC2CCCCC2
CCCCN(CCCC)C(=O)C(C)NC(=O)c1ccc(Cl)cc1
CC(NS(=O)(=O)c1ccccc1)C(=O)NC2CCCCC2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)NC2CCCCC2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)Nc2ccccc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2ccccc2
CC(NS(=O)(=O)c1ccccc1)C(=O)Nc2ccccc2
Cc2ccc(S(=O)(=O)NC(C)C(=O)Nc1ccccc1)cc2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)Nc2ccccc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2ccccc2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)Nc2ccccc2
CC(NC(=O)c1cccc(N(=O)=O)c1)C(=O)N2CCCC2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)N2CCCC2
CC(NS(=O)(=O)c1cccc(N(=O)=O)c1)C(=O)N2CCCC2
CCOC(=O)c2ccc(NC(=O)C(C)NC(=O)c1ccc(Cl)cc1)cc2
CC(NC(=O)c1ccc(Cl)cc1)C(=O)Nc2cccc(C(F)(F)F)c2
CC(NS(=O)(=O)c1ccccc1)C(=O)Nc2cccc(C(F)(F)F)c2

```

Size(nm): Diameter: 40 ± 10

Length: 250 ± 120

Other properties:

Coating compounds were reported by SMILES notation.

The number of walls of the nanotube will have little, if any, impact on the conformational behaviour of the surface attached decorator groups. The diameter(s) of the carbon nanotubes were not reported by Zhou, then two different diameters of 1 nm or 1.3 nm of diameter each 6.5 nm in length were defined.

The nanotube–decorator complex was geometry optimized using the molecular dynamics simulation (MDS) package GROMACS (version 4.5.2 for Linux) with the ffgmx force field (a derivative of the GROMOS87 force field).

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

- DTB:

$$R^2 = 0.971$$

- DTF:

$$R^2 = 0.927$$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The statistical parameters given might bring confusion, since for example Q_{F1}^2 should be equal to R^2 if we look at the equation. Then, those values were not used in the discussion, and also the values to be compared with the previous studies were those which come from whole complete data

EL: Ensemble learning

DTF: Decision tree forest

DTB: Decision treeboost

CV-RMSE: Cross validation root-mean-square error

R^2 : correlation coefficient

NP: Nanoparticle

AD: Applicability domain

9.2. Bibliography:

Zhou, H., Mu, Q., Gao, N., Liu, A., Xing, Y., Gao, S., ... Yan, B. (2008). A nano-combinatorial library strategy for the discovery of nanotubes with reduced protein-binding, cytotoxicity, and immune response. *Nano Letters*, 8(3), 859–865.
<http://doi.org/10.1021/nl0730155>

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, THP-1 (human monocytic cell line) differentiate into Macrophages, QSAR, - The third Kier and Hall kappa molecular shape indices (Kier 3)


- Molecular distance edge between all secondary carbons (MDEC-22)
- Simple path descriptor of order 5 (SP-5),
- Weighted holistic invariant molecular descriptor (WTunity)
- X log P constitutional
- Moment of inertia along y/z-axis (MOMI-YZ), In all cases two Ensemble Learning (EL) based nano-QSAR models were applied:

Decision Tree Forest (DTF) - implementing bagging

Decision Tree Boost (DTB) - implementing boosting

Both models were applied to obtain a regression model.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells by MLR plus
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells by MLR plus PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunal Roy

kunalroy_in@yahoo.com

kroy@pharma.jdvu.ac.in

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Kar, S., Gajewicz, A., Puzyn, T., & Roy, K. (2014). Nano-quantitative structure-activity relationship modelling using easily computable and interpretable descriptors for uptake of magnetofluorescent engineered nanoparticles in pancreatic cancer cells. Tox

<http://doi.org/10.1016/j.tiv.2013.12.018>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as Percentage of damaged cells by Propidium Iodide uptake of

BEAS-2B

3.3.Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Stepwise-MLR (Multiple Linear Regression) followed by PLS

employed software:

STATISTICA 7.0

SPSS 9.0

MINITAB 14

SIMPCA-P 10.0

4.3.Descriptors in the model:

- (1 - Atype - N - 66): refers to the hydrophobicity of the N atom in primary aliphatic amine (A1-NH2) fragment.
- (Atype - N - 67): refers to the hydrophobicity of the N atom in a secondary aliphatic amine (A12-NH) fragment.
- (0.600 - $\Sigma\beta'$): indicates the measure of electronic features of the molecule relative to molecular size.
- (Jurs - RPCS): the relative positive charge surface area. It is the solvent-accessible surface area of most positively charged atom divided by the relative positive charge.
- Wap: defined as all-path Wiener index which is the sum of the lengths of the shortest paths between all pairs of vertices in the chemical graph.
- nRNO2: the number of aliphatic nitro groups whose; 6

4.4.Descriptor selection:

The chemical structures were drawn in MarvinSketch 5.10.0 software.

The pool of descriptors were generated by different software:

Cerius 2 v4.10

Dragon 6

PaDEL-Descriptor v2.11

GFA applied to find out the most occurring descriptors. Within the building model, the final descriptors were screened and classified by the Variable Importance Projection (VIP)

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

89/6

Descriptor: Chemical ratio :6:89 ~ 1:15

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Two different approaches to assess AD:

- leverage approach

$$h^* = 0.236$$

- DModX

$$D\text{-crit} = 2.621$$

Identified an outlier. The 95% of the test set can be reliably predicted.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

89 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_mShape: NACoating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride
 4 3,3-Dimethyldihydrofuran-2,5-dione
 Furan-2,5-dione
 3-Methylfuran-2,5-dione
 7 3,4-Dimethylfuran-2,5-dione
 Hexanoic anhydride
 3-Methyldihydrofuran-2,5-dione
 5,5'-Carbonylbis(2-benzofuran-1,3-dione)
 5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride

5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid

NCCCCC(N)C(O)=O

Amino(4-chlorophenyl)acetic acid

NC(C(O)=O)c1ccc(Cl)cc1

2-Aminopropanoic acid

2-Amino-5-carbamimidamidopentanoic acid

2-Aminobutanedioic acid

2,5-Diamino-5-oxopentanoic acid

2-Aminopentanedioic acid

2-Amino-3-(1Himidazol-4-yl)propanoic acid

2-Amino-4-(methylsulfanyl)butanoic acid

2-Amino-3-phenylpropanoic acid

Dihydrofuran-2,5-dione

Acetic anhydride

3-Methylidenedihydrofuran-2,5-dione

1,4-Dioxane-2,6-dione

2-Benzofuran-1,3-dione

(2,5-Dioxotetrahydrofuran-3-yl)acetic acid

4,7-Difluoro-2-benzofuran-1,3-dione

{Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

Entire dataset was divided randomly into training and test sets keeping the highest and lowest toxic compounds in the training set. The training set and the test set contain 89 and 20 compounds respectively.

A PCA score plot in publication's supplementary material (Figure S2) shows that each test set compound is near to at least one training set compound.

For validation purposes leave-one-out (LOO) and leave-many-out (LMO) (10% and 25%) were applied to the training set.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.806$

SEE = 0.2

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

$Q^2_{\text{Leave-one-out}} = 0.758$

$Q^2_{\text{Leave-10\%-out}} = 0.634$

$Q^2_{\text{Leave-25\%-out}} = 0.648$

100 round of Y-randomization:

(c) $R^2_p = 0.806$ No chance correlation considered.

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20 Metal Oxide

List

$(\text{Fe}_2\text{O}_3)_n(\text{Fe}_3\text{O}_4)_m$

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.02,6]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.02,6]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione

Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid

2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$Q^2_{F1} = R^2_{Pred} = 0.879$

SEP = 0.12

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the

majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models.

Is in consonance with the guidelines of OECD WSAR model Development.

There is a Mechanistic Interpretation

PCA: Principal Component Analysis

GFA: Genetic Function Approximation

MLR: Multiple Linear Regression

PLS: Partial Least of Squares

R^2 : correlation coefficient

$Q^2_{F1} = R^2_{pred}$: correlation coefficient of external validation

SEE: standard error of

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - (1 - Atype - N - 66): refers to the hydrophobicity of the N atom in primary aliphatic amine (Al-NH₂) fragment.

- (Atype - N - 67): refers to the hydrophobicity of the N atom in a secondary aliphatic amine (Al₂-NH) fragment.

- (0.600 - $\Sigma\beta'$): indicates the measure of electronic features of the molecule relative to molecular size.

- (Jurs - RPCS): the relative positive charge surface area. It is the solvent-accessible surface area of most positively charged atom divided by the relative positive charge.

- Wap: defined as all-path Wiener index which is the sum of the lengths of the shortest paths between all pairs of vertices in the chemical graph.

- nRNO₂: the number of aliphatic nitro groups whose, Stepwise-MLR (Multiple Linear Regression) followed by PLS

employed software:


STATISTICA 7.0

SPSS 9.0

MINITAB 14

SIMPCA-P 10.0

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictive model of Mutagenicity of fullerene C60 by SMILES-based
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictive model of Mutagenicity of fullerene C60 by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software

package:

Toropov, A. A., & Toropova, A. P. (2014). Optimal descriptor as a translator of eclectic data into endpoint prediction: Mutagenicity of fullerene as a mathematical function of conditions. Chemosphere, 104, 262–264.

Toropova, A. P., Toropov, A. A., Vesel

<http://doi.org/10.1016/j.chemosphere.2013.10.079>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Salmonella typhimurium TA100

3.2. Endpoint:

In vitro - Mutagenicity - measured as the number of observed colonies

3.3.Comment on endpoint:

The mutagenicity of a substance is proportional to the number of observed colonies.
Data on the bacterial reverse mutation test that was conducted using *Salmonella typhimurium* strain TA100 in the presence and absence of metabolic activation under dark conditions and irradiation are taken in the literature (Shinohara et al., 2009).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Linear regression model
based on SMILES-based optimal descriptors.

4.3.Descriptors in the model:

- "0": Dark condition
- "1": Irradiation condition
- "+": Presence of S9
- "-": Absence of S9
- Dose(g/plate):
 - B: 100
 - C: 200
 - D: 400
 - E: 1000; 8

4.4.Descriptor selection:

Optimal descriptors and Monte-Carlo optimization

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/8

Descriptor: Chemical ratio :8:20 ~ 1:2.5

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

We can apply the range of the final descriptors of training data.

For fullerenes C60 up to 100 nm of diameter

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): Up to 100 nm of diameter

Other info: The specific surface area of purchased C60 before pulverization with beads was 0.92 m²/g

6.6. Pre-processing of data before modelling:

The data are split into the training, calibration, and validation sets according to the following principles: (i) the split is random; and (ii) the ranges of endpoint for the above-mentioned sets are similar

6.7. Statistics for goodness-of-fit:

- Training

$r^2 = 0.7549$

$s = 7.67$

- Calibration

$r^2 = 0.8987$

s = 18.4

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Y-randomization:

- Training set

(c)R²_p = 0.645

- Calibration set

(c)R²_p = 0.759

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

5 MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): Up to 100 nm of diameter

Other properties:

The specific surface area of purchased C60 before pulverization with beads was 0.92 m²/g

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2 = 0.6968$

$s = 10.9$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

There are not structural or molecular descriptors, thus could be not considered as QSAR model.

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models.

It could be noticed that they also applied the same method to E. coli strain WP2 uvrA/pKM101 in a posterior paper:

Toropova, A. P., Toropov, A. A., Veselinović, A. M., Veselinović, J. B., Benfenati, E., Leszczynska, D., & Leszczynski, J. (2016). Nano-QSAR: Model of mutagenicity of fullerene as a mathematical function of different conditions. Ecotoxicology and Environmental Safety, 124, 32–36. 10.1016/j.ecoenv.2015.09.038

r^2 : correlation coefficient

s : root-mean-square error

9.2.Bibliography:

Shinohara, N., Matsumoto, K., Endoh, S., Maru, J., & Nakanishi, J. (2009). In vitro and in vivo genotoxicity tests on fullerene C60 nanoparticles. Toxicology Letters, 191(2-3), 289–296. <http://doi.org/10.1016/j.toxlet.2009.09.012>

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Salmonella typhimurium TA100, QSAR, - "0": Dark condition

- "1": Irradiation condition

- "+": Presence of S9

- "-": Absence of S9

- Dose(g/plate):

· B: 100


· C: 200

· D: 400

· E: 1000,Linear regression model

based on SMILES-based optimal descriptors.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Ecotoxicity model under three organo-coated silver NPs by GLM
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Ecotoxicity model under three organo-coated silver NPs by GLM
(E. coli case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Brajesh Dubey

bdubey@uoguelph.ca

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Silva, T., Pokhrel, L. R., Dubey, B., Tolaymat, T. M., Maier, K. J., & Liu, X. (2014). Particle size, surface charge and concentration dependent ecotoxicity of three organo-coated silver nanoparticles: Comparison between general linear model-predicted and
<http://doi.org/10.1016/j.scitotenv.2013.09.006>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as EC50

3.3. Comment on endpoint:

Concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

GLM: General Linear Model

4.3.Descriptors in the model:

- TEMdiameter (TEM_dia) : representing the primary particle
- Zeta potential (ζ) : function of surface charge
- Interaction term: (TEMdia) x (ζ); 3

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

3/3

Descriptor: Chemical ratio :3:3 ~ 1:1

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

We can apply the range of the final descriptors of training data.

- citrate-coated AgNP (Citrate-AgNP)
- polyvinylpyrrolidone-coated AgNP (PVP-AgNP)
- branched polyethyleneimine- coated AgNP(BPEI-AgNP)

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

3 Metal

List: Ag

Shape: Spherical

Coating: citrate

polyvinylpyrrolidone

branched polyethyleneimine

Size (nm): TEM diameter Mean \pm standard deviation

- (Citrate-AgNP) = 10.0 ± 4.6

- (PVP-AgNP) = 56.0 ± 14.0

- (BPEI-AgNP) = 72.0 ± 24.0

Other info: Synthesis details at publication's supplementary material.

Characterization values at Table 1 in the publication.

Volume-weighted hydrodynamic diameter measured in the test matrix (moderately hardwater) using the DLS method before and after the toxicity tests were conducted; particle circularity of 1 indicates that the particle is a perfect circle in a 2D TEMimagery; PVP-AgNP, polyvinylpyrrolidone-coated AgNPs; Citrate-AgNP, citrate-coated AgNP; BPEI-AgNP, branched polyethyleneimine-coated AgNPs. ImageJ 1.44 program was used to analyze particle size distributions and circularity of the AgNPs from the representative TEM imageries

6.6.Pre-processing of data before modelling:

Since the authors present data as the average of descriptors for each of the three types of NPs, which could be used the averages for the fitting model, or the original ones. Since the particle size distribution was measured from TEM images, it seems that only the averages for the 3 types of NPs have been used in the model, but it is not clear

6.7.Statistics for goodness-of-fit:

$R^2 = 0.999$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA Metal

List

Ag

Shape: SphericalCoating: citrate

polyvinylpyrrolidone

branched polyethyleneimine

Size(nm): TEM diameter Mean \pm standard deviation- (Citrate-AgNP) = 10.0 ± 4.6 - (PVP-AgNP) = 56.0 ± 14.0 - (BPEI-AgNP) = 72.0 ± 24.0 Other properties:

Synthesis details at publication's supplementary material.

Characterization values at Table 1 in the publication.

Volume-weighted hydrodynamic diameter measured in the test matrix (moderately hardwater) using the DLS method before and after the toxicity tests were conducted; particle circularity of 1 indicates that the particle is a perfect circle in a 2D TEMimagery; PVP-AgNP, polyvinylpyrrolidone-coated AgNPs; Citrate-AgNP, citrate-coated AgNP; BPEI-AgNP, branched polyethyleneimine-coated AgNPs. ImageJ 1.44 program was used to analyze particle size distributions and circularity of the AgNPs from the representative TEM imageries

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It doesn't have a good statistical reliability:

It is not clear if it used only 3 data points for building the model, which will be a poor set of data; there are not validation techniques, only the regression for the training data, which could be strong influenced by an overfitting.

Comparison of the means for E. coli EC50 and D. magna LC50 among different treatments was performed using the one-way ANOVA followed by Dunnett t test (2-tailed posthoc) for multiple comparisons.

The toxicity evaluation due free Ag ions was also included, which was greater the three NPs on E.Coli case.

There is a Mechanistic Interpretation

NP: Nanoparticle

GLM: General Linear Model

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

R^2 : correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - TEMdiameter (TEM_dia) : representing the primary particle

- Zeta potential (ζ) : function of surface charge

- Interaction term: (TEMdia) x (ζ), GLM: General Linear Model

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Ecotoxicity model under three organo-coated silver NPs by GLM
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Ecotoxicity model under three organo-coated silver NPs by GLM
(D. Magna case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Brajesh Dubey

bdubey@uoguelph.ca

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Silva, T., Pokhrel, L. R., Dubey, B., Tolaymat, T. M., Maier, K. J., & Liu, X. (2014). Particle size, surface charge and concentration dependent ecotoxicity of three organo-coated silver nanoparticles: Comparison between general linear model-predicted and
<http://doi.org/10.1016/j.scitotenv.2013.09.006>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as LC50

3.3. Comment on endpoint:

Non-renewable D.magna 48 h bioassay was performed following the standard USEPA guidelines, including the culture and main tenance of the daphnids (USEPA, 1987)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

GLM: General Linear Model

4.3.Descriptors in the model:

- TEMdiameter (TEM_dia) : representing the primary particle
- Zeta potential (ζ) : function of surface charge
- Interaction term: (TEMdia) x (ζ); 3

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

3/3

Descriptor: Chemical ratio :3:3 ~ 1:1

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

We can apply the range of the final descriptors of training data.

- citrate-coated AgNP (Citrate-AgNP)
- polyvinylpyrrolidone-coated AgNP (PVP-AgNP)
- branched polyethyleneimine- coated AgNP(BPEI-AgNP)

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

3 Metal

List: Ag

Shape: Spherical

Coating: citrate

polyvinylpyrrolidone

branched polyethyleneimine

Size (nm): TEM diameter Mean \pm standard deviation

- (Citrate-AgNP) = 10.0 ± 4.6

- (PVP-AgNP) = 56.0 ± 14.0

- (BPEI-AgNP) = 72.0 ± 24.0

Other info: Synthesis details at publication's supplementary material.

Characterization values at Table 1 in the publication.

Volume-weighted hydrodynamic diameter measured in the test matrix (moderately hardwater) using the DLS method before and after the toxicity tests were conducted; particle circularity of 1 indicates that the particle is a perfect circle in a 2D TEMimagery; PVP-AgNP, polyvinylpyrrolidone-coated AgNPs; Citrate-AgNP, citrate-coated AgNP; BPEI-AgNP, branched polyethyleneimine-coated AgNPs. ImageJ 1.44 program was used to analyze particle size distributions and circularity of the AgNPs from the representative TEM imageries

6.6.Pre-processing of data before modelling:

Since they present data as the average of descriptors for each of the three types of NPs, which could be used the averages for the fitting model, or the original ones. Since the particle size distribution was measured from TEM images, it seems that only the averages for the 3 types of NPs have been used in the model, but it is not clear

6.7.Statistics for goodness-of-fit:

$R^2 = 0.998$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA M Metal

List

Ag

Shape: SphericalCoating: citrate

polyvinylpyrrolidone

branched polyethyleneimine

Size(nm): TEM diameter Mean \pm standard deviation- (Citrate-AgNP) = 10.0 ± 4.6 - (PVP-AgNP) = 56.0 ± 14.0 - (BPEI-AgNP) = 72.0 ± 24.0 Other properties:

Synthesis details at publication's supplementary material.

Characterization values at Table 1 in the publication.

Volume-weighted hydrodynamic diameter measured in the test matrix (moderately hardwater) using the DLS method before and after the toxicity tests were conducted; particle circularity of 1 indicates that the particle is a perfect circle in a 2D TEMimagery; PVP-AgNP, polyvinylpyrrolidone-coated AgNPs; Citrate-AgNP, citrate-coated AgNP; BPEI-AgNP, branched polyethyleneimine-coated AgNPs. ImageJ 1.44 program was used to analyze particle size distributions and circularity of the AgNPs from the representative TEM imageries

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It doesn't have a good statistical reliability:

It is not clear if it used only 3 data points for building the model, which will be a poor set of data; there are not validation techniques, only the regression for the training data, which could be strong influenced by an overfitting.

Comparison of the means for E. coli EC50 and D. magna LC50 among different treatments was performed using the one-way ANOVA followed by Dunnett t test (2-tailed posthoc) for multiple comparisons.

The toxicity evaluation due free Ag ions was also included, which was greater the three NPs on E.Coli case.

There is a Mechanistic Interpretation

NP: Nanoparticle

GLM: General Linear Model

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

R^2 : correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Organism, Bacteria Escherichia Coli (E. Coli), QSAR, - TEMdiameter (TEM_dia) : representing the primary particle

- Zeta potential (ζ) : function of surface charge

- Interaction term: (TEMdia) x (ζ), GLM: General Linear Model

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cytotoxicity of metal oxide to bacteria E.Coli models by Periodic table-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cytotoxicity of metal oxide to bacteria E.Coli models by Periodic table-based descriptors and stepwise-MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunal Roy

kunalroy_in@yahoo.com

kroy@pharma.jdvu.ac.in

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Kar, S., Gajewicz, A., Puzyn, T., Roy, K., & Leszczynski, J. (2014). Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach. *Ecotoxicology and Environmental Safety*, 107, 162–169.

(MLR ca

<http://doi.org/10.1016/j.ecoenv.2014.05.026>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as pEC50

3.3.Comment on endpoint:

Negative logarithm of concentration of drug/toxic material for the 50 percent of its effect.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Stepwise-MLR (Multiple Linear Regression)

4.3.Descriptors in the model:

- Xox: charge of metal cation corresponding to a given oxide; 0

4.4.Descriptor selection:

The 7 initial descriptors obtained from a periodic-table information, were reduced within the building methods:

- Stepwise-MLR with a "stepping criteria" . The F-value used for inclusion or exclusion of a variable in the stepwise regression process is a test for partial regression coefficient and it is obtained by dividing the difference between reductions of sum of squares with and without the variable being included or excluded with error mean square of the equation.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

11/0

Descriptor: Chemical ratio :01:11

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

AD was verified with leverage approach and Williams plot. (For specific details see (in the publication) Table S4 in Supplementary material)

$h^* = 0.55$

Any outlier was detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

11 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

6.6.Pre-processing of data before modelling:

The dataset was divided randomly into ten different combinations of training and test sets comprising eleven and six compounds respectively.

Splitting of the dataset was performed based on random selection. Compounds present in the training and the test sets for each model is demonstrated in publication's Table S1 (Supplementary material section)

6.7.Statistics for goodness-of-fit:

$R^2 = 0.81-0.90$

RMSE = 0.16-0.22

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2 = 0.74-0.85$

(c) $R^2_p = 0.77-0.86$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

6 Metal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size(nm): 15-90

Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

$R^2_{pred} = 0.72-0.91$

RMSE = 0.15-0.26

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models.

Statistical filled data was used to compare the performance and predictivity of the models against previous studies. Other statistic can be checked in the publication's Table 2 (for PLS) and the publication's Table S2 from Supplementary material (for stepwise-MLR).

The Eq2 and Eq3 in the paper are models which include all the NPs in the training set (without external validation), which means that those one don't have as much reliability as the 10 randomized and splitted data of the other presented models.

There is a Mechanistic Interpretation

r^2 : correlation coefficient

RMSE: root-mean-square-error

Q^2 : leave-one-out cross-validation correlation coefficient

MLR: Multiple Linear Regression

AD: Applicability Domain

h^* : leverage threshold

9.2. Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ... Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nature Nanotechnology, 6(3), 175–178.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC


10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - Xox: charge of metal cation corresponding to a given oxide, Stepwise-MLR (Multiple Linear Regression)

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cytotoxicity of metal oxide to bacteria E.Coli models by Periodic table-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cytotoxicity of metal oxide to bacteria E.Coli models by Periodic table-based descriptors and PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Kunal Roy

kunalroy_in@yahoo.com

kroy@pharma.jdvu.ac.in

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Kar, S., Gajewicz, A., Puzyn, T., Roy, K., & Leszczynski, J. (2014). Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach. *Ecotoxicology and Environmental Safety*, 107, 162–169.

(PLS ca

<http://doi.org/10.1016/j.ecoenv.2014.05.026>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as pEC50

3.3.Comment on endpoint:

Negative logarithm of concentration of drug/toxic material for the 50 percent of its effect.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- Xox: charge of metal cation corresponding to a given oxide
- X: metal electronegativity; 2

4.4.Descriptor selection:

The 7 initial descriptors obtained from a periodic-table information, were reduced within the building methods:

- PLS uses the 7 descriptor and then omitting the less significant.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

11/2

Descriptor: Chemical ratio :2:11 ~ 1:6

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

AD was verified with leverage approach and Williams plot. (For specific details see (in the publication) Table 4)

$h^* = 0.82$

Any outlier was detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

11 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material

in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

6.6.Pre-processing of data before modelling:

The dataset was divided randomly into ten different combinations of training and test sets comprising eleven and six compounds respectively.

Splitting of the dataset was performed based on random selection. Compounds present in the training and the test sets for each model is demonstrated in publication's Table S1 (Supplementary material section)

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.73-0.87$$

$$RMSE = 0.19-0.27$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$$Q^2 = 0.55-0.78$$

$$(c)R^2_p = 0.72-0.81$$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

6 M Metal Oxide

List

ZnO

CuO

Al₂O₃Fe₂O₃SnO₂TiO₂V₂O₃Y₂O₃Bi₂O₃In₂O₃Sb₂O₃SiO₂ZrO₂

CoO

NiO

Cr₂O₃La₂O₃Shape:NACoating:NASize(nm): 15-90Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation: $R^2_{pred} = 0.65-0.88$

RMSE = 0.17-0.29

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models.

Statistical filled data was used to compare the performance and predictivity of the models against previous studies. Other statistic can be checked in the publication's Table 2 (for PLS) and the publication's Table S2 from Supplementary material (for stepwise-MLR).

The Eq2 and Eq3 in the paper are models which include all the NPs in the training set (without external validation), which means that those one don't have as much reliability as the 10 randomized and splitted data of the other presented models.

There is a Mechanistic Interpretation

r^2 : correlation coefficient

RMSE: root-mean-square-error

Q^2 : leave-one-out cross-validation correlation coefficient

PLS: Partial Least Squares

AD: Applicability Domain

h^* : leverage threshold

9.2. Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ... Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nature Nanotechnology, 6(3), 175–178.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - Xox: charge of metal cation corresponding to a given oxide

- X: metal electronegativity, PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: cytotoxicity of SiO₂ NPs on human lung fibroblast model by SMILES-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

cytotoxicity of SiO₂ NPs on human lung fibroblast model by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software

package:

Toropova, A. P., Toropov, A. A., Benfenati, E., & Korenstein, R. (2014). QSAR model for cytotoxicity of SiO₂ nanoparticles on human lung fibroblasts. Journal of Nanoparticle Research, 16(2).

<http://doi.org/10.1007/s11051-014-2282-9>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human lung fibroblast

3.2. Endpoint:

In vitro - Cytotoxicity - measured as Inhibition Ratio (IR%)

3.3. Comment on endpoint:

MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromide] assay has been used as the endpoint for the QSAR analysis.

HFL-I cell line was maintained in logarithmic phase of growth by subculturing at appropriate stages and in MEM with glutamine supplemented with 10% FBS and 1% penicillin streptomycin (Gibco, Invitrogen, NY, USA). The cell line was maintained at 37 °C in a 5% CO₂ incubator. Appropriately assigned concentrations of SiO₂ NPs were then added to the cell cultures at the stage of 80% confluency and incubated for 24 h.

The exponentially growing HFL-I s (1 × 10⁵ per well) were seeded into 96-well culture plates and incubated with various concentrations of SiO₂ NPs (20 and 80 nm) for 48 h. Four hours before termination, the supernatants were replaced with 90 ml fresh medium and 10 ml MTT (1 mg ml⁻¹) solution. After incubating for 4 h at 37 °C, the medium was aspirated and the formazan crystals were solubilised in acidified isopropanol. The reduction of absorbance was measured spectrophotometrically at 570 and 630 nm, and the Cellular viability was expressed as a percentage over the untreated control

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

Split1:

1, 2, a, g, h

Split2:

1, 2, a, b, h

Split3:

1, 2, a, b, g, h

Codes:

- Size of NP:

20 nm (1),

80 nm (2)

- Concentrations:

250 µg/ml (a),

500 µg/ml (b),

750 µg/ml (c),

1,000 µg/ml (d),

1,250 µg/ml (e),

1,500 µg/ml (f),

1,750 µg/ml (g),

2,000 µg/ml (h); 0

4.4.Descriptor selection:

Optimal descriptors and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :Split1:

5:2

Split2:

5:2

Split3:

6:2

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper

It suggested that could be applicable to other NPs, but without any probe we can only say that AD should be for SiO₂ NPs within the size range and the concentration range (range of descriptors)

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: SiO₂

Shape: NA

Coating: NA

Size (nm): 20 and 80

Other info: The physical and chemical properties of SiO₂ NPs were characterized. The SiO₂ NP's structure was confirmed by transmission electron microscopy (TEM; JEM-2010, Jeol Ltd, Japan). The size distribution was analyzed by laser scattering (ELS-Z, Otsuka Electronics, Japan).

6.6.Pre-processing of data before modelling:

These 16 systems (size–concentration–endpoint) were randomly split into the training set (systems which are involved in model development) and test set (systems which are not involved in model development). Three various splits were examined in the present study

6.7.Statistics for goodness-of-fit:

Split1:

$R^2 = 0.8843$

$s = 6.67 \%$

Split2:

$R^2 = 0.8694$

$s = 6.56 \%$

Split3:

$R^2 = 0.9837$

$s = 2.53 \%$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

Split1:

5

Split2:

5

Split3:

6 Metal Oxide

List

SiO₂

Shape: NA

Coating: NA

Size(nm): 20 and 80

Other properties:

The physical and chemical properties of SiO₂ NPs were characterized. The SiO₂ NP's structure was confirmed by transmission electron microscopy (TEM; JEM-2010, Jeol Ltd, Japan). The size distribution was analyzed by laser scattering (ELS-Z, Otsuka Electronics, Japan).

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

Split1:

$R^2 = 0.6639$

$s = 7.63 \%$

Split2:

$R^2 = 0.8074$

$s = 9.09 \%$

Split3:

$R^2 = 0.9269$

$s = 7.94 \%$

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Really small data set, lack in reliability of the model.

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

CORAL: CORrelation And Logic

NPs: Nanoparticles

R²: Correlation coefficient

s: Normalized root-mean-square error

9.2. Bibliography:

Xu Z, Chou L, Sun J (2012) Effects of SiO₂ nanoparticles on HFL-I activating ROS-mediated apoptosis via p53 path- way. J Appl Toxicol 32:358–364

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Human lung fibroblast, QSAR, Split1:

1, 2, a, g, h

Split2:

1, 2, a, b, h

Split3:

1, 2, a, b, g, h

Codes:

- Size of NP:

20 nm (1),

80 nm (2)

- Concentrations:

250 µg/ml (a),

500 µg/ml (b),

750 µg/ml (c),

1,000 µg/ml (d),

1,250 µg/ml (e),


1,500 µg/ml (f),

1,750 µg/ml (g),

2,000 µg/ml (h), Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Photo-induced toxicity of metal oxide NPs to E. Coli by MLR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Photo-induced toxicity of metal oxide NPs to E. Coli by MLR
(dark condition case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

M.-J. Huang

H.-M. Hwang

ming-ju.huang@jsums.edu

huey-min.hwang@jsums.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Pathakoti, K., Huang, M.-J., Watts, J. D., He, X., & Hwang, H.-M.
(2014). Using experimental data of Escherichia coli to develop a
QSAR model for predicting the photo-induced cytotoxicity of metal
oxide nanoparticles. Journal of Photochemistry and Photobi
<http://doi.org/10.1016/j.jphotobiol.2013.11.023>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as $-\log(\text{LC}_{50})$

3.3.Comment on endpoint:

The samples in Pyrex and quartz test tubes were exposed to sunlight for 30 min with agitation in a water bath at 150 r.p.m. Similarly, corresponding samples were exposed under dark conditions, by wrapping the test samples in the test tubes with aluminium foils (solar irradiation outdoors: irradiance: UVA range = 3.979–4.652 mW/cm²; UVB range = 3.1–3.7 MED/h, where an MED is defined as the minimum erythral dose or the amount of UV radiation to produce barely perceptible erythema; 1 MED/h = 0.05833 W/m²).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- MELECT: the absolute electronegativity of the metal atom
 - LZELEHHO: the absolute electronegativity of the metal oxide
- ; 4

4.4.Descriptor selection:

Set of descriptors were obtained from density functional theory (DFT) with B3LYP functional and the LANL2Z basis sets, also the semi-empirical molecular orbital method PM6 was used. On the metal atoms' calculations the CCSD(T) method and the QZVP basis set were applied. All calculations were performed with Gaussian 2009 program.

Also tabulated data was included in the study.

The workflow to select the final descriptors was to select the pair of descriptors in the MLR which more reduces the correlation coefficient.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

13/4

Descriptor: Chemical ratio :2:17 ~ 1:9

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

It should be considered at least for metal oxide NPs within the range of

the characterization data and the applied descriptors.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

13 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr2O3

La2O3

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Primary particle was measured by using transmission electron microscopy (TEM). Samples were prepared by drop-coating the NP suspension onto a carbon-coated copper grid (Ted Pella, CA) and then the samples were dried overnight at room temperature. The samples were observed using a TEM (JEOL JEM-1011). The hydrodynamic diameters (z-average) were measured in distill water (at a concentration of 100 ppm in water) and zeta potentials of the MNPs were measured in both distill water and 1mM KCl solution using Malvern Zeta Sizer (Nano-ZS, Malvern Instruments, UK). All measurements were conducted in triplicate at 25 °C and an average values was determined.

6.6.Pre-processing of data before modelling:

Among the 17 MNPs, there are three different types of metal oxides (MO,MO2, and M2O3). They have four MO, four MO2, and nine M2O3. They chose one MO, oneMO2, and two M2O3 as our prediction set. In addition, they made sure the LC50 values for those four MNPs are not the largest or smallest values in their subsets.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.87$

SD = 0.48

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

4 M Metal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size(nm): 15-90

Other properties:

Primary particle was measured by using transmission electron microscopy (TEM). Samples were prepared by drop-coating the NP suspension onto a carbon-coated copper grid (Ted Pella, CA) and then the samples were dried overnight at room temperature. The samples were observed using a TEM (JEOL JEM-1011). The hydrodynamic diameters (z-average) were measured in distilled water (at a concentration of 100 ppm in water) and zeta potentials of the MNPs were measured in both distilled water and 1mM KCl solution using Malvern Zeta Sizer (Nano-ZS, Malvern Instruments, UK). All measurements were conducted in triplicate at 25 °C and an average values was determined.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

Experimental vs. calculated values in publication's table 3

The percentage of average error from the prediction set:

13.86 %

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Although the prediction set was used to compare the experimental and the calculated values of $-\log LC_{50}$, no statistical value was computed. The percentage of average error in the prediction set, was computed to include in this table.

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

DFT: Density Functional Theory

MLR: Multiple Linear Regression

R^2 : Correlation coefficient

SD: Standard deviation

F: Fisher's

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC


10.3. Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - MELECT: the absolute electronegativity of the metal atom

- LZELEHHO: the absolute electronegativity of the metal oxide

,MLR: Multiple Linear Regression

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Photo-induced toxicity of metal oxide NPs to E. Coli by MLR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Photo-induced toxicity of metal oxide NPs to E. Coli by MLR
(Photo-induced (light) case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

M.-J. Huang

H.-M. Hwang

ming-ju.huang@jsums.edu

huey-min.hwang@jsums.edu

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Pathakoti, K., Huang, M.-J., Watts, J. D., He, X., & Hwang, H.-M.
(2014). Using experimental data of Escherichia coli to develop a
QSAR model for predicting the photo-induced cytotoxicity of metal
oxide nanoparticles. Journal of Photochemistry and Photobi
<http://doi.org/10.1016/j.jphotobiol.2013.11.023>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as $-\log(\text{LC}_{50})$

3.3.Comment on endpoint:

The samples in Pyrex and quartz test tubes were exposed to sunlight for 30 min with agitation in a water bath at 150 r.p.m. Similarly, corresponding samples were exposed under dark conditions, by wrapping the test samples in the test tubes with aluminium foils (solar irradiation outdoors: irradiance: UVA range = 3.979–4.652 mW/cm²; UVB range = 3.1–3.7 MED/h, where an MED is defined as the minimum erythral dose or the amount of UV radiation to produce barely perceptible erythema; 1 MED/h = 0.05833 W/m²).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- C_p is the literature molar heat capacity of the metal oxide at 298.15 K.
 - ALZLUMO is the average of the alpha and beta LUMO energies of the metal oxide.
- ; 4

4.4.Descriptor selection:

Set of descriptors were obtained from density functional theory (DFT) with B3LYP functional and the LANL2Z basis sets, also the semi-empirical molecular orbital method PM6 was used. On the metal atoms' calculations the CCSD(T) method and the QZVP basis set were applied. All calculations were performed with Gaussian 2009 program.

Also tabulated data was included in the study.

The workflow to select the final descriptors was to select the pair of descriptors in the MLR which more reduces the correlation coefficient.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

13/4

Descriptor: Chemical ratio :2:17 ~ 1:9

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

It should be considered at least for metal oxide NPs within the range of

the characterization data and the applied descriptors.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for the dependent variable for the training set:

Yes

6.4.Other information about the training set:

13 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Primary particle was measured by using transmission electron microscopy (TEM). Samples were prepared by drop-coating the NP suspension onto a carbon-coated copper grid (Ted Pella, CA) and then the samples were dried overnight at room temperature. The samples were observed using a TEM (JEOL JEM-1011). The hydrodynamic diameters (z-average) were measured in distill water (at a concentration of 100 ppm in water) and zeta potentials of the MNPs were measured in both distill water and 1mM KCl solution using Malvern Zeta Sizer (Nano-ZS, Malvern Instruments, UK). All measurements were conducted in triplicate at 25 °C and an average values was determined.

6.5.Data for each descriptor variable for the training set:

Yes

6.6.Pre-processing of data before modelling:

Among the 17 MNPs, there are three different types of metal oxides (MO, MO₂, and M₂O₃). They have four MO, four MO₂, and nine M₂O₃. They chose one MO, one MO₂, and two M₂O₃ as our prediction set. In addition, they made sure the LC₅₀ values for those four MNPs are not the largest or smallest values in their subsets.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.804$

SD = 0.63

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

4 M Metal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size(nm): 15-90

Other properties:

Primary particle was measured by using transmission electron microscopy (TEM). Samples were prepared by drop-coating the NP suspension onto a carbon-coated copper grid (Ted Pella, CA) and then the samples were dried overnight at room temperature. The samples were observed using a TEM (JEOL JEM-1011). The hydrodynamic diameters (z-average) were measured in distilled water (at a concentration of 100 ppm in water) and zeta potentials of the MNPs were measured in both distilled water and 1mM KCl solution using Malvern Zeta Sizer (Nano-ZS, Malvern Instruments, UK). All measurements were conducted in triplicate at 25 °C and an average value was determined.

7.6. Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Experimental vs.calculated values in publication's table 5

The percentage of average error from the prediction set:

20.59 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

Although the prediction set was used to compare the experimental and the calculated values of -logLC50, no statistical value was computed. The percentage of average error in the prediction set, was computed to include in this table.

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

DFT: Density Functional Theory

MLR: Multiple Linear Regression

R²: Correlation coefficient

SD: Standard deviation

F: Fisher's

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - Cp is the literature molar heat capacity of the metal oxide at 298.15 K.

- ALZLUMO is the average of the alpha and beta LUMO energies of the metal oxide.

,MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: quasi-QSPR model for photocatalytic decolourization rate constants by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

quasi-QSPR model for photocatalytic decolourization rate constants by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software

package:

Toropova, A. P., Toropov, A. A., & Benfenati, E. (2015). A quasi-QSPR modelling for the photocatalytic decolourization rate constants and cellular viability (CV%) of nanoparticles by CORAL. SAR and QSAR in Environmental Research, 26(1), 29–40.

Case stud

<http://doi.org/10.1080/1062936X.2014.984327>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Photocatalytic decolourization rate constants DRC (10^{-5} /s) of methylene blue dye

3.3.Comment on endpoint:

The photocatalytic performance of each sample was tested by dye (methylene blue) decolourisation under simulated solar light illumination (mixture of UV and visible), with regular measurements of the dye concentration (determined by absorbance at 664 nm). (Goodall, 2010) The dye decolourisation rate is defined as $d \ln[C(t)/C_0]/dt$, where C_0 is the initial dye concentration and $C(t)$ is the dye concentration at time t .

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSPR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

A: concentration of 0.5 mol %,

B: concentration of 2.5 mol%

C : concentration of 5.0 mol%

Ag

Ce

Co

Cr

Fe

Er

Ga

Gd

La

Nd

Mn

Ni

V

Pr

Y

Sr

Zn; 20

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

0/20

Descriptor: Chemical ratio :20:36 ~ 1:2

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

It should be considered within the range of the characterization data and the applied descriptors to doped TiO₂ NPs

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Metal Oxide

List: Doped TiO₂ nanopowder by:

Ag

Ce

Co

Cr

Fe

Er

Ga

Gd

La

Nd

Mn

Ni

V

Pr

Y

Sr

Zn

Shape: NACoating: NASize (nm): NAOther info: NA**6.6.Pre-processing of data before modelling:**

The experimental data set of dopants was randomly distributed into the training, test and validation sets.

Three splits built up according to the above-mentioned principles are examined in the present study

6.7.Statistics for goodness-of-fit:

Split1:

$$R^2 = 0.8974$$

$$s = 2.26$$

Split2:

$$R^2 = 0.8741$$

$$s = 2.77$$

Split3:

$$R^2 = 0.8759$$

$$s = 2.91$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Split1:

$$q^2 = 0.8629$$

Split2:

$q^2 = 0.8426$

Split3:

$q^2 = 0.8488$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

split1 : 9

split2 : 8

split 3 : 8 Metal Oxide

List

Doped TiO₂ nanopowder by:

Ag

Ce

Co

Cr

Fe

Er

Ga

Gd

La

Nd

Mn

Ni

V

Pr

Y

Sr

Zn

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Split1:

$R^2 = 0.6810$

$s = 3.91$

Split2:

$R^2 = 0.7633$

$s = 3.38$

Split3:

$R^2 = 0.9766$

$s = 1.41$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

This is the case study 1 of the paper. As they explain in the introduction, could be not classified as QSPR/QSAR, due to the descriptors that they choose, which are not structural or physicochemical properties.

In that particular case, it will be considered as QSPR.

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm^2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the majority of the classified models, we have decided only to mention

that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models.

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

R^2 : correlation coefficient

s: root-mean-square error

q^2 : cross validation coefficient

CORAL: CORrelation And Logic

9.2. Bibliography:

Y. Yang, T. Lin, X.L. Weng, J.A. Darr, and X.Z. Wang, Data flow modelling, data mining and QSAR in high-throughput discovery of functional nanomaterials, Comput. Chem. Engin. 35 (2011), pp. 671–678

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, A: concentration of 0.5 mol %,

B: concentration of 2.5 mol%

C : concentration of 5.0 mol%

Ag

Ce

Co

Cr

Fe

Er

Ga

Gd

La

Nd

Mn

Ni

V

Pr


Y

Sr

Zn, Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: quasi-QSAR model for cellular viability by SMILES-based optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

quasi-QSAR model for cellular viability by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software

package:

Toropova, A. P., Toropov, A. A., & Benfenati, E. (2015). A quasi-QSPR modelling for the photocatalytic decolourization rate constants and cellular viability (CV%) of nanoparticles by CORAL. SAR and QSAR in Environmental Research, 26(1), 29–40.

Case stud

<http://doi.org/10.1080/1062936X.2014.984327>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human kidney (HK-2) cells

and

Porcine tubular LLC-PK₁ cells

3.2.Endpoint:

In vitro - Cytotoxicity - measured as percentage of cellular viability

3.3.Comment on endpoint:

The effects of SiO₂ nanoparticles on HK-2 and LLC-PK1 cells were determined by WST-1.

The WST-1 assay evaluates cellular mitochondrial activity. It is based on cleavage of the tetrazolium salt to a soluble formazan dye by succinate-tetrazolium reductase, a mitochondrial enzyme that is active only in viable cells.

% of cellular viability was calculated using the formula :

(absorbance treated sample × 100/absorbance control sample)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

- '1' and '2' : 20 nm and 100 nm size respectively
- 'A', 'B', 'C', 'D', 'E', 'F', and 'G' : concentration (5, 10, 40, 70, 100, 400, 700 µg/ml)
- 'x', 'y', and 'z' :exposure times (24, 48, and 72 h); 12

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/12

Descriptor: Chemical ratio :12:50 ~ 1:4

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

It should be considered within the range of the characterization data and the applied descriptors to SiO₂ NPs

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: SiO₂

Shape: Spherical

Coating: NA

Size (nm): 20 and 100

Other info: NA

6.6.Pre-processing of data before modelling:

The experimental data set of dopants was randomly distributed into the training, test and validation sets.

6.7.Statistics for goodness-of-fit:

Split1:

$R^2 = 0.8289$

$s = 13.4$

$F = 44$

Split2:

$R^2 = 0.8420$

$s = 15.5$

F = 37

Split3:

$R^2 = 0.8866$

s = 10.4

F = 70

Calibration:

Split1:

$R^2 = 0.7679$

s = 13.9

Split2:

$R^2 = 0.8424$

s = 2.77

Split3:

$R^2 = 0.8890$

s = 12.4

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Split1:

$q^2 = 0.7070$

Split2:

$q^2 = 0.7769$

Split3:

$q^2 = 0.8169$

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

split1 : 9

split2 : 12

split 3 : 9 Metal Oxide

List

SiO₂

Shape:Spherical

Coating:NA

Size(nm): 20 and 100

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Split1:

$R^2 = 0.8680$

$s = 14.5$

Split2:

$R^2 = 0.8662$

$s = 15.7$

Split3:

$R^2 = 0.8459$

$s = 14.0$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

This is the case study 2 of the paper. As they explain in the introduction, could be not classified as QSPR/QSAR, due to the descriptors that they choose, which are not structural or physicochemical properties.

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models.

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

The same model was developed in posterior publication, with a different source of data, and AD was computed:

Manganelli, S., Leone, C., Toropov, A. A., Toropova, A. P., & Benfenati, E. (2016). QSAR model for predicting Cellular viability of human embryonic kidney cells exposed to SiO₂ nanoparticles. Chemosphere, 144, 995–1001. 10.1016/j.chemosphere.2015.09.086

R²: correlation coefficient

s: root-mean-square error

q²: cross-validation correlation coefficient

CORAL: CORrelation And Logic

9.2. Bibliography:

I. Passagne, M. Morille, M. Rousset, I.L. Pujalté, and B. L'Azou, Implication of oxidative stress in size-dependent toxicity of silica nanoparticles in kidney cells, Toxicology 299 (2012), pp. 112–124

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC


10.3. Keywords:

Cell, Human kidney (HK-2) cells
and

Porcine tubular LLC-PK₁ cells, QSAR, - '1' and '2' : 20 nm and 100 nm size respectively
- 'A', 'B', 'C', 'D', 'E', 'F', and 'G' : concentration (5, 10, 40, 70, 100, 400, 700 µg/ml)
- 'x', 'y', and 'z' : exposure times (24, 48, and 72 h), Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Determine the impact of surface modifications on MF phenotype by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS
(Arginase:iNOS case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

K.M. Bratlie

kbratlie@iastate.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Bygd, H. C., Forsmark, K. D., & Bratlie, K. M. (2015). Altering in vivo macrophage responses with modified polymer properties. *Biomaterials*, 56, 187–197.

(Arginase:iNOS case)

<http://doi.org/10.1016/j.biomaterials.2015.03.042>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Macrophage

on 6-week-old female SKH-1 mice

3.2. Endpoint:

In vivo - Cell differentiation response - measured by concentration of Arginase:iNOS

3.3.Comment on endpoint:

The library of materials used by Wang et al., was injected into SKH1-E mice and monitored for 7 days through in vivo fluorescence imaging. Histological analysis and ex vivo assays were performed on tissues collected 7 days post-injection, when the highest-level of MF presence is expected.

The different evaluated endpoints are related with the differentiation of macrophages in M1 or M2.

Cathepsin activity measured at 7 days was negatively correlated with TNF-a and positively correlated with arginase:iNOS, demonstrating that cathepsin is positively correlated with increased M2 MFs and negatively correlated with M1 MFs.

High levels of IL-10 are commonly associated with M2 MFs and tumour progression, and high levels of TNF-a are commonly associated with M1 MFs and tumour suppression

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- H_A: H-bond acceptors
- N_(sp²): Number of sp² carbon atoms
- N: Number of non-hydrogen atoms
- N_O: Number of oxygen atoms
- N_K: $N_K = 5N_{\text{amide}} + 4N_{\text{hydroxyl}} - 3N_{\text{ether}} - 5N_{\text{(C=C)}} - 3N_{\text{sulfone}}$
- N_group: $N_{\text{group}} = 12N_{\text{hydroxyl}} - 12N_{\text{amide}} - 2N_{\text{(non-amide-NH-unite)}} - N_{\text{(alkyl ether - O -)}} - N_{\text{(C=C)}} - 4N_{\text{(non-amide-(C=O)-next to a nitrogen)}} - 7N_{\text{((C=O) in carboxylic acid, ketone or aldehyde)}} - 2N_{\text{(other -(C=O)-)}}; 0$

4.4.Descriptor selection:

Original descriptors were selected from Bicerano J. Prediction of polymer properties. CRC Press; 2002.

Within the model building algorithm PLS the most contributing descriptors were selected for the final models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

14/0

Descriptor: Chemical ratio :6:14 ~1:2

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Should be considered p(NIPAm-co-AAc) with other functionalizations that fall in the range of applied descriptors.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

14 Polymeric

List: p(NIPAm-co-AAc)

Shape: NA

Coating: 3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4-aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic acid and ethylene ketal
 2-aminoethylmethylsulfone hydrochloride
 3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)
 Aminomethylphosphonic acid (Alfa Aesar, Ward Hill, MA)
 and without functionalization

Size (nm): 600

Other info: N-iso-propylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

6.6.Pre-processing of data before modelling:

No splitting data is reported, but cross-validation was applied. NA in "NPs used as test set" was defined because there is no specification about if it was an LOO or LMO (the number of folds was not specified).

6.7.Statistics for goodness-of-fit:

$R^2Y = 0.923$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2Y = 0.719$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MPolymeric

List

p(NIPAm-co-AAc)

Shape:NA

Coating:3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4- aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic and ethylene ketal

2-aminoethylmethylsulfone hydrochloride

3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)

Aminomethylphosphonic acid (Alfa Aesar,Ward Hill, MA)

and without functionalization

Size(nm): 600

Other properties:

N-iso-propylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

No external validation is provided, it is commented as future work, with more data. We should take into account the possible overfitting.

ANOVA test was performed to confirm the statistical significance with $p < 0.05$.

There is a Mechanistic Interpretation

They are not nanomaterials according to EU definition

PLS: Partial least squares

R^2Y : correlation coefficient

Q^2Y : cross-validation correlation coefficient

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Macrophage

on 6-week-old female SKH-1 mice, QSAR, - H_A: H-bond acceptors

- N_(sp²): Number of sp² carbon atoms

- N: Number of non-hydrogen atoms


- N_O: Number of oxygen atoms

- N_K: $N_K = 5N_{amide} + 4N_{hydroxyl} - 3N_{ether} - 5N_{(C=C)} - 3N_{sulfone}$

- N_{group}: $N_{group} = 12N_{hydroxyl} - 12N_{amide} - 2N_{(non-amide-NH-unite)} - N_{(alkyl\ ether - O -)}$

- N_(C=C) - 4N_{(non-amide-(C=O)-next to a nitrogen)} - 7N_{((C=O) in carboxylic acid, ketone or aldehyde)} - 2N_{(other -(C=O)-)}, PLS: Partial Least Squares

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Determine the impact of surface modifications on MF phenotype by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS
(IL-10 case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

K.M. Bratlie

kbratlie@iastate.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Bygd, H. C., Forsmark, K. D., & Bratlie, K. M. (2015). Altering in vivo macrophage responses with modified polymer properties. *Biomaterials*, 56, 187–197.

(IL-10 case)

<http://doi.org/10.1016/j.biomaterials.2015.03.042>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Macrophage

on 6-week-old female SKH-1 mice

3.2. Endpoint:

In vivo - Cell differentiation response - measured by concentration of IL-10

3.3.Comment on endpoint:

The library of materials used by Wang et al., was injected into SKH1-E mice and monitored for 7 days through in vivo fluorescence imaging. Histological analysis and ex vivo assays were performed on tissues collected 7 days post-injection, when the highest-level of MF presence is expected.

The different evaluated endpoints are related with the differentiation of macrophages in M1 or M2. Cathepsin activity measured at 7 days was negatively correlated with TNF- α and positively correlated with arginase:iNOS, demonstrating that cathepsin is positively correlated with increased M2 MFs and negatively correlated with M1 MFs.

High levels of IL-10 are commonly associated with M2 MFs and tumour progression, and high levels of TNF- α are commonly associated with M1 MFs and tumour suppression

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- (1) χ : Intensive connectivity index 1
- (0) χ : Intensive atomic index 1
- N_C: Number of carbon atoms
- N_(sp²): Number of sp² carbon atoms
- N_{dc}: N_{dc} = 19N_N - 12N(side group O, S) - 52N_{sulfone} - 14N_{cyc}
- T: Transition temperature; 0

4.4.Descriptor selection:

Original descriptors were selected from Bicerano J. Prediction of polymer properties. CRC Press; 2002.

Within the model building algorithm PLS the most contributing descriptors were selected for the final models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

14/0

Descriptor: Chemical ratio :6:14 ~ 1:2

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Should be considered p(NIPAm-co-AAc) with other functionalizations that fall in the range of applied descriptors.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

14 Polymeric

List: p(NIPAm-co-AAc)

Shape: NA

Coating: 3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4- aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic and ethylene ketal

2-aminoethylmethylsulfone hydrochloride

3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)

Aminomethylphosphonic acid (Alfa Aesar, Ward Hill, MA)

and without functionalization

Size (nm): 600

Other info: (N-isopropylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

6.6.Pre-processing of data before modelling:

No splitting data is reported, but cross-validation was applied. NA in "NPs used as test set" was defined because there is no specification about if it was an LOO or LMO (the number of folds was not specified).

6.7.Statistics for goodness-of-fit:

$R^2Y = 0.936$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2Y = 0.861$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MPolymeric

List

p(NIPAm-co-AAc)

Shape:NA

Coating:3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4- aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic and ethylene ketal

2-aminoethylmethylsulfone hydrochloride

3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)

Aminomethylphosphonic acid (Alfa Aesar,Ward Hill, MA)

and without functionalization

Size(nm): 600

Other properties:

(N-isopropylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

No external validation is provided, it is commented as future work, with more data. We should take into account the possible overfitting.

ANOVA test was performed to confirm the statistical significance with $p < 0.05$.

There is a Mechanistic Interpretation

They are not nanomaterials according to EU definition

PLS: Partial least squares

R^2Y : correlation coefficient

Q^2Y : cross-validation correlation coefficient

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Macrophage

on 6-week-old female SKH-1 mice, QSAR, - (1) ξ : Intensive connectivity index 1

- (0) ξ : Intensive atomic index 1


- N_C : Number of carbon atoms

- $N_{(sp^2)}$: Number of sp^2 carbon atoms

- N_{dc} : $N_{dc} = 19N_N - 12N(\text{side group O, S}) - 52N_{\text{sulfone}} - 14N_{\text{cyc}}$

- T: Transition temperature, PLS: Partial Least Squares

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Determine the impact of surface modifications on MF phenotype by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS
(TNF- α case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

K.M. Bratlie

kbratlie@iastate.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Bygd, H. C., Forsmark, K. D., & Bratlie, K. M. (2015). Altering in vivo macrophage responses with modified polymer properties. *Biomaterials*, 56, 187–197.

(TNF- α case)

<http://doi.org/10.1016/j.biomaterials.2015.03.042>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Macrophage

on 6-week-old female SKH-1 mice

3.2. Endpoint:

In vivo - Cell differentiation response - measured by concentration of TNF- α

3.3.Comment on endpoint:

The library of materials used by Wang et al., was injected into SKH1-E mice and monitored for 7 days through in vivo fluorescence imaging. Histological analysis and ex vivo assays were performed on tissues collected 7 days post-injection, when the highest-level of MF presence is expected.

The different evaluated endpoints are related with the differentiation of macrophages in M1 or M2.

Cathepsin activity measured at 7 days was negatively correlated with TNF- α and positively correlated with arginase:iNOS, demonstrating that cathepsin is positively correlated with increased M2 MFs and negatively correlated with M1 MFs.

High levels of IL-10 are commonly associated with M2 MFs and tumour progression, and high levels of TNF- α are commonly associated with M1 MFs and tumour suppression

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- (1) χ^v : Intensive connectivity index 2
- N_C: Number of carbon atoms
- N_group
- N_PS: $N_{PS} = 3N_{(sp^3 \text{ carbon atoms})} - 6N_{(\text{carbonyl groups})}$
- N_vdW: $N_{vdW} = N_{\text{menonar}} - N_{\text{alamid}} - N_{OH} - 4N_{cyc} - 2N_{(C=C)}$, (N_{menonar} is the number of methyl groups attached to non-aromatic atoms.)
- Vw = $3.8618030X - 13.7484351X_v$, van der Waals volume
- T: Transition temperature
- WCA: Water contact angle; 0

4.4.Descriptor selection:

Original descriptors were selected from Bicerano J. Prediction of polymer properties. CRC Press; 2002.

Within the model building algorithm PLS the most contributing descriptors were selected for the final models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

14/0

Descriptor: Chemical ratio :8:14 ~1:2

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Should be considered p(NIPAm-co-AAc) with other functionalizations that fall in the range of applied descriptors.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

14 Polymeric

List: p(NIPAm-co-AAc)

Shape: NA

Coating: 3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4-aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime
 2,4-dinitro-phenyl-hydroxylamine
 1-amino-4-oxocyclohexanecarboxylic acid ethylene ketal
 2-aminoethylmethylsulfone hydrochloride
 3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)
 Aminomethylphosphonic acid (Alfa Aesar, Ward Hill, MA)
 and without functionalization

Size (nm): 600

Other info: (N-isopropylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

6.6.Pre-processing of data before modelling:

No splitting data is reported, but cross-validation was applied. NA in "NPs used as test set" was defined because there is no specification about if it was an LOO or LMO (the number of folds was not specified).

6.7.Statistics for goodness-of-fit:

$R^2Y = 0.968$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2Y = 0.768$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MPolymeric

List

p(NIPAm-co-AAc)

Shape:NA

Coating:3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4- aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic and ethylene ketal

2-aminoethylmethylsulfone hydrochloride

3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)

Aminomethylphosphonic acid (Alfa Aesar,Ward Hill, MA)

and without functionalization

Size(nm): 600

Other properties:

(N-isopropylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

No external validation is provided, it is commented as future work, with more data. We should take into account the possible overfitting.

ANOVA test was performed to confirm the statistical significance with $p < 0.05$.

There is a Mechanistic Interpretation

They are not nanomaterials according to EU definition

PLS: Partial least squares

R^2_Y : correlation coefficient

Q^2_Y : cross-validation correlation coefficient

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Macrophage

on 6-week-old female SKH-1 mice, QSAR, - (1) ξ^v : Intensive connectivity index 2

- N_C : Number of carbon atoms

- N_{group}

- N_{PS} : $N_{PS} = 3N_{(sp^3 \text{ carbon atoms})} - 6N_{(carbonyl \text{ groups})}$


- N_{vdW} : $N_{vdW} = N_{menonar} - N_{alamid} - N_{OH} - 4N_{cyc} - 2N_{(C=C)}$, ($N_{menonar}$ is the number of methyl groups attached to non-aromatic atoms.)

- $V_w = 3.8618030X - 13.7484351X_v$, van der Waals volume

- T : Transition temperature

- WCA: Water contact angle, PLS: Partial Least Squares

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Determine the impact of surface modifications on MF phenotype by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS
(Cathepsin_3day case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

K.M. Bratlie

kbratlie@iastate.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Bygd, H. C., Forsmark, K. D., & Bratlie, K. M. (2015). Altering in vivo macrophage responses with modified polymer properties. *Biomaterials*, 56, 187–197.

(Cathepsin_3day case)

<http://doi.org/10.1016/j.biomaterials.2015.03.042>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Macrophage

on 6-week-old female SKH-1 mice

3.2. Endpoint:

In vivo - Cell differentiation response - measured by concentration of Cathepsin_3day

3.3.Comment on endpoint:

The library of materials used by Wang et al., was injected into SKH1-E mice and monitored for 7 days through in vivo fluorescence imaging. Histological analysis and ex vivo assays were performed on tissues collected 7 days post-injection, when the highest-level of MF presence is expected.

The different evaluated endpoints are related with the differentiation of macrophages in M1 or M2. Cathepsin activity measured at 7 days was negatively correlated with TNF- α and positively correlated with arginase:iNOS, demonstrating that cathepsin is positively correlated with increased M2 MFs and negatively correlated with M1 MFs.

High levels of IL-10 are commonly associated with M2 MFs and tumour progression, and high levels of TNF- α are commonly associated with M1 MFs and tumour suppression

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- N_(1°C): Number of 1⁰ carbon atoms
 - (1)X: Connectivity index 1
 - N: Number of non-hydrogen atoms
 - N_C: Number of carbon atoms
 - WCA: Water contact angle
 - (1) ξ^v : Intensive connectivity index 2
 - (0) ξ : Intensive atomic index 1
- ; 0

4.4.Descriptor selection:

Original descriptors were selected from Bicerano J. Prediction of polymer properties. CRC Press; 2002.

Within the model building algorithm PLS the most contributing descriptors were selected for the final models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

14/0

Descriptor: Chemical ratio :7:14 ~ 1:2

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Should be considered p(NIPAm-co-AAc) with other functionalizations that fall in the range of applied descriptors.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

14 Polymeric

List: p(NIPAm-co-AAc)

Shape: NA

Coating: 3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4-aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic and ethylene ketal
 2-aminoethylmethylsulfone hydrochloride
 3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)
 Aminomethylphosphonic acid (Alfa Aesar, Ward Hill, MA)
 and without functionalization

Size (nm): 600

Other info: (N-isopropylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

6.6.Pre-processing of data before modelling:

No splitting data is reported, but cross-validation was applied. NA in "NPs used as test set" was defined because there is no specification about if it was an LOO or LMO (the number of folds was not specified).

6.7.Statistics for goodness-of-fit:

$R^2Y = 0.826$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2Y = 0.723$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MPolymeric

List

p(NIPAm-co-AAc)

Shape:NA

Coating:3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4- aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic and ethylene ketal

2-aminoethylmethylsulfone hydrochloride

3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)

Aminomethylphosphonic acid (Alfa Aesar,Ward Hill, MA)

and without functionalization

Size(nm): 600

Other properties:

(N-isopropylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

No external validation is provided, it is commented as future work, with more data. We should take into account the possible overfitting.

ANOVA test was performed to confirm the statistical significance with $p < 0.05$.

There is a Mechanistic Interpretation

They are not nanomaterials according to EU definition

PLS: Partial least squares

R^2Y : correlation coefficient

Q^2Y : cross-validation correlation coefficient

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Macrophage

on 6-week-old female SKH-1 mice, QSAR, - $N_{(1^0C)}$: Number of 1^0 carbon atoms

- $(1)X$: Connectivity index 1

- N: Number of non-hydrogen atoms

- N_C : Number of carbon atoms


- WCA: Water contact angle

- $(1)\xi^v$: Intensive connectivity index 2

- $(0)\xi$: Intensive atomic index 1

, PLS: Partial Least Squares

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Determine the impact of surface modifications on MF phenotype by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Determine the impact of surface modifications on MF phenotype by examining the effect of surface functionalities of materials on cytokine profiles and in vivo imaging results by PLS
(Cathepsin_7day case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

K.M. Bratlie

kbratlie@iastate.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Bygd, H. C., Forsmark, K. D., & Bratlie, K. M. (2015). Altering in vivo macrophage responses with modified polymer properties. *Biomaterials*, 56, 187–197.

(Cathepsin_7day case)

<http://doi.org/10.1016/j.biomaterials.2015.03.042>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Macrophage

on 6-week-old female SKH-1 mice

3.2. Endpoint:

In vivo - Cell differentiation response - measured by concentration of Cathepsin_7day

3.3.Comment on endpoint:

The library of materials used by Wang et al., was injected into SKH1-E mice and monitored for 7 days through in vivo fluorescence imaging. Histological analysis and ex vivo assays were performed on tissues collected 7 days post-injection, when the highest-level of MF presence is expected.

The different evaluated endpoints are related with the differentiation of macrophages in M1 or M2. Cathepsin activity measured at 7 days was negatively correlated with TNF- α and positively correlated with arginase:iNOS, demonstrating that cathepsin is positively correlated with increased M2 MFs and negatively correlated with M1 MFs.

High levels of IL-10 are commonly associated with M2 MFs and tumour progression, and high levels of TNF- α are commonly associated with M1 MFs and tumour suppression

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- N_(1°C): Number of 1° carbon atoms
- (0)X^v: Atomic index 2
- N_O: Number of oxygen atoms
- N_H: Number of hydrogen atoms
- N_CH2: Number of CH2
- N_group
- WCA: Water contact angle; 7

4.4.Descriptor selection:

Original descriptors were selected from Bicerano J. Prediction of polymer properties. CRC Press; 2002.

Within the model building algorithm PLS the most contributing descriptors were selected for the final models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

14/7

Descriptor: Chemical ratio :7:14 ~ 1:2

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Should be considered p(NIPAm-co-AAc) with other functionalizations that fall in the range of applied descriptors.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

14 Polymeric

List: p(NIPAm-co-AAc)

Shape: NA

Coating: 3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4-aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic and ethylene ketal

2-aminoethylmethylsulfone hydrochloride

3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)

Aminomethylphosphonic acid (Alfa Aesar, Ward Hill, MA)

and without functionalization

Size (nm): 600

Other info: (N-isopropylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

6.6.Pre-processing of data before modelling:

No splitting data is reported, but cross-validation was applied. NA in "NPs used as test set" was defined because there is no specification about if it was an LOO or LMO (the number of folds was not specified).

6.7.Statistics for goodness-of-fit:

$R^2Y = 0.852$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2Y = 0.647$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MPolymeric

List

p(NIPAm-co-AAc)

Shape:NA

Coating:3-butenylamine (Santa Cruz Biotechnology, Dallas, TX)

1,4-dioxan-2-ylmethanamine

glycidamide

4-amino-3-penten-2-one

malonamide (Fisher, Pittsburgh, PA)

tert-butyl 4- aminobutanoate (VWR, Radnor, PA)

aminoacetaldehyde dimethyl acetal (Alfa Aesar, Ward Hill, MA)

3-aminobenzamide oxime

2,4-dinitro-phenyl-hydroxylamine

1-amino-4-oxocyclohexanecarboxylic and ethylene ketal

2-aminoethylmethylsulfone hydrochloride

3-amino-1-propanesulfonic acid (Fisher, Pittsburgh, PA)

Aminomethylphosphonic acid (Alfa Aesar,Ward Hill, MA)

and without functionalization

Size(nm): 600

Other properties:

(N-isopropylacrylamide-co-acrylic acid) = p(NIPAm-co-AAc)

To ensure successful modification, NMR spectra were collected for each modified particle using a Bruker Avance III Spectrometer. Particles were then examined with scanning electron microscopy using a FEI Quanta 250 FE-SEM. Water contact angles (WCA) were measured for each p(NIPAm-co-AAc) modified particle using the captive bubble technique. The zeta potential of each material was found using a Zetasizer Nano Z (Malvern, Westborough, MA). Transition temperatures were determined from differential scanning calorimetry measurements taken with a Perkin 1model (PerkinElmer Inc., Waltham, MA). Lastly, alternative activation of complement was assessed using the ASTM protocol F2065.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

No external validation is provided, it is commented as future work, with more data. We should take into account the possible overfitting.

ANOVA test was performed to confirm the statistical significance with $p < 0.05$.

There is a Mechanistic Interpretation

They are not nanomaterials according to EU definition

PLS: Partial least squares

R^2_Y : correlation coefficient

Q^2_Y : cross-validation correlation coefficient

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Macrophage

on 6-week-old female SKH-1 mice, QSAR, - $N_{(1^\circ C)}$: Number of 1^o carbon atoms

- $(0)X^v$: Atomic index 2

- N_O : Number of oxygen atoms


- N_H : Number of hydrogen atoms

- N_{CH2} : Number of CH₂

- N_{group}

- WCA: Water contact angle, PLS: Partial Least Squares

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal Oxide NPs toxicity towards BEAS-2B and RAW 264.7 by RF and
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal Oxide NPs toxicity towards BEAS-2B and RAW 264.7 by RF and causal inference interpretation
(BEAS-2B case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tomasz Puzyn

t.puzyn@qsar.eu.org

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Sizochenko, N., Rasulev, B., Gajewicz, A., Mokshyna, E., Kuz'min, V. E., Leszczynski, J., & Puzyn, T. (2015). Causal inference methods to assist in mechanistic interpretation of classification nano-SAR models. RSC Advances, 5(95), 77739–77745.

(BEAS-2B

<http://doi.org/10.1039/c5ra11399g>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human bronchial epithelial cells (BEAS-2B)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as binary classification in toxic("1") or non-toxic ("0") transformed from log(EC50)

3.3.Comment on endpoint:

Originally measured data were characterized as slopes of the dose-response based on doses of nanoparticles (in µg/mL) and responses (the decimal logarithm of EC50). In current study the final endpoints' data set was transformed to binary rank scale: slopes with negative values were marked as "0", while positive slopes were marked as "1". Thus, "0" means non-toxic compound and "1" means toxic compound in given conditions.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RF: Random Forest

4.3.Descriptors in the model:

- Mass density
- Covalent index
- Cation polarizing power
- Wigner-Seitz radius
- Surface-area-to-volume ratio
- Aggregation parameter
- Tri-atomic descriptor of atomic charges; 7

4.4.Descriptor selection:

Descriptors calculated within this contribution can be divided into four groups:

- Simplex Representations of Molecular Structure (SiRMS)- based descriptors
- Metal-ligand binding descriptors
- "Liquid drop" model (LDM) derived descriptors
- Integral (constitutional) descriptors for each molecule, such as molecular weight, mass density and aligned electronegativity of oxide

Explained detailed descriptors at supplementary material.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

18/7

Descriptor: Chemical ratio :7:18 ~ 1:3

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Domain applicability (DA) was measured based on minimum-cost-tree of variable importance values in space of descriptors considering their relative importance

Reference:

P. G. Polishchuk, E. N. Muratov, A. G. Artemenko, O. G. Kolumbin, N. N. Muratov and V. E. Kuz'min, J. Chem. Inf. Model., 2009, 49, 2481–2488

No outliers were detected.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

18 Metal Oxide

List: Al₂O₃

CuO

CeO₂

Co₃O₄

CoO
 Cr₂O₃
 Fe₂O₃
 Fe₃O₄
 Gd₂O₃
 HfO₂
 In₂O₃
 La₂O₃
 Mn₂O₃
 NiO
 Ni₂O₃
 Sb₂O₃
 SiO₂
 SnO₂
 R-TiO₂
 WO₃
 Y₂O₃
 Yb₂O₃
 ZnO
 ZrO₂

Shape: NA

Coating: NA

Size (nm): 10-100

Other info: exceptions outside range of sizes:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

6.6.Pre-processing of data before modelling:

Nanoparticles were splitted into training and test set (18 and

6 compounds, respectively) in the following way - the splitting of the dataset to training and test sets fulfilled three conditions:

- (1) metal oxides from each activity group should be presented in both training and test sets;
- (2) metal oxides presented in the test set should cover all types of oxides (MeO, Me₂O₃, MeO₂), similarly to the training set;
- (3) the list of oxides in each test set should be identical for both toxicity endpoints.

6.7. Statistics for goodness-of-fit:

Sensitivity = 100.00 %

Specificity = 100.00%

Balanced Accuracy = 100.00 %

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

After calculating all the descriptors, variables having zero-variance, and highly cross-correlating variables (with the Pearson's pair correlation coefficient $|r| > 0.9$) were eliminated.

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

6 Metal Oxide

List

Al₂O₃

CuO

CeO₂

Co₃O₄

CoO

Cr₂O₃

Fe₂O₃

Fe₃O₄

Gd₂O₃

HfO₂

In₂O₃

La₂O₃

Mn₂O₃

NiO

Ni₂O₃

Sb₂O₃

SiO₂

SnO₂

R-TiO₂

WO₃

Y₂O₃

Yb₂O₃

ZnO

ZrO₂

Shape:NA

Coating:NA

Size(nm): 10-100

Other properties:

exceptions outside range of sizes:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis

spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 100.00 %

Specificity = 66.00%

Balanced Accuracy = 83.00 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

They implement an additional criteria for quality evaluation and also a tool to elucidate underlying structure of nanotoxicity (Mechanistic Interpretation), which was the development of causal structures. For specific details see the publication's Figure 4.

The conclusion was the following: In fact, there is no direct link between toxicity and any descriptor. It means there are only particular causal links and the developed models are the collection of the most important descriptors, which only represent the conditions for the emergence of particular cause of action.

Correlation between target properties was evaluated via ϕ -coefficient. $S_\phi = 0.51$ which means that there is an average degree of association between two types of toxicity.

There is a Mechanistic Interpretation

NA

9.2.Bibliography:

(already reported in this table)

Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., Liu, R., ... Nel, A. E. (2012). Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano*, 6(5), 4349–4368

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Human bronchial epithelial cells (BEAS-2B), QSAR, - Mass density

- Covalent index
- Cation polarizing power
- Wigner-Seitz radius
- Surface-area-to-volume ratio
- Aggregation parameter
- Tri-atomic descriptor of atomic charges,RF: Random Forest

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal Oxide NPs toxicity towards BEAS-2B and RAW 264.7 by RF and
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal Oxide NPs toxicity towards BEAS-2B and RAW 264.7 by RF and causal inference interpretation
(RAW264.7 case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tomasz Puzyn

t.puzyn@qsar.eu.org

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Sizochenko, N., Rasulev, B., Gajewicz, A., Mokshyna, E., Kuz'min, V. E., Leszczynski, J., & Puzyn, T. (2015). Causal inference methods to assist in mechanistic interpretation of classification nano-SAR models. RSC Advances, 5(95), 77739–77745.

(RAW264.7

<http://doi.org/10.1039/c5ra11399g>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Rat alveolar macrophage cells (RAW264.7)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as binary classification in toxic("1") or non-toxic ("0") transformed from log(EC50)

3.3.Comment on endpoint:

Originally measured data were characterized as slopes of the dose-response based on doses of nanoparticles (in µg/mL) and responses (the decimal logarithm of EC50). In current study the final endpoints' data set was transformed to binary rank scale: slopes with negative values were marked as "0", while positive slopes were marked as "1". Thus, "0" means non-toxic compound and "1" means toxic compound in given conditions.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RF: Random Forest

4.3.Descriptors in the model:

- Molecular weight
- Aligned electronegativity
- Covalent index
- Surface area
- Surface-area-to-volume ratio
- Two-atomic descriptor of van-der-Waals interactions
- Tetra-atomic descriptor of atomic charges
- Size in DMEM; 9

4.4.Descriptor selection:

Descriptors calculated within this contribution can be divided into four groups:

- Simplex Representations of Molecular Structure (SiRMS)- based descriptors
- Metal-ligand binding descriptors
- "Liquid drop" model (LDM) derived descriptors
- Integral (constitutional) descriptors for each molecule, such as molecular weight, mass density and aligned electronegativity of oxide

Explained detailed descriptors at supplementary material.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

18/9

Descriptor: Chemical ratio :9:18 ~ 1:2

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Domain applicability (DA) was measured based on minimum-cost-tree of variable importance values in space of descriptors considering their relative importance

Reference:

P. G. Polishchuk, E. N. Muratov, A. G. Artemenko, O. G. Kolumbin, N. N. Muratov and V. E. Kuz'min, J. Chem. Inf. Model., 2009, 49, 2481–2488

No outliers were detected.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

18 Metal Oxide

List: Al₂O₃

CuO

CeO₂

Co₃O₄
 CoO
 Cr₂O₃
 Fe₂O₃
 Fe₃O₄
 Gd₂O₃
 HfO₂
 In₂O₃
 La₂O₃
 Mn₂O₃
 NiO
 Ni₂O₃
 Sb₂O₃
 SiO₂
 SnO₂
 R-TiO₂
 WO₃
 Y₂O₃
 Yb₂O₃
 ZnO
 ZrO₂

Shape: NA

Coating: NA

Size (nm): 10-100

Other info: exceptions outside range of sizes:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

6.6.Pre-processing of data before modelling:

Nanoparticles were splitted into training and test set (18 and 6 compounds, respectively) in the following way - the splitting of the dataset to training and test sets

fulfilled three conditions:

- (1) metal oxides from each activity group should be presented in both training and test sets;
- (2) metal oxides presented in the test set should cover all types of oxides (MeO, Me₂O₃, MeO₂), similarly to the training set;
- (3) the list of oxides in each test set should be identical for both toxicity endpoints.

6.7. Statistics for goodness-of-fit:

Sensitivity = 88.00 %

Specificity = 100.00%

Balanced Accuracy = 94 %

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

After calculating all the descriptors, variables having zero-variance, and highly cross-correlating variables (with the Pearson's pair correlation coefficient $|r| > 0.9$) were eliminated.

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

6 Metal Oxide

List

Al₂O₃

CuO

CeO₂
 Co₃O₄
 CoO
 Cr₂O₃
 Fe₂O₃
 Fe₃O₄
 Gd₂O₃
 HfO₂
 In₂O₃
 La₂O₃
 Mn₂O₃
 NiO
 Ni₂O₃
 Sb₂O₃
 SiO₂
 SnO₂
 R-TiO₂
 WO₃
 Y₂O₃
 Yb₂O₃
 ZnO
 ZrO₂

Shape:NA

Coating:NA

Size(nm): 10-100

Other properties:

exceptions outside range of sizes:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 100.00 %

Specificity = 75.00%

Balanced Accuracy = 88.00 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

They implement an additional criteria for quality evaluation and also a tool to elucidate underlying structure of nanotoxicity (Mechanistic Interpretation), which was the development of causal structures. For specific details see the publication's Figure 5.

The conclusion was the following: In fact, there is no direct link between toxicity and any descriptor. It means there are only particular causal links and the developed models are the collection of the most important descriptors, which only represent the conditions for the emergence of particular cause of action.

Correlation between target properties was evaluated via ϕ -coefficient. $S_\phi = 0.51$ which means that there is an average degree of association between two types of toxicity.

NA

9.2.Bibliography:

(already reported in this table)

Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., Liu, R., ... Nel, A. E. (2012). Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano*, 6(5), 4349–4368

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Rat alveolar macrophage cells (RAW264.7), QSAR, - Molecular weight

- Aligned electronegativity
- Covalent index
- Surface area
- Surface-area-to-volume ratio
- Two-atomic descriptor of van-der-Waals interactions
- Tetra-atomic descriptor of atomic charges
- Size in DMEM,RF: Random Forest

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Toxicological modelling of diverse nanomaterials using the embryonic
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Toxicological modelling of diverse nanomaterials using the embryonic zebrafish (EZ) metric toxicity by multiple non-linear regression

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Stacey Harper

stacey.harper@oregonstate.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software

package:

Harper, B., Thomas, D., Chikkagoudar, S., Baker, N., Tang, K., Heredia-Langner, A., ... Harper, S. (2015). Comparative hazard analysis and toxicological modelling of diverse nanomaterials using the embryonic zebrafish (EZ) metric of toxicity. Journal of Nan

<http://doi.org/10.1007/s11051-015-3051-0>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Zebrafish (Danio rerio) embryos

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as Mod EZ Metric

3.3.Comment on endpoint:

Exposures' concentrations were typically fivefold serial dilutions of nanomaterials ranging from approximately 250 parts per million (ppm) down to *16 parts per billion (ppb) prepared in fishwater. Control exposures comprised fishwater alone (without NPs). Embryos were incubated at 26 °C under 14/10 light cycle and were evaluated visually at 24 hpf for viability, developmental progression, and spontaneous movement (earliest behaviour in zebrafish). At 120 hpf, behavioural endpoints (motility, tactile response) were thoroughly evaluated in vivo and larval morphology (body axis, eye, snout, jaw, otic vesicle, notochord, heart, brain, somite, fin, yolk sac, trunk, circulation, pigment, swim bladder) was evaluated visually and scored in a binary fashion (present or absent) (Harper et al., 2008a, b; Truong et al., 2011).

To summarize the 21 measured toxicity endpoints for each dose applied to the embryonic zebrafish, we define the EZ Metric score to provide a relative comparison of nanomaterial-elicited effects. Original EZ Metric responses were transformed by multiplying the score by 100 and adding 0.1 to avoid the discontinuity resulting from taking the natural logarithm of zero scores (denoted as Mod EZ Metric in publication's equation "Eq. 3")

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Multiple non-linear regression

4.3.Descriptors in the model:

- log(Conc): represents the natural logarithm of the NP concentration used in the tests
- Size: is the primary NP diameter in nanometers
- SASA/Polar : where SASA is the Solvent-Accessible Surface Area (\AA^2), and Polar represents the surface area formed by all the polar atoms of a molecule (\AA^2)
- Refractivity: the molar refractivity, a measure of the volume occupied by an atom or functional group (m^3/mol)
- Band Gap: the energy difference between the highest occupied molecular orbital (HOMO) and the lowest occupied molecular orbital (LUMO)
- $\log(\text{Conc}) \times \text{SASA/Polar}$
- $\text{Size} \times \text{Band Gap}$
- $(\log(\text{Conc}))^2$
- $\text{Size}^2; 9$

4.4.Descriptor selection:

The surface functional group chemical characteristics were built using the extensible computational chemistry environmental program (Black et al., 2003). The compounds were geometry optimized at the Hartree–Fock/ 6-31G* level of theory using the NWChem 5.1 program (Bylaska et al., 2007; Kendall et al., 2000) and the band gaps calculated. The remaining topographical and physicochemical molecular descriptors were calculated using the Cerius2/Discovery Studio program (Accelrys 2006).

4.5.Algorithm and descriptor generation:

No information available

4.6. Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

16/9

Descriptor: Chemical ratio :9:16 ~1:2

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Should be considered Au NPs with other functionalizations that fall in the range of applied descriptors.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

16 Metal

List: Au

Shape: NA

Coating: TMAT: N,N,N-trimethylammonium-mercaptoethanol

MEE: 2-(2-mercaptoethoxy)ethanol

MEEE: 2,2,2-[mercaptoethoxy(ethoxy)]ethanol

MES: 2-mercaptoethanesulfonate

Size (nm): 0.8, 2 and 10

Other info: Nanomaterials were acquired from a variety of commercial sources including Sigma Aldrich (St. Louis, MO, USA), Nanocomposix (San Diego, CA, USA), Dendritic Technologies (San Francisco, CA USA), and non-commercial research labs. Details of nanomaterial manufacturers and material composition are available in Online Resource 1, online at nbi.oregonstate.edu and in previous publications on selected materials (Harper et al., 2007, 2008a; Pryor et al., 2014; Usenko et al., 2007, 2008)

6.6.Pre-processing of data before modelling:

340 data points generated from 68 NPs x fivefold dilutions.

After a cluster analysis and a classification analysis, with the obtained information they decide to focus their final model in a subset of four Au nanoparticles that varied in surface chemistry and size. 16 NPs

6.7.Statistics for goodness-of-fit:

$R^2 = 0.88$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal

List

Au

Shape:NACoating:TMAT: N,N,N-trimethylammonium-umethanethiol

MEE: 2-(2-mercaptoethoxy)ethanol

MEEE: 2,2,2-[mercaptoethoxy(ethoxy)]ethanol

MES: 2-mercaptoethanesulfonate

Size(nm): 0.8, 2 and 10Other properties:

Nanomaterials were acquired from a variety of commercial sources including Sigma Aldrich (St. Louis, MO, USA), Nanocomposix (San Diego, CA, USA), Dendritic Technologies (San Francisco, CA USA), and non-commercial research labs. Details of nanomaterial manufacturers and material composition are available in Online Resource 1, online at nbi.oregonstate.edu and in previous publications on selected materials (Harper et al., 2007, 2008a; Pryor et al., 2014; Usenko et al., 2007, 2008)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

With whole data set:

Spearman rank correlation by SigmaPlot v12.0

Hierarchical clustering algorithm with Euclidean distance measure

and Wardlinkage rule by MATLAB

Classification decision tree C4.5 with ten-fold cross-validation using WEKA software (v3.6.3) Decision tree (J48) (94% of accuracy)

68 different NPs were analyzed and studied, which can be seen at Online resource 1. The 4 types of Au NPs were selected after a clustering and classification analysis to be involved in the final model.

Information about the classification decision tree was not clear (mistake at publication: the dendrogram was not published; and confusing cross-validation information), hence we decide to not classify it.

There is not any robustness nor external validation. Then, an overfitting could be present, especially if we check the plot of the predicted vs. the measured EZ Metric (Online resource 3), we could see an agglomeration of data close to 0, which could be a noise data and give us a fake correlation idea.

NPs: Nanoparticles

R²: Correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Organism, Zebrafish (Danio rerio) embryos, QSAR, - log(Conc): represents the natural logarithm of the NP concentration used in the tests

- Size: is the primary NP diameter in nanometers
- SASA/Polar : where SASA is the Solvent-Accessible Surface Area (Å²), and Polar represents the surface area formed by all the polar atoms of a molecule (Å²)
- Refractivity: the molar refractivity, a measure of the volume occupied by an atom or functional group (m³/mol)
- Band Gap: the energy difference between the highest occupied molecular orbital (HOMO) and the lowest occupied molecular orbital (LUMO)
- log(Conc) x SASA/Polar
- Size x Band Gap
- (log(Conc))²
- Size², Multiple non-linear regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction of multiple antibacterial profiles of nanoparticles by a
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction of multiple antibacterial profiles of nanoparticles by a perturbational approach and LDA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A. Speck-Planche

alejspivanovich@gmail.com

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Speck-Planche, A., Kleandrova, V. V., Luan, F., & Ds Cordeiro, M. N. (2015). Computational modelling in Nanomedicine: Prediction of multiple antibacterial profiles of nanoparticles using a quantitative structure-activity relationship perturbation model. Na <http://doi.org/10.2217/nnm.14.96>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

34 different Bacteria

3.2. Endpoint:

In vitro - Cytotoxicity - measured as binary classification into antibacterial active ("1") or antibacterial inactive ("-1") class which was obtained by different antibacterial activities (MBC, MIC, Microb-Eff)

3.3. Comment on endpoint:

The NPs were tested against different bacteria, by considering several measures of antibacterial activity and diverse assay times.

Diverse measures of antibacterial activity (m_e)

MBC: Minimum bactericidal concentration

MIC: Minimum inhibitory concentration

Microb-Eff: Microbicidal effect

300 NPs/cases were assigned to 1 out of 2 possible groups related to their antibacterial activities in a specific condition c_j [$AB_i(c_j)$]. In this sense, $AB_i(c_j)$ is a binary classification variable that accounts for the antibacterial activity of the NPs. Thus, the NPs/cases were considered as active [$AB_i(c_j) = 1$], when they exhibited high antibacterial activities (low levels of growth inhibition); otherwise, they were selected as inactive [$AB_i(c_j) = -1$]

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

LDA applied to the perturbation approach obtained equation.

by software STATISTICA 6.0

The final model will be a consensus prediction, after apply the desired NP to the 300 NPs used as reference NP.

4.3.Descriptors in the model:

$AB_i(c_j)_{rf}$: Binary function of the antibacterial activity of the nanoparticle used in the reference (initial) state

$\Delta\Delta AE(m_e)$: Perturbation term accounting for the changes in the electronegativity between the reference (initial) and output (final) nanoparticles, and also depending on the measures of the antibacterial activities

$\Delta\Delta AP(b_t)$: Perturbation term accounting for the changes in the polarizability between the reference (initial) and output (final) nanoparticles, and also depending on the types of bacteria (strain included) on which the assays were performed

$\Delta\Delta aps(b_t)$: Perturbation term accounting for the changes in particle size between the reference (initial) and output (final) nanoparticles, and also depending on the types of bacteria (strain included) on which the assays were performed

$\Delta\Delta AMV(n_{sh})$: Perturbation term accounting for the changes in the molar volume between the reference (initial) and output (final) nanoparticles, and also depending on the shapes of the nanoparticles

$\Delta\Delta aps(t_e)$: Perturbation term accounting for the changes in particle size between the reference (initial) and output (final) nanoparticles, and also depending on the times during which the bacteria were exposed to the nanoparticles

$(\Delta SM_2)^{Std}(s_c)$: Perturbation term focused on the spectral moment of order 2 weighted by the standard bond distance, which depends on changes in the chemical structures of the coating agents used in the reference (initial) and output (final) states future; 7

4.4.Descriptor selection:

Physicochemical properties (PPs) were retrieved from the website Chemicool Periodic Table. For the specific case of those NPs with chemical compositions involving more than one element, these properties were normalized.

The fourth descriptor was the NP size (aps), which was expressed in nanometers, being measured experimentally.

Descriptors characterizing the chemical structures of the coating agents were computed by MODESLAB 1.5

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

51979/7

Descriptor: Chemical ratio :7:300 ~ 1:43

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal and metal oxide NPs within the range of parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

51979 Metal

Metal Oxide

List: CdS

CdO

ZnO

CuO

La₂O₃

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

Au

CuI

Ag

Cu

Fe₃O₄

SiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Core NPs were synthesized against different experimental conditions (exposure time, coatings, bacteria, etc), different sizes and shapes

The raw data were retrieved from the public source OChem. In addition, other complementary data were also extracted from the literature.

6.6.Pre-processing of data before modelling:

The dataset containing 69,231 NPs/cases (pairs NP-NP from 300 NPs) was randomly divided into two series: training and prediction (validation or test) sets. The training set was used to create the QSAR perturbation model, comprising 51,979 cases, with 28,715 of them considered as active and 23,264 inactive. The prediction (validation or test) set contained 17,252 cases, of which 9439 were active and 7813 inactive.

Predicted this Cu–Ag nanoalloy 20,400 times as a result of considering the two different measures of antibacterial activities (MIC and MBC) that were experimentally reported, as well as the two possible sizes (30 and 50 nm), the 17 bacterial strains (belonging to *S. aureus* and *E. coli*) that were used and the 300 (original) cases employed as reference NPs

6.7.Statistics for goodness-of-fit:

Training set:

- Accuracy = 98.01 %
- Sensitivity = 98.21 %
- Specificity = 97.76 %
- ROC_AUC = 0.999

Test set:

- Accuracy = 98.04 %
- Sensitivity = 98.42 %
- Specificity = 97.58%

- ROC_AUC = 0.999

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

1 M Metal

Metal Oxide

List

CdS

CdO

ZnO

CuO

La₂O₃

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

Au

CuI

Ag

Cu

Fe₃O₄

SiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Core NPs were synthesized against different experimental conditions (exposure time, coatings, bacteria, etc), different sizes and shapes

The raw data were retrieved from the public source OChem. In addition, other complementary data were also extracted from the literature.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Cu-Ag nanoalloy:

- Accuracy = 77.73 %

Specific detail predictions due specific antibacterial measures or sizes:

- MIC Accuracy = 79.75 %

- MBC Accuracy = 75.71 %

- 30 nm Accuracy = 83.33 %

- 50 nm Accuracy = 72.12 %

For different bacteria:

Accura

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The methodology is related with the previous reported publication in the table (Kleandrova, V. V, Luan, F., González-Díaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., & Cordeiro, M. N. D. S. (2014).

Computational ecotoxicology: Simultaneous prediction of ecotoxic

effects of nanoparticles under different experimental conditions.

Environment International, 73, 288–294.)

Physicochemical interpretations of the descriptors were provided

NPs: Nanoparticles

MBC: Minimum bactericidal concentration

MIC: Minimum inhibitory concentration

Microb-Eff: Microbicidal effect

ROC_AUC: Area under the curve of receiving operating characteristic curve

9.2. Bibliography:

The raw data were retrieved from the public source OChem:

- Sushko I, Novotarskyi S, Korner R et al., Online chemical modelling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J. Comput. Aided Mol. Des. 25(6), 533–554 (2011).

- OChem: online chemical database with modelling environment.

<https://ochem.eu/home/show.do>

In addition, other complementary data were also extracted from the literature:

- From references 19 to 29 in the publication.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, 34 different Bacteria, QSAR, AB_i(c_j)_{rf}: Binary function of the antibacterial activity of the nanoparticle used in the reference (initial) state

$\Delta\Delta E(m_e)$: Perturbation term accounting for the changes in the electronegativity between the reference (initial) and output (final) nanoparticles, and also depending on the measures of the antibacterial activities

$\Delta\Delta AP(b_t)$: Perturbation term accounting for the changes in the polarizability between the reference (initial) and output (final) nanoparticles, and also depending on the types of bacteria (strain included) on which the assays were performed

$\Delta\Delta aps(b_t)$: Perturbation term accounting for the changes in particle size between the reference (initial) and output (final) nanoparticles, and also depending on the types of bacteria (strain included) on which the assays were performed

$\Delta\Delta AMV(n_{sh})$: Perturbation term accounting for the changes in the molar volume between the reference (initial) and output (final) nanoparticles, and also depending on the shapes of the nanoparticles

$\Delta\Delta aps(t_e)$: Perturbation term accounting for the changes in particle size between the reference (initial) and output (final) nanoparticles, and also depending on the times during which the bacteria were exposed to the nanoparticles


$(ASM_2)^{Std}(s_c)$: Perturbation term focused on the spectral moment of order 2 weighted by the standard bond distance, which depends on changes in the chemical structures of the coating agents

used in the reference (initial) and output (final) states future, LDA applied to the perturbation approach obtained equation.

by software STATISTICA 6.0

The final model will be a consensus prediction, after apply the desired NP to the 300 NPs used as reference NP.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Toxicity model of engineered zinc oxide nanoparticles to embryonic
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Toxicity model of engineered zinc oxide nanoparticles to embryonic zebrafish by PCA and Kriging estimation

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Stacey Harper

stacey.harper@oregonstate.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Zhou, Z., Son, J., Harper, B., Zhou, Z., & Harper, S. (2015).

Influence of surface chemical properties on the toxicity of engineered zinc oxide nanoparticles to embryonic zebrafish.

Beilstein Journal of Nanotechnology, 6(1), 1568–1579.

<http://doi.org/10.3762/bjnano.6.160>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Embryonic zebrafish

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as percentage of Mortality

3.3. Comment on endpoint:

Wild-type 5D zebrafish (*Danio rerio*) embryos were obtained from group spawns of adult fish housed at the Sinnhuber Aquatic Research Laboratory at Oregon State University (Corvallis, OR). All NP dilutions and exposures were conducted in fish water (FW). The FW was prepared with 0.26 g/L Instant Ocean salts (Aquatic Ecosystem, Apopka, FL) combined with approximately 0.01g NaHCO₃ pH buffer in reverse osmosis water (pH 7.0–7.4, conductivity 450–600 µS)

Despite assessing a suite of 19 different developmental, behavioural and morphological endpoints in addition to mortality in this study, mortality was the most common endpoint observed for all of the ZnO NP types tested

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

The ordinary kriging was conducted in R using the additional “Kriging” and “gstat” packages.

The PCs vs mortality were applied in the Kriging estimation.

4.3.Descriptors in the model:

- PC1 (0.735 of explained variance) : First Principal Component (8 descriptors all have moderately similar weights)

- PC2 (0.138 of explained variance) : Second Principal Component (LogD, PS, and SZ have outstanding weights)

8 descriptors used in the PCA:

-size (SZ)

-LogD: distribution coefficient

-polarizability (PL)

-polar surface area (PS)

-van der Waals surface (VS)

-solvent-accessible surface area (SASA)

-molar refractivity (RF)

-Dreiding energy (DE)

; 0

4.4.Descriptor selection:

Fisher's exact test (Sigma Plot v12.0, San Jose, CA) was used to analyze individual endpoints recorded at 24 and 120 hpf. P-value was calculated based on two-tailed test and a $p \leq 0.05$ significance level was maintained for all analyses. Mortality data was compared between NPs with the same capping agent but different sizes using two-way analysis of variance (R v3.1.0)

Principal component analysis (PCA) was conducted in R using the primary particle size and seven intrinsic properties of NPs' surface chemistry.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :2:17 ~1:8

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of ZnO NPs within the range of parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: ZnOShape: NACoating: uncoated

oleic acid

octanoic acid

para-nitrobenzoic acid

cyclohexanecarboxylic acid

benzoic acid

Size (nm): 4-70

Other info: The ZnO NPs with different capping agents and sizes were obtained from a variety of commercial and research laboratories (Table 1 in the publication). More detailed characterization of the nanomaterials are also available on the open-source Nanomaterial-Biological Interactions Knowledgebase provided by Oregon State University

6.6.Pre-processing of data before modelling:

137 rows became from 17 NPs at 8 different concentrations. The study was developed for each of the concentrations (from 0.016 to 250 mg/L)

6.7.Statistics for goodness-of-fit:

$R^2 = 0.702$ (for 250mg/L)

$R^2 = 0.702-0.778$

(range for the different concentrations)

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA Metal Oxide

List

ZnO

Shape:NA

Coating:uncoated

oleic acid

octanoic acid

para-nitrobenzoic acid

cyclohexanecarboxylic acid

benzoic acid

Size(nm): 4-70

Other properties:

The ZnO NPs with different capping agents and sizes were obtained from a variety of commercial and research laboratories (Table 1 in the publication). More detailed characterization of the nanomaterials are also available on the open-source Nanomaterial-Biological Interactions Knowledgebase provided by Oregon State University

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

External validation must be required to ensure the predictivity of the model. Neither there is any robustness evaluation technique.

Even the lack in the reliability of the model, a new methodology is presented, which had not been applied in the previous classified papers.

A brief mechanistic interpretation of the descriptors is presented.

NPs: Nanoparticles

PCA: Principal Component Analysis

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Organism, Embryonic zebrafish, QSAR, - PC1 (0.735 of explained variance) : First Principal Component (8 descriptors all have moderately similar weights)

- PC2 (0.138 of explained variance) : Second Principal Component (LogD, PS, and SZ have outstanding weights)

8 descriptors used in the PCA:

-size (SZ)

-LogD: distribution coefficient

-polarizability (PL)

-polar surface area (PS)

-van der Waals surface (VS)

-solvent-accessible surface area (SASA)


-molar refractivity (RF)

-Dreiding energy (DE)

,The ordinary kriging was conducted in R using the additional “Kriging” and “gstat” packages.

The PCs vs mortality were applied in the Kriging estimation.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by linear regression method (MLR case for TiO₂ and ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(MLR case for TiO₂ and ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression
by QSARINS software

4.3.Descriptors in the model:

- X4: Concentration
- X0: Engineered size
- X2: Size in PBS; 3

4.4.Descriptor selection:

'The variable selection' procedure generates a 'population' of models within the building model (MLR), ranked according to decreasing r^2 values. The best models were chosen by using Q^2 leave-one-out (Q^2_{loo}) as the optimization value, and taking into account the parsimony principle regarding the complexity of the models, which should be as small as possible. For this reason, only up to three descriptors were included in the QSARs generated in this study.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

31/3

Descriptor: Chemical ratio :3:31 ~ 1:10

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 2)

$h^* = 0.387$

Any outlier was detected

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the performance of the models, if we compare it with the previous study (data source)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

31 Metal Oxide

List: TiO₂

ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 42 to 31.

Not splitting applied to the final model. Ten different random splitting of 20 % out, and one 50% out were generated in order to identify the most predictive and stable model among the possible combinations of available descriptors.

6.7.Statistics for goodness-of-fit:

$r^2 = 0.82$

RMSE = 0.10

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LOO} = 0.74$

RMSE_{LOO} = 0.12

$Q^2_{MLO(30\%)} = 0.76$

$r^2_{YS} = 0.10$

From 11 random splits the range of parameters:

$r^2_{tr} = 0.78-0.85$

RMSE_{tr} = 0.09-0.11

$r^2_{ext} = 0.65-0.98$

RMSE_{ext} = 0.08-0.17

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA M Metal Oxide

List

TiO₂

ZnO

Shape:NA

Coating:NA

Size(nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that "the equation for the best model among the split models generated for each training set was newly calibrated on all of the available data for each dataset (full models)." lets the final model without a proper external validation, which should be taken into account. Then, the "external" validation statistics were classified in the Robustness column.

NPs: Nanoparticles

PCA: Principal Component Analysis

MLR : Multiple Linear Regression

r^2 : correlation coefficient

r^2_Y : correlation coefficient for Y-scrambling evaluation

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-on

9.2.Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Immortalized rat L2 lung epithelial cells
and
rat lung alveolar macrophages, QSAR, - X4: Concentration
- X0: Engineered size
- X2: Size in PBS,MLR: Multiple Linear Regression
by QSARINS software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by TiO₂ NPs by linear
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by TiO₂ NPs by linear regression method
(MLR case for TiO₂)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(MLR case for TiO₂)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression
by QSARINS software

4.3.Descriptors in the model:

- X0: Engineered size
 - X4: Concentration
- ; 2

4.4.Descriptor selection:

'The variable selection' procedure generates a 'population' of models within the building model (MLR), ranked according to decreasing r^2 values. The best models were chosen by using Q^2 leave-one-out (Q^2_{loo}) as the optimization value, and taking into account the parsimony principle regarding the complexity of the models, which should be as small as possible. For this reason, only up to three descriptors were included in the QSARs generated in this study.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

22/2

Descriptor: Chemical ratio :2:22 ~ 1:11

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure S4)

$h^* = 0.409$

Any outlier was detected

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the performance of the models, if we compare it with the previous study (data source)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

22 Metal Oxide

List: TiO₂

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 24 to 22.

Not splitting applied to the final model. Ten different random splitting of 20 % out were generated in order to identify the most predictive and stable model among the possible combinations of available descriptors.

6.7.Statistics for goodness-of-fit:

$r^2 = 0.84$

RMSE = 0.11

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LOO} = 0.79$

RMSE_{LOO} = 0.13

$Q^2_{MLO(30\%)} = 0.78$

$r^2_{YS} = 0.10$

From 10 random splits the range of parameters:

$r^2_{tr} = 0.82-0.91$

RMSE_{tr} = 0.09-0.12

$r^2_{ext} = 0.32-0.98$

RMSE_{ext} = 0.09-0.19

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA Metal Oxide

List

TiO₂

Shape:NA

Coating:NA

Size(nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that "the equation for the best model among the split models generated for each training set was newly calibrated on all of the available data for each dataset (full models)." lets the final model without a proper external validation, which should be taken into account. Then, the "external" validation statistics were classified in the Robustness column.

NPs: Nanoparticles

PCA: Principal Component Analysis

MLR : Multiple Linear Regression

r^2 : correlation coefficient

r^2_Y : correlation coefficient for Y-scrambling evaluation

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-on

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Immortalized rat L2 lung epithelial cells
and


rat lung alveolar macrophages, QSAR, - X0: Engineered size

- X4: Concentration

,MLR: Multiple Linear Regression

by QSARINS software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO NPs by linear
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO NPs by linear regression method
(MLR case for ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(MLR case for ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression
by QSARINS software

4.3.Descriptors in the model:

- X1: Size in water
- X2: Size in PBS
- X4: Concentration; 3

4.4.Descriptor selection:

'The variable selection' procedure generates a 'population' of models within the building model (MLR), ranked according to decreasing r^2 values. The best models were chosen by using Q^2 leave-one-out (Q^2_{loo}) as the optimization value, and taking into account the parsimony principle regarding the complexity of the models, which should be as small as possible. For this reason, only up to three descriptors were included in the QSARs generated in this study.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

15/3

Descriptor: Chemical ratio :3:15 ~ 1:15

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure S4)

$h^* = 0.800$

Any outlier was detected

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the performance of the models, if we compare it with the previous study (data source)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

15 Metal Oxide

List: ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 18 to 15.

Not splitting applied to the final model. Ten different random splitting of 20 % out were generated in order to identify the most predictive and stable model among the possible combinations of available descriptors.

6.7.Statistics for goodness-of-fit:

$r^2 = 0.91$

RMSE = 0.07

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LOO} = 0.80$

RMSE_{LOO} = 0.10

$Q^2_{MLO(30\%)} = 0.76$

$r^2_{YS} = 0.22$

From 10 random splits the range of parameters:

$r^2_{tr} = 0.88-0.92$

RMSE_{tr} = 0.09-0.12

$r^2_{ext} = 0.45-0.99$

RMSE_{ext} = -1.44-0.15

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA Metal Oxide

List

ZnO

Shape:NA

Coating:NA

Size(nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that "the equation for the best model among the split models generated for each training set was newly calibrated on all of the available data for each dataset (full models)." lets the final model without a proper external validation, which should be taken into account. Then, the "external" validation statistics were classified in the Robustness column.

NPs: Nanoparticles

PCA: Principal Component Analysis

MLR : Multiple Linear Regression

r^2 : correlation coefficient

r^2_Y : correlation coefficient for Y-scrambling evaluation

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-on

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Immortalized rat L2 lung epithelial cells
and


rat lung alveolar macrophages, QSAR, - X1: Size in water

- X2: Size in PBS

- X4: Concentration, MLR: Multiple Linear Regression

by QSARINS software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by non-linear regression method (SVM-linear case for TiO₂ and ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(SVM-linear case for TiO₂ and ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM-linear: Support Vector Machine linear
by Caret package of the Cran-R software

4.3.Descriptors in the model:

- X4: Concentration
- X0: Engineered size
- X2: Size in PBS; 3

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

27/3

Descriptor: Chemical ratio :3:31 ~ 1:10

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 31 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

27 Metal Oxide

List: TiO₂

ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 42 to 31.

From initial 31 NPs ten different random splitting of 20 % out, and one 50% out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$r^2_{tr} = 0.77-0.85$

$RMSE_{tr} = 0.09-0.12$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$Q^2_{LOO} = 0.62-0.81$

$RMSE_{LOO} = 0.10-0.15$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

4 M Metal Oxide

List

TiO₂

ZnO

Shape: NA

Coating: NA

Size(nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$r^2_{\text{ext}} = 0.65-0.96$

RMSE_{ext} = 0.08-0.17

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

SVM: Support Vector Machines

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross-validation

9.2.Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X4: Concentration

- X0: Engineered size

- X2: Size in PBS,SVM-linear: Support Vector Machine linear
by Caret package of the Cran-R software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by TiO2 NPs by non-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by TiO2 NPs by non-linear regression method
(SVM-linear case for TiO2)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO2 and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(SVM-linear case for TiO2)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM-linear: Support Vector Machine linear
by Caret package of the Cran-R software

4.3.Descriptors in the model:

- X0: Engineered size
 - X4: Concentration
- ; 2

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

18/2

Descriptor: Chemical ratio :2:22 ~ 1:11

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 22 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

18 Metal Oxide

List: TiO₂

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 24 to 22.

From initial 22 NPs ten different random splitting of 20 % out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 10 splittings:

$r^2_{tr} = 0.78-0.90$

$RMSE_{tr} = 0.09-0.13$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 10 splittings:

$Q^2_{LOO} = 0.66-0.89$

$RMSE_{LOO} = 0.10-0.17$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

4 M Metal Oxide

List

TiO₂

Shape:NA

Coating:NA

Size(nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Range of statistics for 10 splittings:

$r^2_{\text{ext}} = 0.30-0.99$

RMSE_{ext} = 0.08-0.19

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

SVM: Support Vector Machines

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross-validation

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Immortalized rat L2 lung epithelial cells
and


rat lung alveolar macrophages, QSAR, - X0: Engineered size

- X4: Concentration

,SVM-linear: Support Vector Machine linear

by Caret package of the Cran-R software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO NPs by non-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO NPs by non-linear regression method
(SVM-linear case for ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(SVM-linear case for ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM-linear: Support Vector Machine linear
by Caret package of the Cran-R software

4.3.Descriptors in the model:

- X1: Size in water
- X2: Size in PBS
- X4: Concentration; 3

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

12/3

Descriptor: Chemical ratio :3:15 ~ 1:15

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 15 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

12 Metal Oxide

List: ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 18 to 15.

From initial 15 NPs ten different random splitting of 20 % out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 10 splittings:

$$r^2_{tr} = 0.85-0.97$$

$$RMSE_{tr} = 0.05-0.09$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 10 splittings:

$$Q^2_{LOO} = 0.44-0.92$$

$$RMSE_{LOO} = 0.08-0.18$$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

3 M Metal Oxide

List

ZnO

Shape:NA

Coating:NA

Size(nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Range of statistics for 10 splittings:

$r^2_{\text{ext}} = 0.33-1.00$

RMSE_{ext} = 0.04-0.20

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

SVM: Support Vector Machines

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross-validation

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X1: Size in water

- X2: Size in PBS

- X4: Concentration, SVM-linear: Support Vector Machine linear
by Caret package of the Cran-R software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by non-linear regression method (SVM-radial case for TiO₂ and ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(SVM-radial case for TiO₂ and ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM-radial: Support Vector Machine radial
by Caret package of the Cran-R software

4.3.Descriptors in the model:

- X4: Concentration
- X0: Engineered size
- X2: Size in PBS; 3

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

27/3

Descriptor: Chemical ratio :3:31 ~ 1:10

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 31 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

27 Metal Oxide

List: TiO₂

ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 42 to 31.

From initial 31 NPs ten different random splitting of 20 % out, and one 50% out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$r^2_{tr} = 0.84-0.99$

$RMSE_{tr} = 0.03-0.10$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$Q^2_{LOO} = 0.18-0.80$

$RMSE_{LOO} = 0.11-0.23$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

4 M Metal Oxide

List

TiO₂

ZnO

Shape: NA

Coating: NA

Size(nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$r^2_{\text{ext}} = 0.50-0.99$

RMSE_{ext} = 0.7-0.17

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

SVM: Support Vector Machines

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross-validation

9.2.Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X4: Concentration

- X0: Engineered size

- X2: Size in PBS,SVM-radial: Support Vector Machine radial
by Caret package of the Cran-R software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by TiO₂ NPs by non-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by TiO₂ NPs by non-linear regression method
(SVM-radial case for TiO₂)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(SVM-radial case for TiO₂)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM-radial: Support Vector Machine radial
by Caret package of the Cran-R software

4.3.Descriptors in the model:

- X0: Engineered size
 - X4: Concentration
- ; 2

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

18/2

Descriptor: Chemical ratio :2:22 ~ 1:11

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 22 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

18 Metal Oxide

List: TiO₂

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 24 to 22.

From initial 22 NPs ten different random splitting of 20 % out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 10 splittings:

$$r^2_{tr} = 0.95-0.99$$

$$RMSE_{tr} = 0.03-0.07$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 10 splittings:

$$Q^2_{LOO} = 0.73-0.85$$

$$RMSE_{LOO} = 0.08-0.16$$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

4 M Metal Oxide

List

TiO₂

Shape:NA

Coating:NA

Size(nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Range of statistics for 10 splittings :

$r^2_{ext} = 0.77-1.00$

RMSE_{ext} = 0.05-0.10

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

SVM: Support Vector Machines

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross-validation

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X0: Engineered size

- X4: Concentration

,SVM-radial: Support Vector Machine radial
by Caret package of the Cran-R software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO NPs by non-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO NPs by non-linear regression method
(SVM-radial case for ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(SVM-radial case for ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM-radial: Support Vector Machine radial
by Caret package of the Cran-R software

4.3.Descriptors in the model:

- X1: Size in water
- X2: Size in PBS
- X4: Concentration; 3

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

12/3

Descriptor: Chemical ratio :3:15 ~ 1:15

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 15 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

12 Metal Oxide

List: ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 18 to 15.

From initial 15 NPs ten different random splitting of 20 % out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 10 splittings:

$r^2_{tr} = 0.94-1.00$

$RMSE_{tr} = 0.02-0.06$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 10 splittings:

$Q^2_{LOO} = 0.29-0.94$

$RMSE_{LOO} = 0.06-0.20$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

3 M Metal Oxide

List

ZnO

Shape:NA

Coating:NA

Size(nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Range of statistics for 10 splittings :

$r^2_{\text{ext}} = 0.49-1.00$

$\text{RMSE}_{\text{ext}} = 0.06-0.24$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

SVM: Support Vector Machines

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross-validation

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X1: Size in water

- X2: Size in PBS

- X4: Concentration, SVM-radial: Support Vector Machine radial
by Caret package of the Cran-R software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by non-linear regression method (RBFNN case for TiO₂ and ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(RBFNN case for TiO₂ and ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RBNN: Radial Basis Function Neural Networks
by MATLAB

4.3.Descriptors in the model:

- X4: Concentration
- X0: Engineered size
- X2: Size in PBS; 3

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

27/3

Descriptor: Chemical ratio :3:31 ~ 1:10

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 31 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

27 Metal Oxide

List: TiO₂

ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 42 to 31.

From initial 31 NPs ten different random splitting of 20 % out, and one 50% out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$r^2_{tr} = 0.78-0.88$

$RMSE_{tr} = 0.08-0.11$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$Q^2_{LOO} = 0.66-0.84$

$RMSE_{LOO} = 0.10-0.14$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

4 M Metal Oxide

List

TiO₂

ZnO

Shape: NA

Coating: NA

Size(nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$r^2_{\text{ext}} = 0.47-0.91$

RMSE_{ext} = 0.7-0.20

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

RBFNN: Radial Basis Function Neural Networks

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out c

9.2.Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X4: Concentration

- X0: Engineered size

- X2: Size in PBS,RBFNN: Radial Basis Function Neural Networks
by MATLAB

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by TiO2 NPs by non-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by TiO2 NPs by non-linear regression method
(RBFNN case for TiO2)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO2 and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(RBFNN case for TiO2)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RBFNN: Radial Basis Function Neural Networks
by MATLAB

4.3.Descriptors in the model:

- X0: Engineered size
 - X4: Concentration
- ; 2

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

18/2

Descriptor: Chemical ratio :2:22 ~ 1:11

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 22 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

18 Metal Oxide

List: TiO₂

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 24 to 22.

From initial 22 NPs ten different random splitting of 20 % out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 10 splittings:

$r^2_{tr} = 0.78-0.96$

$RMSE_{tr} = 0.06-0.15$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 10 splittings:

$Q^2_{LOO} = 0.73-0.93$

$RMSE_{LOO} = 0.07-0.13$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

4 M Metal Oxide

List

TiO₂

Shape:NA

Coating:NA

Size(nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Range of statistics for 10 splittings :

$r^2_{\text{ext}} = 0.11-0.98$

RMSE_{ext} = 0.05-0.18

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

RBFNN: Radial Basis Function Neural Networks

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out c

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Immortalized rat L2 lung epithelial cells
and


rat lung alveolar macrophages, QSAR, - X0: Engineered size

- X4: Concentration

, RBFNN: Radial Basis Function Neural Networks

by MATLAB

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO NPs by non-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO NPs by non-linear regression method
(RBFNN case for ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(RBFNN case for ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RBFNN: Radial Basis Function Neural Networks
by MATLAB

4.3.Descriptors in the model:

- X1: Size in water
- X2: Size in PBS
- X4: Concentration; 3

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

12/3

Descriptor: Chemical ratio :3:15 ~ 1:15

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 15 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

12 Metal Oxide

List: ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 18 to 15.

From initial 15 NPs ten different random splitting of 20 % out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 10 splittings:

$r^2_{tr} = 0.86-0.99$

$RMSE_{tr} = 0.02-0.08$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 10 splittings:

$Q^2_{LOO} = 0.66-0.96$

$RMSE_{LOO} = 0.05-0.17$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

3 M Metal Oxide

List

ZnO

Shape:NA

Coating:NA

Size(nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Range of statistics for 10 splittings :

$r^2_{\text{ext}} = 0.36-0.99$

$\text{RMSE}_{\text{ext}} = 0.04-0.20$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

RBFNN: Radial Basis Function Neural Networks

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out c

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X1: Size in water

- X2: Size in PBS

- X4: Concentration, RBFNN: Radial Basis Function Neural Networks
by MATLAB

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by non-linear regression method (GRNN case for TiO₂ and ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(GRNN case for TiO₂ and ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

GRNN: General Regression Neural Networks
by MATLAB

4.3.Descriptors in the model:

- X4: Concentration
- X0: Engineered size
- X2: Size in PBS; 3

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

27/3

Descriptor: Chemical ratio :3:31 ~ 1:10

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 31 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

27 Metal Oxide

List: TiO₂

ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 42 to 31.

From initial 31 NPs ten different random splitting of 20 % out, and one 50% out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$r^2_{tr} = 0.79-0.99$

$RMSE_{tr} = 0.02-0.11$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$Q^2_{LOO} = 0.44-0.76$

$RMSE_{LOO} = 0.11-0.17$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

4 Metal Oxide

List

TiO₂

ZnO

Shape: NA

Coating: NA

Size(nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

TiO₂ : Anatase/Rutile

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

Range of statistics for 11 splittings (10x 20% and 1x 50% of training set as test set):

$r^2_{\text{ext}} = 0.67-0.94$

RMSE_{ext} = 0.7-0.17

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

GRNN: General Regression Neural Networks

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross

9.2.Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. Risk Analysis, 30(11), 1723–1734.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X4: Concentration

- X0: Engineered size

- X2: Size in PBS, GRNN: General Regression Neural Networks
by MATLAB

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by TiO2 NPs by non-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by TiO2 NPs by non-linear regression method
(GRNN case for TiO2)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO2 and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(GRNN case for TiO2)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

GRNN: General Regression Neural Networks
by MATLAB

4.3.Descriptors in the model:

- X0: Engineered size
 - X4: Concentration
- ; 2

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

18/2

Descriptor: Chemical ratio :2:22 ~ 1:11

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 22 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

18 Metal Oxide

List: TiO₂

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 24 to 22.

From initial 22 NPs ten different random splitting of 20 % out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 10 splittings:

$r^2_{tr} = 0.80-0.90$

$RMSE_{tr} = 0.06-0.11$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 10 splittings:

$Q^2_{LOO} = 0.64-0.83$

$RMSE_{LOO} = 0.12-0.15$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

4 M Metal Oxide

List

TiO₂

Shape:NA

Coating:NA

Size(nm): Engineered Size: 30, 45, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Range of statistics for 10 splittings :

$r^2_{\text{ext}} = 0.50-0.98$

RMSE_{ext} = 0.11-0.18

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

GRNN: General Regression Neural Networks

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Immortalized rat L2 lung epithelial cells
and


rat lung alveolar macrophages, QSAR, - X0: Engineered size

- X4: Concentration

, GRNN: General Regression Neural Networks

by MATLAB

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO NPs by non-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO NPs by non-linear regression method (GRNN case for ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(GRNN case for ZnO)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

GRNN: General Regression Neural Networks
by MATLAB

4.3.Descriptors in the model:

- X1: Size in water
- X2: Size in PBS
- X4: Concentration; 3

4.4.Descriptor selection:

The best modelling variables selected by MLR case, were selected to develop non-linear regression models.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

12/3

Descriptor: Chemical ratio :3:15 ~ 1:15

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

From the initial data set (42 NPs) PCA analysis was performed on the available descriptors, and the large ZnO NPs (>1000 nm at concentration of 50 mg/L) were isolated from the rest of NPs. An analysis in the supplementary material was presented about the agglomeration rate vs the LDH, where two TiO₂ NPs (45nm at concentration of 100 mg/L, IDs 13 and 14) presented an anomaly behaviour. The large ZnO NPs and those two TiO₂ (ID 13 and 14) NPs were extracted from the data, which increased the perform of the models, if we compare it with the previous study (data source)

PCA was applied to the 15 NPs and not highlighted outliers were identified.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

12 Metal Oxide

List: ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

After an applicability domain assessment, the initial data set was reduced from 24 to 22.

From initial 15 NPs ten different random splitting of 20 % out were generated

6.7.Statistics for goodness-of-fit:

Range of statistics for 10 splittings:

$r^2_{tr} = 0.94-0.98$

$RMSE_{tr} = 0.03-0.06$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Range of statistics for 10 splittings:

$Q^2_{LOO} = 0.79-0.94$

$RMSE_{LOO} = 0.06-0.11$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

3 M Metal Oxide

List

ZnO

Shape:NA

Coating:NA

Size(nm): Engineered Size: 50, 60, 70

Size in water: 55-172

Size in PBS: 158-385

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Q water by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Range of statistics for 10 splittings :

$r^2_{\text{ext}} = 0.52-1.00$

$\text{RMSE}_{\text{ext}} = 0.04-0.12$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

Despite of the 11 split results give an idea of stability, and the increase of data increases the robustness of the model, the fact that all the data was used in the final model, hence letting the final model without an external validation, which should be taken into account.

NPs: Nanoparticles

PCA: Principal Component Analysis

GRNN: General Regression Neural Networks

r^2 : correlation coefficient

RMSE: root-mean-square error

RMSE_LOO : root-mean-square error for leave-one-out cross validation

Q^2_{LOO} : leave-one-out cross

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X1: Size in water

- X2: Size in PBS

- X4: Concentration, GRNN: General Regression Neural Networks
by MATLAB

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Modelling of membrane disruption mediated by ZnO and TiO₂ NPs by classification tree method
(Classification tree case for TiO₂ and ZnO)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., & Doucet-Panaye, A. (2015). Linear and non-linear modelling of the cytotoxicity of TiO₂ and ZnO nanoparticles by empirical descriptors. SAR and QSAR in Environmental Research, 26(7-9), 647–665.

(Classification tree case for TiO₂)

<http://doi.org/10.1080/1062936X.2015.1080186>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Immortalized rat L2 lung epithelial cells

and

rat lung alveolar macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

Cell culture systems were cultured in F-12K medium (Kaighn's modification of Ham's F-12 medium) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cellular membrane damage was collected at 80–85% confluency. Tests for cellular membrane damage were done in triplicate.

Characterize the culture media by using Olympus Lactate Dehydrogenase reagents (absorbance method at 340 nm). The release [units/L] was classified in:

- $y < 0.99$ --> Dense cell membrane
- $0.99 < y < 1.09$ --> Normal cell membrane
- $1.09 < y < 1.25$ --> Leaky cell membrane
- $1.25 < y$ --> Disrupted cell membrane

In this paper they only used one cutoff value (i.e. LDH > 1.09) to distinguish between toxic (i.e. causing leaking or disruption of the membrane) and non-toxic effects due to exposure to TiO₂ and ZnO NPs

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

J48 classification tree
by WEKA software

4.3.Descriptors in the model:

- X1: Size in water
- X4: Concentration; 2

4.4.Descriptor selection:

The final descriptors were selected within the model building algorithm (Classification tree)

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

28/2

Descriptor: Chemical ratio :2:28 ~ 1:14

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

PCA was performed to evaluate the distribution of the studied NPs. As

well as the presence of areas of partial overlapping among the two activity classes, where the misclassified NPs were (red circles in the publication's Figure S12)

The applicability domain is defined as ZnO and TiO₂ NPs within the range of descriptors of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

28 Metal Oxide

List: TiO₂

ZnO

Shape: NA

Coating: NA

Size (nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other info: To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater

by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

6.6.Pre-processing of data before modelling:

For the complete set of NPs (n = 42). The 30% of the available data was not used to train the model, but only to perform the external validation (28 NPs were included in the training set, and 14 in the prediction set).

10 cross-validation groups for the training set were developed (ratio of subtraining and subtest groups of each one, was not specified)

6.7.Statistics for goodness-of-fit:

Sensitivity = 88 %

Specificity = 92 %

Accuracy = 89 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

For 10 cross-validation groups:

Sensitivity = 81 %

Specificity = 83 %

Accuracy = 82 %

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

14 Metal Oxide

List

TiO₂

ZnO

Shape:NA

Coating:NA

Size(nm): Engineered Size: 30, 45, 50, 60, 70, 125

Size in water: 101-967

Size in PBS: 961-3871

Other properties:

To reduce particle settlement, Tween 20 (~1% v/v) was added to each nanoparticle stock suspension.

Size and size distribution was determined by transmission electron microscopy (TEM) and dynamic light scattering (DLS) spectroscopy (dry powder and aqueous). Zeta potential was measured in ultrapure Milli-Qwater by DLS.

Particle size was determined via DLS, on a ZetaSizer Nano-ZS instrument (Malvern Inc., Worcestershire, UK). The instrument measures the size of the suspended particles through Brownian motion. Malvern DLS software version 5.03 was used to analyze the results.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 100 %

Specificity = 86 %

Accuracy = 93 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1. Comments:

Obtained data was analyzed and better results against the previous work were presented.

NPs: Nanoparticles

J48: open source Java implementation of the C4.5 (an algorithm used to generate a decision tree) in the Weka data mining tool

PCA: Principal Component Analysis

9.2. Bibliography:

(already reported in this table)

Sayes, C., & Ivanov, I. (2010). Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 30(11), 1723–1734.

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:


Cell, Immortalized rat L2 lung epithelial cells
and

rat lung alveolar macrophages, QSAR, - X1: Size in water

- X4: Concentration, J48 classification tree

by WEKA software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Toxicity model for metal oxide NPs by PLS
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Toxicity model for metal oxide NPs by PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Xue Z. Wang

x.z.wang@leeds.ac.uk

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Oksel, C., Ma, C. Y., & Wang, X. Z. (2015). Structure-activity relationship models for hazard assessment and risk management of engineered nanomaterials. *Procedia Engineering*, 102(0), 1500–1510. <http://doi.org/10.1016/j.proeng.2015.01.284>

<http://doi.org/10.1016/j.proeng.2015.01.284>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial cells

3.2. Endpoint:

In vitro - Cytotoxicity - measured as cellular viability by determining the mitochondrial dehydrogenases' activity

3.3. Comment on endpoint:

Initially, 6 different toxicity assay at different doses expositions were analyzed, but due to the bad performance of the model applied to different endpoints at a time, they decide to focus the study on a single toxicity assay.

Cellular viability was assessed in A549 human lung epithelial cells exposed to NPs (24 h; 1-100 µg/ml) in terms of mitochondrial function, using a commercially available assay (MTT). The basis of this assay viable cell-mediated conversion of colourless, water soluble (3-(4,5-dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide; MTT) to an insoluble formazan that can be measured spectrophotometrically ($\lambda=570$ nm). Conversion of the formazan to MTT is dependent on NAD(P)H reductase enzymes, and hence is related to the metabolic activity of the cells.

The mitochondrial impact of the NP panel was explored in more detail using a flow cytometry-based assessment of mitochondrial membrane potential using the fluorescent dye, DiOC6, to detect collapse of this potential. Traditionally, this test is used as an early marker of apoptosis and is conducted in parallel with a test for secondary necrosis (PI in this case). This test was conducted in a different, but equally relevant, cell type (THP-1 human monocyte-like cells) because the assay is better-suited to non-adherent cells that remain in suspension.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- x32: zinc content
- x33: cadmium content
- x13 and x14: oxygen-centred free radical activities
- x3: specific surface area
- x28: reactivity; 6

4.4.Descriptor selection:

The final descriptors were selected within the model building algorithm (PLS with 3 three principal components)

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

10/6

Descriptor: Chemical ratio :6:10 ~ 1:2

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

The authors say: that "Although some case-specific correlations between the properties of ENMs and their biological activity were observed, it was not possible to generalize these findings for external ENMs."

It was expected an applicability domain of metal oxide NPs within the range of parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

10 Metal

Metal Oxide

List: Al₂O₃

CeO₂

NiO

SiO₂

ZnO

Ag

Shape: NA

Coating: NA

Size (nm): NA

Other info: TiO₂ MPs have rutile and anatase crystal conformations.

Al₂O₃ sizes of 7, 50, 300 nm

6.6.Pre-processing of data before modelling:

Cross-validation was performed but the ratio of subtraining and subtest groups were not specified

6.7.Statistics for goodness-of-fit:

$R^2 = 0.99$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2 = 0.80$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal

Metal Oxide

List

Al₂O₃

CeO₂

NiO

SiO₂

ZnO

Ag

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

TiO₂ MPs have rutile and anatase crystal conformations.

Al₂O₃ sizes of 7, 50, 300 nm

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

There is not an external validation procedure, and the data size is not enough to ensure the reliability of the model. Also the final regression function obtained from the PLS technique was not provided, which gives a confusing idea of the final used descriptors.

ENMs: Engineered nanomaterials

NPs: Nanoparticles

PLS: Partial Least Squares

R²_Y: correlation coefficient

Q²: cross-validation correlation coefficient

9.2.Bibliography:

Wang, X. Z., Yang, Y., Li, R., McGuinness, C., Adamson, J., Megson, I. L., & Donaldson, K. (2014). Principal component and causal analysis of structural and acute in vitro toxicity data for nanoparticles. *Nanotoxicology*, 8(5), 465–476.
10.3109/17435390.2013.796534

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial cells, QSAR, - x32: zinc content


- x33: cadmium content

- x13 and x14: oxygen-centred free radical activities

- x3: specific surface area

- x28: reactivity,PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cytotoxicity of metal oxide nanoparticles on E. coli prediction by nano-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cytotoxicity of metal oxide nanoparticles on E. coli prediction by nano-read-across (t-HCA)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tomasz Puzyn

t.puzyn@qsar.eu.org

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Gajewicz, A., Cronin, M. T. D., Rasulev, B., Leszczynski, J., & Puzyn, T. (2015). Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: Nano-read-across. Nanotechnology, 26(1).

(case study 1)

<http://doi.org/10.1088/0957-4484/26/1/015701>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as -log(LC50)

3.3. Comment on endpoint:

Determined the cytotoxicity of the metal oxide nanoparticles in terms of EC50 (concentration which cytotoxicity reduces bacteria viability up to 50%) based on the curve fitting least squares procedure.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

nano-read-across by:

t-HCA: two dimensional hierarchical cluster analysis using Euclidean distance and Ward's method of linkage

4.3.Descriptors in the model:

- ΔH_{Me+} : represents the enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure.; 1

4.4.Descriptor selection:

Selection of independent variable(s) that define the similarity of MeOx was conducted based on the value of the Pearson's correlation coefficient calculated between the matrix of all descriptors (X) and the vector of the dependent variable (y) for the training set. Only those descriptors that had been found to contribute significantly to understanding the mechanism of toxicity and having a Pearson's correlation coefficient with the endpoint with an absolute value greater than 0.8. It should be mentioned, however, that sometimes a model can include several descriptors that may have a low individual correlation with activity, but in combination provide a good model. Nevertheless, it is generally assumed that the proposed nano-read-across approach should be maximally simplified and based on the minimal number of used descriptor(s).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

10/1

Descriptor: Chemical ratio :1:10

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Can be interpreted as those NPs which will be closely related with Metal Oxide NPs (also within the range of the applied descriptor) used in the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

10 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

6.6.Pre-processing of data before modelling:

The splitting algorithm was as follows:

- (1). 13 metal oxides for which toxicity data had been either taken from the previous paper, or they had been tested in Batch I were sorted based on decreasing toxicity.
- (2). In a next step they were split into two sets: the training set (T) and the validation set (V1) in a way ensured that the points from V1 were evenly distributed within the range of the toxicity of the training set compounds (T). We utilized the following pattern of splitting: T-T-V1-T-T-T-V1-T-T-T-V1-T-T.
- (3). Finally, three additional compounds tested in Batch II and La₂O₃ were additionally included in the validation set (those compounds are indicated with V2).

We split the data in an above discussed way because of three reasons:

- (i) to ensure that the compounds V1 are evenly distributed within the range of toxicity log (1/EC₅₀),
- (ii) to have both experimental batches represented in the validation set, whereas only compounds from the Batch I were used for training,
- (iii) to include to the validation set some additional compounds (V2) having toxicity not necessarily within the range of the training set (this would be impossible, if we have merged compounds from Batch I and II together and then labeled every third compound as a member of the validation set). Indeed, observed toxicity of CoO was higher than toxicity of the most toxic compound in the training set (ZnO).

6.7.Statistics for goodness-of-fit:

A_training = 100.00 %

E_training = 0.00 %

(experiment) $p_s = 0.955$

(experiment) critical $p_s = 0.414$

$p = 0.0001$

(nano-QSAR) $p_s = 0.961$

(nano-QSAR) critical $p_s = 0.279$

$p = 0.0001$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

7 MMetal Oxide

List

ZnO

CuO

Al₂O₃Fe₂O₃SnO₂TiO₂V₂O₃Y₂O₃Bi₂O₃In₂O₃Sb₂O₃SiO₂ZrO₂

CoO

NiO

Cr₂O₃La₂O₃Shape:NACoating:NASize(nm): 15-90Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the

source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

A_validation = 100.00 %

E_validation = 0.00 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Actually it could be interpreted as a filling gap of data procedure, but at a time is applying t-HCA it became a classification model.

The obtained results were compared with a previous nano-QSAR model (source data) and also to the experimental results.

A_training: Accuracy value for training set

A_validation: Accuracy value for validation set

E_training: Error value for training set

E_validation: Error value for validation set

EC50 : concentration of a drug, antibody or toxicant which induces a res

9.2.Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ...

Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, 6(3), 175–178.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - ΔH_{Me+} : represents the enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure., nano-read-across by:

t-HCA: two dimensional hierarchical cluster analysis using Euclidean distance and Ward's method of linkage

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cytotoxicity of metal oxide nanoparticles on HaCaT cell line prediction
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cytotoxicity of metal oxide nanoparticles on HaCaT cell line prediction by nano-read-across (t-HCA)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tomasz Puzyn

t.puzyn@qsar.eu.org

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Gajewicz, A., Cronin, M. T. D., Rasulev, B., Leszczynski, J., & Puzyn, T. (2015). Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: Nano-read-across. Nanotechnology, 26(1).

(case study 1)

<http://doi.org/10.1088/0957-4484/26/1/015701>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human keratinocyte cell line (HaCaT)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(LC50)

3.3. Comment on endpoint:

Cell viability was measured using the CytoTox-Glo Cytotoxicity Assay from Promega (Madison, WI). LC50 values for all MeOx were extrapolated using the third order polynomial equation of the log transformed data with the least squares fit in GraphPad (GraphPad Software, Inc., La Jolla, CA)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

nano-read-across by:

t-HCA: two dimensional hierarchical cluster analysis

using Euclidean distance and Ward's method of linkage

4.3.Descriptors in the model:

- X^c : Mulliken's electronegativity; 1

4.4.Descriptor selection:

Selection of independent variable(s) that define the similarity of MeOx was conducted based on the value of the Pearson's correlation coefficient calculated between the matrix of all descriptors (X) and the vector of the dependent variable (y) for the training set. Only those descriptors that had been found to contribute significantly to understanding the mechanism of toxicity and having a Pearson's correlation coefficient with the endpoint with an absolute value greater than 0.8. It should be mentioned, however, that sometimes a model can include several descriptors that may have a low individual correlation with activity, but in combination provide a good model. Nevertheless, it is generally assumed that the proposed nano-read-across approach should be maximally simplified and based on the minimal number of used descriptor(s).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

10/1

Descriptor: Chemical ratio :01:10

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Can be interpreted as those NPs which will be closely related with Metal Oxide NPs (also within the range of the applied descriptor) used in the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

10 Metal Oxide

List: Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): 15-150

Other info: Calculated and measured descriptors are listed in tables S8 and S9 of publication's supplementary material.

6.6.Pre-processing of data before modelling:

The metal oxides for which both kinds of data (describing the toxicity and the structure) had been available were split into two sets: the training set (T) and the validation set (V)

6.7.Statistics for goodness-of-fit:

A_training = 80.00 %

E_training = 20.00 %

(experiment) $p_s = 0.732$

(experiment) critical $p_s = 0.401$

$p = 0.0001$

(nano-QSAR) $p_s = 0.866$

(nano-QSAR) critical $p_s = 0.337$

$p = 0.0001$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

8 M Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape:NA

Coating:NA

Size(nm): 15-150

Other properties:

Calculated and measured descriptors are listed in tables S8 and S9 of publication's supplementary material.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

A_validation = 87.50 %

E_validation = 12.50 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Actually it could be interpreted as a filling gap of data procedure, but at a time is applying t-HCA it became a classification model.

The obtained results were compared with a previous nano-QSAR model (source data) and also to the experimental results.

MeOx: Metal Oxide

A_training: Accuracy value for training set

A_validation: Accuracy value for validation set

E_training: Error value for training set

E_validation: Error value for validation set

LC50: for a substance is the dose required to kill ha

9.2. Bibliography:

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325. 10.3109/17435390.2014.930195

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC


10.3. Keywords:

Cell, Human keratinocyte cell line (HaCaT), QSAR, - X^c: Mulliken's electronegativity, nano-read-across by:

t-HCA: two dimensional hierarchical cluster analysis

using Euclidean distance and Ward's method of linkage

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction of Toxicity of metal oxide NPs to E. Coli with/without photo-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction of Toxicity of metal oxide NPs to E. Coli with/without photo-inducing by SMILES-based optimal descriptor and Monte Carlo technique

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Toropova, A. P., Toropov, A. A., Rallo, R., Leszczynska, D., & Leszczynski, J. (2015). Optimal descriptor as a translator of eclectic data into prediction of cytotoxicity for metal oxide nanoparticles under different conditions. Ecotoxicology and Environm

<http://doi.org/10.1016/j.ecoenv.2014.10.003>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as -log(LC50)

3.3.Comment on endpoint:

The samples in Pyrex and quartz test tubes were exposed to sunlight for 30 min with agitation in a water bath at 150 r.p.m. Similarly, corresponding samples were exposed under dark conditions, by wrapping the test samples in the test tubes with aluminium foils (solar irradiation outdoors: irradiance: UVA range = 3.979–4.652 mW/cm²; UVB range = 3.1–3.7 MED/h, where an MED is defined as the minimum erythral dose or the amount of UV radiation to produce barely perceptible erythema; 1 MED/h = 0.05833 W/m²).

The dark cytotoxicity and photo-induced cytotoxicity are examined as an united endpoint

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

Split1:

= , Al, Bi, Co, Cr, Cu, Fe, O, In, La, Ni, V, Sb, Y, Sn, Ti, '[', '^', Zn

Split2:

= , Bi, Co, Cr, Cu, Fe, O, In, La, V, Sb, Si, Y, Sn, Ti, '[', '^', Zn

Split3:

= , Al, Bi, Co, Cr, Cu, Fe, O, In, La, Ni, V, Sb, Si, Y, Sn, Ti, '[', '^', Zn

Split4:

= , Bi, Co, Cr, Cu, Fe, O, In, La, Ni, V, Sb, Si, Y, Ti, '[', '^', Zn, Zr

Split5:

= , Al, Bi, Co, Cr, Cu, Fe, O, In, La, Ni, V, Sb, Si, Y, Sn, Ti, '[', '^', Zn, Zr

Split6:

= , Al, Bi, Co, Cr, Cu, O, La, Ni, V, Sb, Si, Y, Ti, '[', '^', Zn

- Chemical Elements (O, V, La, etc)

- '=' : represents double bonds

- '^' : represents photo-inducing, Also called 'widehat'

- '[' or ']' are used in SMILES for encoding special group or metals

; 0

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :Split1:

19:23

Split2:

18:21

Split3:

20:21

Split4:

19:20

Split5:

21:22

Split6:

17:20

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The measure of statistical quality of attributes (descriptors) which are involved to build up model were estimated by Equation 9: $\text{defect}(A_k)$.

Having the numerical data on the $\text{defect}(A_k)$ one can estimate reliability of the model for a representation of metal oxide nanoparticles by a quasi-SMILES (Table 2 in the publication): the basic hypothesis is “the probability of the quasi-SMILES to be in the domain of applicability is inversely proportional of sum of A_k -defects

2^* average(Defect-quasi-SMILES) is used as indicator. Carried out with the CORAL software (CORAL, 2014). The percentage of the domain of applicability, according to the analysis revealed by this software is 100%, 76%, 76%, 71%, 71%, and 71%, for splits 1, 2, 3, 4, 5, and 6 respectively.

Domain of applicability of more than 50% should be considered as satisfactory.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No
 Chemical Name: not applicable
 SMILES: not applicable
 Formula: not applicable
 INChI: not applicable
 MOL file: not applicable
 Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes
 NP size: Yes
 NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Obtained data from Pathakoti et al., 2014 (already reported in this table)

Primary particle was measured by using transmission electron microscopy (TEM). Samples were prepared by drop-coating the NP suspension onto a carbon-coated copper grid (Ted Pella, CA) and then the samples were dried overnight at room temperature. The samples were observed using a TEM (JEOL JEM-1011). The hydrodynamic diameters (z-average) were measured in distilled water (at a concentration of 100 ppm in water) and zeta potentials of

the MNPs were measured in both distilled water and 1mM KCl solution using Malvern Zeta Sizer (Nano-ZS, Malvern Instruments, UK). All measurements were conducted in triplicate at 25 °C and an average value was determined.

6.6.Pre-processing of data before modelling:

Six random distributions of the available data into training and calibration sets (these metal oxide nanoparticles are used to build up the model) and validation set (these metal oxide nanoparticles are not involved to build up the model, they are used to check up predictability of the model) are examined. All these splits are prepared according to the following principles:

- (i) they are random;
- (ii) the range of endpoints in each sub-set is similar to ranges for other sub-sets;
- (iii) these splits are not identical (Table 1 in the publication)

6.7.Statistics for goodness-of-fit:

- Training :

Split1: $r^2 = 0.9250$ $s = 0.347$

Split2: $r^2 = 0.9464$ $s = 0.317$

Split3: $r^2 = 0.9469$ $s = 0.293$

Split4: $r^2 = 0.9276$ $s = 0.339$

Split5: $r^2 = 0.9081$ $s = 0.354$

Split6: $r^2 = 0.9160$ $s = 0.370$

- Calibration:

Split1: $r^2 = 0.7279$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Split1:

$q^2 = 0.9115$

(c) $R^2_p = 0.62$

Split2:

$q^2 = 0.9384$

(c) $R^2_p = 0.89$

Split3:

$q^2 = 0.9396$

(c) $R^2_p = 0.84$

Split4:

$q^2 = 0.9127$

(c) $R^2_p = 0.74$

Split5:

$$q^2 = 0.8925$$

$$(c)R^2_p = 0.91$$

Split6:

$$q^2 = 0.9006$$

$$(c)R^2_p = 0.87$$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

Split1:

6

Split2:

6

Split3:

6

Split4:

7

Split5:

6

Split6:

7 MMetal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃
 Y₂O₃
 Bi₂O₃
 In₂O₃
 Sb₂O₃
 SiO₂
 ZrO₂
 CoO
 NiO
 Cr₂O₃
 La₂O₃

Shape:NA

Coating:NA

Size(nm): 15-90

Other properties:

Obtained data from Pathakoti et al., 2014 (already reported in this table)

Primary particle was measured by using transmission electron microscopy (TEM). Samples were prepared by drop-coating the NP suspension onto a carbon-coated copper grid (Ted Pella, CA) and then the samples were dried overnight at room temperature. The samples were observed using a TEM (JEOL JEM-1011). The hydrodynamic diameters (z-average) were measured in distill water (at a concentration of 100 ppm in water) and zeta potentials of the MNPs were measured in both distill water and 1mM KCl solution using Malvern Zeta Sizer (Nano-ZS, Malvern Instruments, UK). All measurements were conducted in triplicate at 25 °C and an average values was determined.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Split1:

$r^2 = 0.7332$

$s = 0.828$

Split2:

$r^2 = 0.7905$

$s = 0.858$

Split3:

$r^2 = 0.8078$

$s = 0.721$

Split4:

$r^2 = 0.8965$

$s = 0.367$

Split5:

$r^2 = 0.9835$

$s = 0.418$

Split6:

$r^2 = 0.8961$

$s = 0.300$

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

SMILES: Simplified Molecular Input Line Entry Specification

R^2 : correlation coefficient

s : root-mean-square error

q^2 : cross-validation correlation coefficient

CORAL: CORrelation And Logic

$(c)R^2_p$ = Parameter computed from correlations coefficient

9.2. Bibliography:

(already reported in this table)

Pathakoti, K., Huang, M.-J., Watts, J. D., He, X., & Hwang, H.-M. (2014). Using experimental data of Escherichia coli to develop a QSAR model for predicting the photo-induced cytotoxicity of metal oxide nanoparticles. Journal of Photochemistry and Photobiology B: Biology, 130, 234–240.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, Split1:

= , Al, Bi, Co, Cr, Cu, Fe, O, In, La, Ni, V, Sb, Y, Sn, Ti, '[', '^' , Zn

Split2:

= , Bi, Co, Cr, Cu, Fe, O, In, La, V, Sb, Si, Y, Sn, Ti, '[', '^' , Zn

Split3:

= , Al, Bi, Co, Cr, Cu, Fe, O, In, La, Ni, V, Sb, Si, Y, Sn, Ti, '[', '^' , Zn

Split4:

= , Bi, Co, Cr, Cu, Fe, O, In, La, Ni, V, Sb, Si, Y, Ti, '[', '^' , Zn, Zr

Split5:

= , Al, Bi, Co, Cr, Cu, Fe, O, In, La, Ni, V, Sb, Si, Y, Sn, Ti, '[', '^' , Zn, Zr

Split6:


= , Al, Bi, Co, Cr, Cu, O, La, Ni, V, Sb, Si, Y, Ti, '[', '^' , Zn

-
- Chemical Elements (O, V, La, etc)
 - '=' : represents double bonds
 - '^' : represents photo-inducing, Also called '\widehat'
 - '[' or ']' are used in SMILES for encoding special group or metals

,Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: modelling cell membrane damage of metal oxide nanoparticles to
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

modelling cell membrane damage of metal oxide nanoparticles to BEAS-2B by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software

package:

Toropova, A. P., Toropov, A. A., Benfenati, E., Korenstein, R., Leszczynska, D., & Leszczynski, J. (2015). Optimal nano-descriptors as translators of eclectic data into prediction of the cell membrane damage by means of nano metal-oxides.

Environmental Sc

<http://doi.org/10.1007/s11356-014-3566-4>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human bronchial epithelial cells (BEAS-2B)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as Percentage of damaged cells by Propidium Iodide uptake of

BEAS-2B

3.3.Comment on endpoint:

The numerical data on this endpoint related to four doses (50, 100, 150, and 200 µg/mL) and seven exposure time (from 1 to 7 h) for all 24 nanometal-oxides are examined. In fact, the percentage of cells which have membrane damage is the measure of impact of nano-oxides (for defined dose and exposure time). The decimal logarithm of these values is examined as the endpoint.

From 672 (24*4*7) data points, after four observation for CMD, were selected those with a dispersion of <10 %

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

- 22 Chemical elements (Al, Ce, Co, Cr, Cu, Fe, Gd, Hf, O, In, La, Mn, Ni, W, Sb, Si, Y, Sn, Ti, Yb, Zn, and Zr)
- Separator for chemical elements (dot) "."
- 4 Doses "A", "B", "C", "D" (200 µg/mL, 150 µg/mL, 100 µg/mL, 50 µg/mL respectively)
- 7 exposure times (1 - 7 h); 34

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/34

Descriptor: Chemical ratio :34:137 ~1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain for an approach is defined according to probabilistic criteria. One should use for nano metal-oxides which are containing only prevalent (in the "visible" training set) codes of various model components (chemical elements, dose and exposure time), associated with cell membrane damage.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: ZrO₂

ZnO

Yb₂O₃

Y₂O₃

WO₃

TiO₂

SnO₂

SiO₂

Sb₂O₃

NiO

Ni₂O₃

MnO₃

La₂O₃

In₂O₃

HfO₂

Gd₂O₃

Fe₃O₄

Fe₂O₃

CuO

Cr₂O₃

CoO

Co₃O₄

CeO₂

Al₂O₃

Shape: NA

Coating: NA

Size (nm): 10-100

Other info: exceptions outside range of sizes:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

6.6.Pre-processing of data before modelling:

The data are split into training, calibration and validation sets according to the following principles:

- (i) the external validation set contains about 15 % of the data
- (ii) these splits are random
- (iii) the identity of these splits is minimal.

Five splits built up according to the above-mentioned principles are examined in the present study. Table 1 (in the publication) contains the percentage of identity for these splits. The training sets are structured into sub-training set (developer of the model) and test set (calibration of the model)

6.7.Statistics for goodness-of-fit:

- Training

Split1: $r^2 = 0.5213$ $s = 0.392$

Split2: $r^2 = 0.5093$ $s = 0.393$

Split3: $r^2 = 0.5014$ $s = 0.393$

Split4: $r^2 = 0.5026$ $s = 0.389$

Split5: $r^2 = 0.5437$ $s = 0.394$

- Calibration:

Split1: $r^2 = 0.8424$ $s = 0.290$

Split2: $r^2 = 0.8628$ s

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Split1:

$q^2 = 0.4947$

$(c)R^2_p = 0.796$

Split2:

$q^2 = 0.4837$

$(c)R^2_p = 0.843$

Split3:

$q^2 = 0.4754$

$(c)R^2_p = 0.760$

Split4:

$q^2 = 0.4767$

$(c)R^2_p = 0.646$

Split5:

$q^2 = 0.5215$

$(c)R^2_p = 0.827$

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

Split1: 20

Split2: 20

Split3: 20

Split4: 20

Split5: 22 Metal Oxide

List

ZrO₂

ZnO

Yb₂O₃

Y₂O₃

WO₃

TiO₂

SnO₂

SiO₂

Sb₂O₃

NiO

Ni₂O₃

MnO₃

La₂O₃

In₂O₃

HfO₂

Gd₂O₃

Fe₃O₄

Fe₂O₃

CuO

Cr₂O₃

CoO

Co₃O₄

CeO₂

Al₂O₃

Shape:NA

Coating:NA

Size(nm): 10-100

Other properties:

exceptions outside range of sizes:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was

used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Split1:

$$r^2 = 0.9174$$

$$s = 0.280$$

Split2:

$$r^2 = 0.8110$$

$$s = 0.315$$

Split3:

$$r^2 = 0.6979$$

$$s = 0.250$$

Split4:

$$r^2 = 0.9268$$

$$s = 0.402$$

Split5:

$$r^2 = 0.7809$$

$$s = 0.348$$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

The average of internal validation $0.5 (r^2)$ gives a moderate idea of an statistical significance, also the cross-validation correlation coefficient are not greater than 0.5 in average.

It is important to notice that there is not a huge variability in the response data, which gives us a not good applicability of the model. If we check the source publication of the obtained data, we can observe that 7 over 24 NPs of the data set present a variability in the response versus the dose and the exposure time. Close to 1/3 of the data which presents variability could be hidden by the non variability response of the rest of the data.

SMILES: Simplified Molecular Input Line Entry Specification

R^2 : correlation coefficient

s: root-mean-square error

q^2 : cross-validation correlation coefficient

CORAL: CORrelation And Logic

(c) R^2_p = Parameter computed from correlations coefficient

9.2. Bibliography:

Patel T, Low-Kam C, Ji ZH, Zhang H, Xia T, Nel AE, Zinc JI, Telesca D (2012) Relating nanoparticle properties to biological outcomes in exposure escalation experiments
COBRA preprint series 2012, Working Paper 101.
<http://biostats.bepress.com/cobra/art101>

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Human bronchial epithelial cells (BEAS-2B), QSAR, - 22 Chemical elements (Al, Ce, Co, Cr, Cu, Fe, Gd, Hf, O, In, La, Mn, Ni, W, Sb, Si, Y, Sn, Ti, Yb, Zn, and Zr)


- Separator for chemical elements (dot) "."

- 4 Doses "A", "B", "C", "D" (200 $\mu\text{g/mL}$, 150 $\mu\text{g/mL}$, 100 $\mu\text{g/mL}$, 50 $\mu\text{g/mL}$ respectively)

- 7 exposure times (1 - 7 h), Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells by GA-PLS
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells by GA-PLS plus MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Wen Dai

dai.wei6@163.com

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Wen DAI, Xue-Ying SHAN, G.-Y. He. and H.-Q. C. (2015).
PREDICTION FOR CELLULAR UPTAKE OF MANUFACTURED
NANOPARTICLES TO PANCREATIC CANCER CELLS. *Revue
Roumaine de Chimie*, 60(4), 367–370.

NA

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3. Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

by the statistical software SPSS

4.3.Descriptors in the model:

- IC1: Information content index
- Hy: Hydrophilic factor
- Mor12u: 3D MoRSE-signal 12/unweighted; 3

4.4.Descriptor selection:

Up to 1666 molecular descriptors were calculated by using E-Dragon 1.0 software for each molecule. E-Dragon 1.0 software for each molecule. The molecular descriptors stayed constant for all molecules were eliminated. Then examine pair wise correlations between descriptors so that only the one with the highest correlation was retained (correlation coefficient > 0.95). With hundreds of descriptors remained GA-PLS was used to find the molecular descriptors closely related to cellular uptake. The GA-PLS programs were implemented using the software package PLS-Algorithm Toolbox written by Leardi and Lupiáñez.

Multicollinearity between the selected descriptors was examined by calculating their variance inflation factor (VIF) values, and the contribution of each descriptor was evaluated by calculating the value of membership functions (MF) parameters.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

87/3

Descriptor: Chemical ratio :3:87 ~ 1:29

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

$h^* = 0.135$

Two chemicals were identified with a higher leverage value than the warning h^* . Those are called structural outliers. Non data outside the standard residual limits (3)

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

87 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol

Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione

Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

Geometries of all molecules attached on the surface of CLIO-NH₂ were optimized by MM+ molecular mechanics force field and the semi-empirical AM1 method.

6.6.Pre-processing of data before modelling:

Not explicitly detailed in the text. Extracted from the Figure 2-3 in the publication, and the reference which was used to compare the obtained statistical results.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.899$

SEE = 0.146

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LOO} = 0.887$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

21 Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape:NA

Coating:Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic dianhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione

4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione

Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride

3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione

Dibenz(c,e)oxepin-5,7-dione

6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone

Lauric anhydride

1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid

5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine

N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein

isothiocyanate)

Overall size (volume weighted) in aqueous solution.

Geometries of all molecules attached on the surface of CLIO-NH₂ were optimized by MM+ molecular mechanics force field and the semi-empirical AM1 method.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$Q^2_{\text{ext}} = 0.834$

SEP = 0.104

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

There is a Mechanistic Interpretation.

Good agreement with the OECD framework.

NP: nanoparticle

GA-PLS: Genetic Algorithm - Partial Least Square

MLR: Multiple Linear Regression

R^2 : Correlation coefficient

SEE: standard deviation

Q^2_{LOO} : cross-validation leave-one-out correlation coefficient

Q^2_{ext} : correlation coefficient f

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - IC1: Information content index

- Hy: Hydrophilic factor

- Mor12u: 3D MoRSE-signal 12/unweighted,MLR: Multiple Linear Regression
by the statistical software SPSS

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical properties by SVR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yoram Cohen

yoram@ucla.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Liu, R., Jiang, W., Walkey, C. D., Chan, W. C. W., & Cohen, Y. (2015). Prediction of nanoparticles-cell association based on corona proteins and physicochemical properties. *Nanoscale*, 7(21), 9664–9675.

<http://doi.org/10.1039/c5nr01537e>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3. Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands)

with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio $\frac{m_{cell}}{m_{well}}$ divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (\log_2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

SVR: Support Vector Regression

4.3.Descriptors in the model:

- APOB: Apolipoprotein B-100
- A1AT: Alpha-1-antitrypsin
- IGLL5: Immunoglobulin lambda-like polypeptide 5
- ZP_syn: Zeta Potential (as synthesized) in mV
- HRG: Histidine-rich glycoprotein
- FA12: Coagulation factor XII
- APOE: Apolipoprotein E; 7

4.4.Descriptor selection:

Descriptor selection was accomplished by sequential forward floating selection (SFFS). At each selection step, SFFS first conducts a forward selection to identify the descriptor that leads to the greatest increase in model performance, then backward elimination to evaluate whether previously selected descriptors should be removed due to the addition of the newly selected one.

Based on the on model performance, with respect to the selected descriptors, the suitable descriptor number was then determined by locating the “turning point” being defined when the addition of a new descriptor led to insignificant improvement (e.g., $\leq 1\%$ increase in R^2) in model performance.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

84/7

Descriptor: Chemical ratio :7:84 ~ 1:12

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

AD was verified with average kernel similarity approach and Williams plot. (For specific details see the publication's Figure 6a)

$g^* = 0.005$

Covering all but two (G15.DTNB and G60.SPP) of the Au NPs.

G60.SPP is out of standard residual limits, which should be considered as an outlier

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

84 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic

anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrenecomaleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Transmission electron microscopy (TEM) confirmed that the nanoparticle cores were monodisperse and had uniform morphology. After surface modification, it was used dynamic light scattering (DLS) to measure the hydrodynamic diameter (HD) of each formulation and absorbance spectrophotometry (AS) to measure the localized surface plasmon resonance index (LSPRi) and LSPR peak wavelength (LSPRpeak). The electrophoretic mobility and zeta potential (ZP) were characterized using light scattering and agarose gel electrophoresis.

The composition of the protein corona around each formulation was characterized qualitatively using poly(acrylamide) gel electrophoresis (PAGE) and semiquantitatively using high-resolution label-free shotgun tandem mass spectrometry (LC-MS/MS). The abundance of several key adsorbed serum proteins was further confirmed by western blotting.

Further experimental details allocated in the Material and Methods section from source publication (Walkey, et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

The applied splitting was in order to perform a k-fold cross-validation test (k=4)

6.7.Statistics for goodness-of-fit:

$R^2_{E632} = 0.895$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

100-round Y-randomization:

- $R^2_{E632} = -0.208 \pm 0.109$

100-round 4-fold cross-validation:

- $R^2_{4cv} = 0.862 \pm 0.026$

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MMetal

List

Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine

Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Transmission electron microscopy (TEM) confirmed that the nanoparticle cores were monodisperse and had uniform morphology. After surface modification, it was used dynamic light scattering (DLS) to measure the hydrodynamic diameter (HD) of each formulation and absorbance spectrophotometry (AS) to measure the localized surface plasmon resonance index (LSPRi) and LSPR peak wavelength (LSPRpeak). The electrophoretic mobility and zeta potential (ZP) were characterized using light scattering and agarose gel electrophoresis.

The composition of the protein corona around each formulation was characterized qualitatively using poly(acrylamide) gel electrophoresis (PAGE) and semiquantitatively using high-resolution label-free shotgun tandem mass spectrometry (LC-MS/MS). The abundance of several key adsorbed serum proteins was further confirmed by western blotting.

Further experimental details allocated in the Material and Methods section from source publication (Walkey, et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The relative abundance of different protein corona were used as descriptors. Although they are not physicochemical descriptors it was concluded that protein corona encodes relevant biological information regarding cell association to the target NP.

Not external validation test was applied, hence we cannot say that the obtained results will be totally reliable.

NP: Nanoparticle

SVR: Support Vector Regression

R^2_{E632} : The 0.632 estimator. Suitable for performance validation of models based on small datasets.

$R^2_{E632} = 0.368 \cdot R^2_{resub} + 0.632 \cdot R^2_{boot}$, where

R^2_{resub} is the model prediction accuracy assessed

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - APOB: Apolipoprotein B-100

- A1AT: Alpha-1-antitrypsin

- IGLL5: Immunoglobulin lambda-like polypeptide 5


- ZP_syn: Zeta Potential (as synthesized) in mV

- HRG: Histidine-rich glycoprotein

- FA12: Coagulation factor XII

- APOE: Apolipoprotein E, SVR: Support Vector Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical properties by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yoram Cohen

yoram@ucla.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Liu, R., Jiang, W., Walkey, C. D., Chan, W. C. W., & Cohen, Y. (2015). Prediction of nanoparticles-cell association based on corona proteins and physicochemical properties. *Nanoscale*, 7(21), 9664–9675.

<http://doi.org/10.1039/c5nr01537e>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3. Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands)

with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (\log_2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- APOB: Apolipoprotein B-100
- ANT3: Antithrombin-III
- KLKB1: Plasma kallikrein
- TTHY: Transthyretin
- AMBP: Protein AMBP
- ITIH4: Inter-alpha-trypsin inhibitor heavy chain H4
- PON1: Serum paraoxonase/arylesterase 1
- HRG: Histidine-rich glycoprotein
- VOL_Au: Single NP Volume nm³
- IC1: Plasma protease C1 inhibitor
- FA10: Coagulation factor X; 11

4.4.Descriptor selection:

Descriptor selection was accomplished by sequential forward floating selection (SFFS). At each selection step, SFFS first conducts a forward selection to identify the descriptor that leads to the greatest increase in model performance, then backward elimination to evaluate whether previously selected descriptors should be removed due to the addition of the newly selected one.

Based on the on model performance, with respect to the selected descriptors, the suitable descriptor number was then determined by locating the “turning point” being defined when the addition of a new descriptor led to insignificant improvement (e.g., $\leq 1\%$ increase in R^2) in model performance.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

84/11

Descriptor: Chemical ratio :11:84 ~ 1:8

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

 $h^* = 0.43$

Covering all but three (G15.DDT-ODA, G15.MES, and G15.AHT) of the 84 Au NP

G15.AHT is out of standard residual limits, which should be considered as an outlier

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

84 Metal

List: AuShape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid

11-Mercaptoundecanoic acid

(11-Mercaptoundecyl)tetra(ethylene glycol)

(11-Mercaptoundecyl)-N,N,N-trimethylammonium

Amino-poly(ethylene glycol)-thiol (5kDa)

Amino-poly(ethylene glycol)-thiol (5kDa) (low density)

2-Napthalenethiol @ deoxycholic acid

2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)

Thiolated poly(ethyleneimine)

L-Phenylalanine

Thiolated L-phenylalanine

Thiolated poly(L-lysine)

Poly(vinyl alcohol)

Poly(vinylpyrrolidone)

Stearic acid

Thiolated L-serine

Bis(p-sulfonatophenyl)phenylphosphine

TWEEN20

Thiolated L-threonine

N-(2-Mercaptopropionyl)glycine

Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Transmission electron microscopy (TEM) confirmed that the nanoparticle cores were monodisperse and had uniform morphology. After surface modification, it was used dynamic light scattering (DLS) to measure the hydrodynamic diameter (HD) of each formulation and absorbance spectrophotometry (AS) to measure the localized surface plasmon resonance index (LSPRi) and LSPR peak wavelength (LSPRpeak). The electrophoretic mobility and zeta potential (ZP) were characterized using light scattering and agarose gel electrophoresis.

The composition of the protein corona around each formulation was characterized qualitatively using poly(acrylamide) gel electrophoresis (PAGE) and semiquantitatively using high-resolution label-free shotgun tandem mass spectrometry (LC-MS/MS). The abundance of several key adsorbed serum proteins was further confirmed by western blotting.

Further experimental details allocated in the Material and Methods section from source publication (Walkey, et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

Sequential forward floating selection (SFFS) was applied to select the most predictive fingerprints.

The applied splitting was in order to perform a k-fold cross-validation test (k=4)

6.7.Statistics for goodness-of-fit:

$R^2_{E632} = 0.850$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

100-round Y-randomization:

$R^2_{E632} = -0.208 \pm 0.109$

100-round 4-fold cross-validation:

$R^2_{4cv} = 0.843 \pm 0.015$

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic

anhydride)
 2-Naphthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Transmission electron microscopy (TEM) confirmed that the nanoparticle cores were monodisperse and had uniform morphology. After surface modification, it was used dynamic light scattering (DLS) to measure the hydrodynamic diameter (HD) of each formulation and absorbance spectrophotometry (AS) to measure the localized surface plasmon resonance index (LSPRi) and LSPR peak wavelength (LSPRpeak). The electrophoretic mobility and zeta potential (ZP) were characterized using light scattering and agarose gel electrophoresis.

The composition of the protein corona around each formulation was characterized qualitatively using poly(acrylamide) gel electrophoresis (PAGE) and semiquantitatively using high-resolution label-free shotgun tandem mass spectrometry (LC-MS/MS). The abundance of several key adsorbed serum proteins was further confirmed by western blotting.

Further experimental details allocated in the Material and Methods section from source publication (Walkey, et al., 2014)

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The relative abundance of different protein corona were used as descriptors. Although they are not physicochemical descriptors it was concluded that protein corona encodes relevant biological information regarding cell association to the target NP.

Not external validation test was applied, hence we cannot say that the obtained results will be totally reliable.

NP: Nanoparticle

MLR: Multiple Linear Regression

R^2_{E632} : The 0.632 estimator. Suitable for performance validation of models based on small datasets.

$R^2_{E632} = 0.368 \cdot R^2_{resub} + 0.632 \cdot R^2_{boot}$, where

R^2_{resub} is the model prediction accuracy assessed

9.2. Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - APOB: Apolipoprotein B-100

- ANT3: Antithrombin-III


- KLKB1: Plasma kallikrein

- TTHY: Transthyretin

- AMBP: Protein AMBP

- ITIH4: Inter-alpha-trypsin inhibitor heavy chain H4
- PON1: Serum paraoxonase/arylesterase 1
- HRG: Histidine-rich glycoprotein
- VOL_Au: Single NP Volume nm³
- IC1: Plasma protease C1 inhibitor
- FA10: Coagulation factor X,MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictive model of Mutagenicity of multi-walled carbon-nanotubes by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictive model of Mutagenicity of multi-walled carbon-nanotubes by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Toropov, A. A., & Toropova, A. P. (2015). Quasi-QSAR for mutagenic potential of multi-walled carbon-nanotubes. Chemosphere, 124(1), 40–46.

Toropova, A. P., Toropov, A. A., Rallo, R., Leszczynska, D., & Leszczynski, J. (2016). Nano-QSAR: Genotoxicity of

<http://doi.org/10.1016/j.chemosphere.2014.10.067>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Salmonella typhimurium TA100

3.2. Endpoint:

In vitro - Mutagenicity - measured as the number of observed colonies

3.3.Comment on endpoint:

The mutagenicity of a substance is proportional to the number of colonies observed.
 The numerical data on mutagenic potential of MWCNTs taken from the literature (Wirnitzer et al., 2009). Mean mutant counts after incubation of Salmonella strains TA100 without and with metabolic activation (S9 mix) in the plate incorporation and in the preincubation (tube) part of the Salmonella microsome test.
 The numerical data on TA100 are converted into negative decimal logarithmic scale.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Linear regression model
 based on SMILES-based optimal descriptors.

4.3.Descriptors in the model:

Split1:

y, n, +, -, A, B, C, D, E.

Split2:

y, n, +, -, D, E.

Split3:

y, n, +, -, A, B, C, D, E, F.

-
- "y": with preincubation
 - "n": without preincubation
 - "+": Presence of S9
 - "-": Absence of S9
 - Dose($\mu\text{g}/\text{plate}$):
 - A: 0
 - B: 50
 - C: 158
 - D: 500
 - E: 1581
 - F: 5000
- ; 0

4.4.Descriptor selection:

Optimal descriptors and Monte-Carlo optimization

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :Split1:

9:13 ~ 1:2

Split2:

6:13 ~ 1:2

Split3:

10:14 ~ 1:1

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

The measure of statistical quality of attributes (descriptors) which are involved to build up model were estimated by Equation 3: $\text{defect}(A_k)$.

Having the numerical data on the $\text{defect}(A_k)$ one can estimate reliability of the model for a representation of metal oxide nanoparticles by a quasi-SMILES (Table 4 in the publication): the basic hypothesis is "the probability of the quasi-SMILES to be in the domain of applicability is inversely proportional of sum of A_k -defects

$\text{defect}(A_k) < 2 \times \text{average}(\text{Defect-quasi-SMILES})$ is used as indicator to decide if A_k fall inside of the applicability domain.

The number of nanoparticles which fall into the domain of applicability:

Split1:

12

Split2:

20

Split3:

16

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable
 Formula: not applicable
 INChI: not applicable
 MOL file: not applicable
 Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes
 NP size: Yes
 NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Carbon-based

List: MWCNTs (Multi-walled carbon nanotubes)

Shape: Fiber

Coating: NA

Size (nm): NA

Other info: The particle size of the bulk material ranges from 100,000 nm to 3,000,000 nm

6.6.Pre-processing of data before modelling:

The data are split into the training, test, and validation sets according to the following principles:

- (i) the split is random;
- (ii) the ranges of endpoint for the above-mentioned sets are similar
- (iii) the identity of these splits is minimal.

Three splits built up according to the above-mentioned principles are examined in the present study

6.7.Statistics for goodness-of-fit:

- Training

Split1:

$$r^2 = 0.8037$$

$$s = 0.033$$

Split2:

$$r^2 = 0.6446$$

$$s = 0.045$$

Split3:

$$r^2 = 0.8087$$

$$s = 0.026$$

- Calibration -

Split1:

$r^2 = 0.9102$

$s = 0.071$

Split2:

$r^2 = 0.6785$

$s = 0.054$

Split3:

$r^2 = 0.9453$

$s = 0.074$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

- Training

Split1:

$(c)R^2_p = 0.7493$

$q^2 = 0.7260$

Split2:

$(c)R^2_p = 0.6341$

$q^2 = 0.4733$

Split3:

$(c)R^2_p = 0.7927$

$q^2 = 0.6975$

- Test

Split1:

$(c)R^2_p = 0.7268$

Split2:

$(c)R^2_p = 0.5159$

Split3:

$(c)R^2_p = 0.7245$

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable
 Formula: not applicable
 INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA
 NP size: NA
 NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

Split1:

6

Split2:

6

Split3:

5 MCarbon-based

List

MWCNTs (Multi-walled carbon nanotubes)

Shape:Fiber

Coating:NA

Size(nm): NA

Other properties:

The particle size of the bulk material ranges from

100,000 nm to 3,000,000 nm

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Split1:

$r^2 = 0.7627$

$s = 0.044$

Split2:

$r^2 = 0.9593$

$s = 0.032$

Split3:

$r^2 = 0.8951$

$s = 0.052$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

There are not structural or molecular descriptors, thus could be not considered as QSAR model.

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models.

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

The same work, with an other source of data, and almost the same descriptors with worst results were done in a posterior work of the same authors:

Toropova, A. P., Toropov, A. A., Rallo, R., Leszczynska, D., & Leszczynski, J. (2016). Nano-QSAR: Genotoxicity of multi-walled carbon nanotubes. International Journal of Environmental Research, 10(1), 59–64.

SMILES: Simplified Molecular Input Line Entry Specification

R^2 : correlation coefficient

s: root-mean-square error

q^2 : cross-validation correlation coefficient

CORAL: CORrelation And Logic

(c) R^2_p = Parameter computed from correlations coefficient

9.2. Bibliography:

Wirnitzer, U., Herbold, B., Voetz, M., & Ragot, J. (2009). Studies on the in vitro genotoxicity of baytubes??, agglomerates of engineered multi-walled carbon-nanotubes (MWCNT). Toxicology Letters, 186(3), 160–165.
<http://doi.org/10.1016/j.toxlet.2008.11.024>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Salmonella typhimurium TA100, QSAR, Split1:

y, n, +, -, A, B, C, D, E.

Split2:

y, n, +, -, D, E.

Split3:


y, n, +, -, A, B, C, D, E, F.

-
- "y": with preincubation
 - "n": without preincubation
 - "+": Presence of S9
 - "-": Absence of S9
 - Dose(μ g/plate):
 - A: 0
 - B: 50
 - C: 158
 - D: 500
 - E: 1581
 - F: 5000

,Linear regression model

based on SMILES-based optimal descriptors.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictive model of Mutagenicity of multi-walled carbon-nanotubes and
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictive model of Mutagenicity of multi-walled carbon-nanotubes and fullerene by SMILES-based optimal descriptor and Monte Carlo technique (CORAL software)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.A Toropov

andrey.toropov@mrionegri.it

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software

package:

Toropov, A. A., & Toropova, A. P. (2015). Quasi-SMILES and nano-QFAR: United model for mutagenicity of fullerene and MWCNT under different conditions. Chemosphere, 139, 18–22.

<http://doi.org/10.1016/j.chemosphere.2015.05.042>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Salmonella typhimurium TA101

3.2. Endpoint:

In vitro - Mutagenicity - measured as the number of observed colonies

3.3. Comment on endpoint:

The mutagenicity of a substance is proportional to the number of colonies observed
The numerical data on mutagenic potential of MWCNTs taken from the literature (Wirnitzer et al.,

2009) and for fullerene (Shinohara et al., 2009). Mean mutant counts after incubation of Salmonella strains TA100 without and with metabolic activation (S9 mix) in the plate incorporation and in the preincubation (tube) part of the Salmonella microsome test.

The numerical data on TA100 are converted into negative decimal logarithmic scale.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

- X: Fullerene
- Z: MWCNT
- "0": Dark condition
- "1" : Irradiation condition
- "Y": with preincubation
- "N": without preincubation
- "+": Presence of S9
- "-": Absence of S9
- Dose(g/plate) C60:
 - A: 50
 - B: 100
 - C: 200
 - D: 400
 - E: 1000
- Dose(µg/plate) MWCNT:
 - F: 0
 - G: 50
 - H: 158
 - I : 500
 - J: 1581
 - K: 5000; 19

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

0/19

Descriptor: Chemical ratio :Split1:

19:25

Split2:

19:29

Split3:

19:26

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

They suggest to apply the method used in their previous papers (already reported in the table):

The measure of statistical quality of attributes (descriptors) which are involved to build up model were estimated by Equation 3: $\text{defect}(A_k)$.

Having the numerical data on the $\text{defect}(A_k)$ one can estimate reliability of the model for a representation of metal oxide nanoparticles by a quasi-SMILES (Table 4 in the publication): the basic hypothesis is "the probability of the quasi-SMILES to be in the domain of applicability is inversely proportional of sum of A_k -defects

$\text{defect}(A_k) < 2 \times \text{average}(\text{Defect-quasi-SMILES})$ is used as indicator to decide if A_k fall inside of the applicability domain.

Results of applicability domain are not presented, but is highlighted that Split2 validation set fall into it.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Carbon-based

List: MWCNTs (Multi-walled carbon nanotubes)

Fullerenes C60

Shape: fiber (MWCNT)

spherical (C60)

Coating: NA

Size (nm): Up to 100 nm of diameter (C60)

Other info: The particle size of the bulk material ranges from

100,000 nm to 3,000,000 nm for MWCNT

The specific surface area of purchased C60 before pulverization with beads was 0.92 m²/g

6.6.Pre-processing of data before modelling:

The data are split into the training, calibration, and validation sets according to the following principles:

(i) the split is random;

(ii) the ranges of endpoint for the above-mentioned sets are similar

(iii) the identity of these splits is minimal.

Three splits built up according to the above-mentioned principles are examined in the present study

6.7.Statistics for goodness-of-fit:

- Training -

Split1:

$r^2 = 0.6031$

$s = 0.043$

Split2:

$r^2 = 0.7772$

$s = 0.035$

Split3:

$r^2 = 0.6269$

$s = 0.044$

- Calibration -

Split1:

$$r^2 = 0.7504$$

$$s = 0.044$$

Split2:

$$r^2 = 0.9147$$

$$s = 0.056$$

Split3:

$$r^2 = 0.7231$$

$$s = 0.047$$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

- Training -

Split1:

$$(c)R^2_p = 0.5798$$

$$q^2 = 0.5260$$

Split2:

$$(c)R^2_p = 0.7518$$

$$q^2 = 0.7451$$

Split3:

$$(c)R^2_p = 0.6137$$

$$q^2 = 0.5571$$

- Test -

Split1:

$$(c)R^2_p = 0.6751$$

Split2:

$$(c)R^2_p = 0.8269$$

Split3:

$$(c)R^2_p = 0.6567$$

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

Split1:

10

Split2:

7

Split3:

10 MCarbon-based

List

MWCNTs (Multi-walled carbon nanotubes)

Fullerenes C60

Shape: fiber (MWCNT)

spherical (C60)

Coating: NA

Size(nm): Up to 100 nm of diameter (C60)

Other properties:

The particle size of the bulk material ranges from

100,000 nm to 3,000,000 nm for MWCNT

The specific surface area of purchased C60 before pulverization with beads was 0.92 m²/g

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Split1:

$r^2 = 0.6429$

$s = 0.051$

Split2:

$r^2 = 0.8341$

$s = 0.091$

Split3:

$r^2 = 0.6951$

$s = 0.044$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The presented model is a combination of two previous models which are already reported in the table

(

Toropov, A. A., & Toropova, A. P. (2015). Quasi-QSAR for mutagenic potential of multi-walled carbon-nanotubes. *Chemosphere*, 124(1), 40–46.

and

Toropov, A. A., & Toropova, A. P. (2014). Optimal descriptor as a translator of eclectic data into endpoint prediction: Mutagenicity of fullerene as a mathematical function of conditions. *Chemosphere*, 104, 262–264.

)

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

The

SMILES: Simplified Molecular Input Line Entry Specification

R²: correlation coefficient

s: root-mean-square error

q²: cross-validation correlation coefficient

CORAL: CORrelation And Logic

(c)R²_p = Parameter computed from correlations coefficient

9.2. Bibliography:

Shinohara, N., Matsumoto, K., Endoh, S., Maru, J., & Nakanishi, J. (2009). In vitro and in vivo genotoxicity tests on fullerene C60 nanoparticles. *Toxicology Letters*, 191(2-3), 289–296. <http://doi.org/10.1016/j.toxlet.2009.09.012>

Wirnitzer, U., Herbold, B., Voetz, M., & Ragot, J. (2009). Studies on the in vitro

genotoxicity of baytubes??, agglomerates of engineered multi-walled carbon-nanotubes (MWCNT). Toxicology Letters, 186(3), 160–165.
<http://doi.org/10.1016/j.toxlet.2008.11.024>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Salmonella typhimurium TA101, QSAR, - X: Fullerene

- Z: MWCNT

- "0": Dark condition

- "1" : Irradiation condition

- "Y": with preincubation

- "N": without preincubation

- "+": Presence of S9

- "-": Absence of S9

- Dose(g/plate) C60:

· A: 50

· B: 100

· C: 200

· D: 400

· E: 1000

- Dose(µg/plate) MWCNT:

· F: 0

· G: 50

· H: 158


· I : 500

· J: 1581

· K: 5000,Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Toxicity model of metal oxide nanoparticles to human keratinocyte cell
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Toxicity model of metal oxide nanoparticles to human keratinocyte cell line (HaCaT) by GA-MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Jerzy Leszczynski

jerzy@icnanotox.org

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–32

<http://doi.org/10.3109/17435390.2014.930195>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human keratinocyte cell line (HaCaT)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as LC50

3.3. Comment on endpoint:

Cell viability was measured using the CytoTox-Glo Cytotoxicity Assay from Promega (Madison, WI). LC50 values for all MeOx were extrapolated using the third order polynomial equation of the log transformed data with the least squares fit in GraphPad (GraphPad Software, Inc., La Jolla, CA)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- $(\Delta H_f)^c$: Standard enthalpy of formation of metal oxide nanocluster
- X^c : Mulliken's electronegativity; 2

4.4.Descriptor selection:

27 parameters quantitatively describing variability of the nanoparticles' structure-nano-descriptors (Table S1 the publication's Supplementary material). These included: 16 quantum-mechanical descriptors (from quantum-chemical calculations, semi-empirical PM6method, implemented in MOPAC 2009 package) and 11 image descriptors (derived from Transmission Electron Microscopy images)

GA (Genetic Algorithm) applied to the model building algorithm

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

10/2

Descriptor: Chemical ratio :2:10 ~ 1:5

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 2B)

$h^* = 0.90$

No outliers were detected.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

10 Metal Oxide

List:

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: Sphericity and circularity

Coating: NA

Size (nm): Average particle size: 15-210

Average particle size in media: 189-2029

Other info: To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images. (For specific details see (in the publication) Table S4 in publication's supplementary material)

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

6.6.Pre-processing of data before modelling:

Cross-validation leave one out and external validation were applied.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.93$$

$$RMSE_c = 0.12$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$$Q^2_{cv} = 0.86$$

$$RMSE_{CV} = 0.16$$

Y-scrambling test was applied. See publication's supplementary material Figure S1

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

8 Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: Sphericity and circularity

Coating: NA

Size(nm): Average particle size: 15-210

Average particle size in media: 189-2029

Other properties:

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images. (For specific details see (in the publication) Table S4 in publication's supplementary material)

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$Q^2_{\text{ext}} = 0.83$

RMSE_P = 0.13

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Good agreement with the OECD framework of QSAR model development.

Not clearly described the obtained results from the Y-scrambling. If the correlation values don't decrease enough with the scrambling of the response variable, a chance correlation can no be discarded. It was discussed through the RMSE, but if a considerable correlation is on it, it could be blamed to a low variability in some of the descriptors and/or endpoint data.

Good and extend Mechanistic interpretation.

GA-MLR: Genetic Algorithm and Multiple Linear Regression

LOO: Leave-One-Out Cross-Validation

R^2 : correlation coefficient

Q^2_{CV} : leave-one-out cross-validation correlation coefficient

Q^2_{ext} : correlation coefficient for external validation

RMSE_C:

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Human keratinocyte cell line (HaCaT), QSAR, - $(\Delta H_f)^c$: Standard enthalpy of formation of metal oxide nanocluster

- X^c : Mulliken's electronegativity, MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictive model of Gold nanoparticles exocytosis on macrophages by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictive model of Gold nanoparticles exocytosis on macrophages by PLSR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Mohammad Reza Hormozi-Nezhad

hormozi@sharif.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Bigdeli, A., Hormozi-Nezhad, M. R., & Parastar, H. (2015). Using nano-QSAR to determine the most responsible factor(s) in gold nanoparticle Exocytosis. RSC Advances, 5(70), 57030–57037. JOUR.

<http://doi.org/10.1039/c5ra06198a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human macrophage-like

U937 cell

3.2. Endpoint:

In vitro - Exocytosis - measured as percentage of observed NPs

3.3. Comment on endpoint:

Undifferentiated (monocytes) or phorbol-12-myristate-13-acetate (PMA)-differentiated U937 cells (macrophages) were treated with each formulation of serum-coated GNPs for 6 h in serum-supplemented media at 37 °C to saturate their endocytosis.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

PLSR: Partial Least Square Regression

by PLS Toolbox v5.8

4.3.Descriptors in the model:

- ChDens_B: Charge density Before protein coating
 - ZP_B: Zeta potential Before protein coating
 - ChAcumm_B: charge accumulation Before protein coating
 - Circle
 - Square
 - ZP_A: Zeta potential After protein coating
 - Corner Count
 - Agg State
 - ChDens_A: Charge density After protein coating
- ; 9

4.4.Descriptor selection:

Within the model building algorithm PLSR, the Variable importance on projections (VIP) was used to see which were the most important descriptors. Also, low SR and regression vector values were applied to ensure the selected variables.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

12/9

Descriptor: Chemical ratio :09:12

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Suggested an applicability domain of GNPs within the range of applied descriptors.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

12 Metal

List: Au

Shape: NA

Coating: Uncoated

Opsonisation by serum proteins

Size (nm): 14 - 51

Other info: GNPs were evaluated before and after an opsonisation by serum protein.

TEM extracted nano-descriptors including size, surface area, aspect ratio, corner count, curvature, aggregation state, and shape performing image processing on the TEM images shown in publication's Figure 1.

6.6.Pre-processing of data before modelling:

Cross-validation leave one out and external validation were applied.

6.7.Statistics for goodness-of-fit:

$R^2_{cal} = 0.971$

RMSEC = 3.456

$R^2_{adj} = 0.78$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

$R^2_{cv} = 0.707$

RMSE_CV = 11.129

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MMetal

List

Au

Shape: NA

Coating: Uncoated

Opsonisation by serum proteins

Size(nm): 14 - 51

Other properties:

GNPs were evaluated before and after an opsonisation by serum protein.

TEM extracted nano-descriptors including size, surface area, aspect ratio, corner count, curvature, aggregation state, and shape performing image

processing on the TEM images shown in publication's Figure 1.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

There is not an external validation procedure, and the data size is not enough to ensure the reliability of the model. Also the final regression function obtained from the PLS technique was not provided, which gives a confusing idea of the final used descriptors.

GNPs: Gold nanoparticles

R^2_{cal} : Correlation coefficient

R^2_{adj} : Adjusted correlation coefficient

LOO: Leave-one-out cross-validation

R^2_{cv} : Cross-correlation leave-one-out correlation coefficient

RMSEC: Root-mean-square error

RMSE_CV: Root-mean-s

9.2.Bibliography:

N. Oh and J. H. Park, Surface chemistry of gold nanoparticles mediates their Exocytosis in macrophages, ACS Nano, 2014, 8, 6232

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Human macrophage-like


U937 cell, QSAR, - ChDens_B: Charge density Before protein coating

- ZP_B: Zeta potential Before protein coating
- ChAcumm_B: charge accumulation Before protein coating
- Circle
- Square
- ZP_A: Zeta potential After protein coating
- Corner Count
- Agg State
- ChDens_A: Charge density After protein coating

,PLSR: Partial Least Square Regression

by PLS Toolbox v5.8

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal oxide toxicity classification by GPTree
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal oxide toxicity classification by GPTree

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Xue Z. Wang

x.z.wang@leeds.ac.uk

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Oksel, C., Winkler, D. A., Ma, C. Y., Wilkins, T., & Wang, X. Z. (2016). Accurate and interpretable nanoSAR models from genetic programming-based decision tree construction approaches. *Nanotoxicology*, 10(7), 1001–1012.

(Case study 1)

<http://doi.org/10.3109/17435390.2016.1161857>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human bronchial epithelial cells (BEAS-2B)

and

Rat alveolar macrophage cells (RAW264.7)

3.2. Endpoint:

In vitro - Cytotoxicity - measured by the the curve of dose-response and consensus Self-Organizing Map clustering on SPS and HTS assay

3.3.Comment on endpoint:

Toxicological responses of twenty-four metal oxide NPs (over a concentration range of 0.39–100 mg L⁻¹) on RAW 264.7 and BEAS-2B cell lines, using both single parameter screening (SPS) assays (MTS, ATP and LDH) and multi-parameter high-throughput screening (HTS) assays (Mito, Fluo4, JC1, and PI over exposure time of 1–24 h)

Toxicity class definition derived based on both dose–response analysis and consensus Self-Organizing Map clustering.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

GPTree: Genetic Program-based decision Tree construction tool

4.3.Descriptors in the model:

- Ec : Computed Valence band energy

- (Z²)/r : Ionic index, where Z and r are the charge number and ionic radius of metal cation in the NP crystals, respectively.; 2

4.4.Descriptor selection:

Within the model building algorithm GPTree. The Genetic algorithm applied with Decision Tree method selects the best trees, then the best descriptors will be chosen till to set conditions are satisfied.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

18/2

Descriptor: Chemical ratio :2:23 ~ 1:12

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper

It should be considered as applicability domain the range of descriptor values of Metal oxide NPs in size range of 10 - 200 nm

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

18 Metal Oxide

List: Al₂O₃

CuO

CeO₂

Co₃O₄

CoO

Cr₂O₃

Fe₂O₃

Fe₃O₄

Gd₂O₃

HfO₂

In₂O₃

La₂O₃

Mn₂O₃

NiO

Ni₂O₃

Sb₂O₃

SiO₂

SnO₂

R-TiO₂

WO₃

Y₂O₃

Yb₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): 10-100

Other info:

exceptions outside range of sizes:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

6.6.Pre-processing of data before modelling:

Nanoparticles were splitted into training and test set in the following way - the splitting of the dataset to training and test sets fulfilled three conditions:

- (1) metal oxides from each activity group should be presented in both training and test sets;
- (2) metal oxides presented in the test set should cover all types of oxides (MeO, Me₂O₃, MeO₂), similarly to the training set;
- (3) the list of oxides in each test set should be identical for both toxicity endpoints.

6.7.Statistics for goodness-of-fit:

Specificity = 100 %

Sensitivity = 100 %

Accuracy = 100 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

4-round Y-scrambling:

Accuracies = 44, 41, 47 and 50 %

No statistically significant obtained models.

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

5 Metal Oxide

List

Al₂O₃

CuO

CeO₂

Co₃O₄

CoO

Cr₂O₃

Fe₂O₃

Fe₃O₄

Gd₂O₃

HfO₂

In₂O₃

La₂O₃

Mn₂O₃

NiO

Ni₂O₃

Sb₂O₃

SiO₂
 SnO₂
 R-TiO₂
 WO₃
 Y₂O₃
 Yb₂O₃
 ZnO
 ZrO₂
Shape:NA
Coating:NA
Size(nm): 10-100
Other properties:

exceptions outside range of sizes:

Cr₂O₃ : 193±90.0 nm and Ni₂O₃ : 140.6±52.5 nm

For specific details, in Crystalline structure information on metal oxide nanoparticles, see Table S1 (supplementary material from source publication)

All of the nanoparticles were provided in powdered form. Transmission electron microscopy (TEM, JEOL 1200 EX, accelerating voltage 80 kV) was used to observe the shapes and primary sizes of the nanoparticles.

X-ray powder diffraction (XRD, Panalytical X'Pert Pro diffractometer, Cu K α radiation) was utilized for identifying the crystal structure of each material.

High-throughput dynamic light scattering (HT-DLS, Dynapro Plate Reader, Wyatt Technology) was performed to determine the particle size and size distribution of the nanoparticles in water and the cell culture media.

Zeta-potential measurement of the nanoparticle suspensions in water was performed using a ZetaPALS instrument (Zeta Potential Analyzer, Brookhaven Instruments Corporation, Holtsville, NY).

Metal dissolution was determined by inductively coupled plasma-mass spectrometry (Perkin-Elmer SCIEX Elan DRCII ICP-MS)

The band gap energies were obtained from diffuse reflectance (DR) UV-vis spectroscopic analysis (Cary 5000 UV-vis-NIR spectrometer equipped with a Praying Mantis accessory). (More details in the publication's section: Materials and Methods - Physicochemical Characterization)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Specificity = 100 %

Sensitivity = 100 %

Accuracy = 100 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

The Model interpretation section ensures that the results are consistent with previous studies source publication and Liu and Rallo et al. 2013

NPs: Nanoparticles

GPTree: Genetic program-based Decision Tree construction tool

9.2. Bibliography:

(already reported in this table)

Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., Liu, R., ... Nel, A. E. (2012). Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano*, 6(5), 4349–4368

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC


10.3. Keywords:

Cell, Human bronchial epithelial cells (BEAS-2B)
and

Rat alveolar macrophage cells (RAW264.7), QSAR, - Ec : Computed Valence band energy

- $(Z^2)/r$: Ionic index, where Z and r are the charge number and ionic radius of metal cation in the NP crystals, respectively., GPTree: Genetic Program-based decision Tree construction tool

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells by GPTree
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells by GPTree
- 1st pool of descriptors -

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Xue Z. Wang

x.z.wang@leeds.ac.uk

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Oksel, C., Winkler, D. A., Ma, C. Y., Wilkins, T., & Wang, X. Z. (2016). Accurate and interpretable nanoSAR models from genetic programming-based decision tree construction approaches. *Nanotoxicology*, 10(7), 1001–1012.

(Case study 2 - 1st pool of descri

<http://doi.org/10.3109/17435390.2016.1161857>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3.Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell
 The cellular uptakes in PaCa2 for the 105 nanoparticles were ranged from 170 to 27 542 nanoparticles per cell. A total of 56 nanoparticles with cellular uptake of more than 5000 nanoparticles per cell were considered to have good/moderate (henceforth referred to as good for brevity) cellular uptake (positive class), while 49 nanoparticles with cellular uptake of less than 5000 nanoparticles per cell were considered to have poor cellular uptake (negative class)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

GPTree: Genetic Program-based decision Tree construction tool

4.3.Descriptors in the model:

- MlogP: Moriguchi octanol-water partition coeff. (Lipophilicity)
- CATS2D_03_AL: CATS2D Acceptor-Lipophilic at lag 03 (Lipophilicity)
- DLS_04: Modified drug-like score (Lipophilicity, H-bonding and molecular weight)
- DLS_cons: DRAGON consensus drug-like score (Lipophilicity, H-bonding and molecular weight)
- AAC: Mean information index on atomic composition (Symmetry associated with structure)
- IDDE: Mean information content on the distance degree equality (Symmetry associated with structure)
- ATSC6m: Centred Broto-Moreau autocorrelation weighted by mass (Atomic masses)
- GGI6: Topological charge index of order 6 (Charge distribution)
- Spmax2Bh(v): Burden largest eigenvalue descriptor weighted by van der Waals volume (Connectivity Index)
- Eig10AEA(ri): Eigenvalue n.10 from edge adjacency mat. weighted by resonance integral (Edge adjacency indices)
- T(N..N): Sum of topological distances between N...N (Connectivity index)
- F04[C-N]: Frequency of C-N at topological distance 4 (Connectivity index); 12

4.4.Descriptor selection:

690 1D and 2D descriptors were calculated using DRAGON 6 software (Mauri et al., 2006). After removing those descriptors with little variation across the NPs, 389 chemical descriptors were retained.

Within the model building algorithm GPTree. The Genetic algorithm applied with Decision Tree method selects the best trees, then the best descriptors will be chosen till to set conditions are satisfied.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

84/12

Descriptor: Chemical ratio :12:84 ~1:7

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of Cross-Linked Iron Oxide (CLIO-NH₂) NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

84 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_mShape: NACoating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione
 7 3,4-Dimethylfuran-2,5-dione
 Hexanoic anhydride
 3-Methyldihydrofuran-2,5-dione
 5,5'-Carbonylbis(2-benzofuran-1,3-dione)
 5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione

4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
 NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
 NC(C(O)=O)c1ccc(Cl)cc1

2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

In order to ensure the quality and accuracy of the data, the 3D structure of each compound was generated by converting the SMILES strings of compounds given in Fourches et al., into 3D structures and then manually inspected and compared with the structures provided by Weissleder et al., 2005. 105 NPs were matched successfully.

6.6.Pre-processing of data before modelling:

The dataset was split into a training set (84 NPs) and test set (21 NPs) that contains NPs distributed across the range of the cellular uptake values

6.7.Statistics for goodness-of-fit:

Specificity = 100 %

Sensitivity = 95 %

Accuracy = 98 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

3-round Y-scrambling:

Accuracies = 39, 44, 55 %

No statistically significant obtained models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

21 Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape:NA

Coating:Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.02,6]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.02,6]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine

2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione

2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

In order to ensure the quality and accuracy of the data, the 3D structure of each compound was generated by converting the SMILES strings of compounds given in Fourches et al., into 3D structures and then manually inspected and compared with the structures provided by Weissleder et al., 2005. 105 NPs were matched successfully.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Specificity = 90 %

Sensitivity = 82 %

Accuracy = 86 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The Model interpretation section ensures that the results are consistent with previous studies Fourches et al. 2010 (already reported in this table)

NPs: Nanoparticles

GPTree: Genetic program-based Decision Tree construction tool

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - MlogP: Moriguchi octanol-water partition coeff. (Lipophilicity)

- CATS2D_03_AL: CATS2D Acceptor-Lipophilic at lag 03 (Lipophilicity)
- DLS_04: Modified drug-like score (Lipophilicity, H-bonding and molecular weight)
- DLS_cons: DRAGON consensus drug-like score (Lipophilicity, H-bonding and molecular weight)
- AAC: Mean information index on atomic composition (Symmetry associated with structure)
- IDDE: Mean information content on the distance degree equality (Symmetry associated with structure)
- ATSC6m: Centred Broto-Moreau autocorrelation weighted by mass (Atomic masses)
- GGI6: Topological charge index of order 6 (Charge distribution)
- Spmax2Bh(v): Burden largest eigenvalue descriptor weighted by van der Waals volume (Connectivity Index)
- Eig10AEA(ri): Eigenvalue n.10 from edge adjacency mat. weighted by resonance integral (Edge adjacency indices)
- T(N..N): Sum of topological distances between N...N (Connectivity index)
- F04[C-N]: Frequency of C-N at topological distance 4 (Connectivity index), GPTree: Genetic Program-based decision Tree construction tool

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells by GPTree
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells by GPTree
- 2nd pool of descriptors -

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Xue Z. Wang

x.z.wang@leeds.ac.uk

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Oksel, C., Winkler, D. A., Ma, C. Y., Wilkins, T., & Wang, X. Z. (2016). Accurate and interpretable nanoSAR models from genetic programming-based decision tree construction approaches. *Nanotoxicology*, 10(7), 1001–1012.

(Case study 2 - 2nd pool of descri

<http://doi.org/10.3109/17435390.2016.1161857>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3.Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell
 The cellular uptakes in PaCa2 for the 105 nanoparticles were ranged from 170 to 27 542 nanoparticles per cell. A total of 56 nanoparticles with cellular uptake of more than 5000 nanoparticles per cell were considered to have good/moderate (henceforth referred to as good for brevity) cellular uptake (positive class), while 49 nanoparticles with cellular uptake of less than 5000 nanoparticles per cell were considered to have poor cellular uptake (negative class)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

GPTree: Genetic Program-based decision Tree construction tool

4.3.Descriptors in the model:

- nN: Number of N atoms (Hydrogen bonding capacity)
- O-058: (atom-centred fragments) =O (Hydrogen bonding capacity)
- SPAM: Average molecular span R
- NCp: Number of terminal primary C(sp³) (Functional group)
- DISPp: Displacement value / weighted by polarizability (Molecular shape and polarizability)
- nHDon: Number of donor atoms for H-bonds (N and O) (Hydrogen bonding capacity)
- ASP: Asphericity (Molecular shape)
- L/Bw Length-to-breadth ratio by WHIM (Molecular shape)
- nSK: Number of non-H atoms (Chemical composition)
- nBT: Number of bonds (Chemical composition)
- nBO: Number of non-H bonds (Degree of unsaturation (hydrogenbonding))
- SCBO: Sum of conventional bond orders (Hdepleted) (Degree of unsaturation (hydrogenbonding))
- G(N...O) Sum of geometrical distances between N..O (Substructure descriptor); 13

4.4.Descriptor selection:

A pool of 147 chemically interpretable descriptors was used (Winkler private communication) (Epa et al., 2012)

Within the model building algorithm GPTree. The Genetic algorithm applied with Decision Tree method selects the best trees, then the best descriptors will be chosen till to set conditions are satisfied.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

84/13

Descriptor: Chemical ratio :13:84 ~ 1:6

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of Cross-Linked Iron Oxide (CLIO-NH₂) NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

84 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione

4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione

Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride

3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione

Dibenz(c,e)oxepin-5,7-dione

6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone

Lauric anhydride

1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid

5-Methyl-2-benzofuran-1,3-dione

4-Nitro-2-benzofuran-1,3-dione

1H-isochromene-1,3(4H)-dione

Dihydro-2H-pyran-2,6(3H)-dione

4,4'-Ethane-1,2-diylmorpholine-2,6-dione

2H-3,1-benzoxazine-2,4(1H)-dione

1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione

4-Methyldihydro-2Hpyran-2,6(3H)-dione

4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione

2,5-Dioxotetrahydrofuran-3,4-diyl diacetate

4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione

Hexahydro-2-benzofuran-1,3-dione

5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione

Iodoacetic anhydride

Chloroacetic anhydride

1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione

Palmitic anhydride

5-amino-1H,3Hbenzo[de]isochromene-1,3-dione

Decanoic anhydride

8-Oxaspiro[4.5]decane-7,9-dione

4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione

1H,3Hbenzo[de]isochromene-1,3-dione

3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(=O)O
 Amino(4-chlorophenyl)acetic acid
NC(C(=O)O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid

2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

In order to ensure the quality and accuracy of the data, the 3D structure of each compound was generated by converting the SMILES strings of compounds given in Fourches et al., into 3D structures and then manually inspected and compared with the structures provided by Weissleder et al., 2005. 105 NPs were matched successfully.

6.6.Pre-processing of data before modelling:

The dataset was split into a training set (84 NPs) and test set (21 NPs) that contains NPs distributed across the range of the cellular uptake values

6.7.Statistics for goodness-of-fit:

Specificity = 98 %

Sensitivity = 100 %

Accuracy = 99 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

3-round Y-scrambling:

Accuracies = 49, 58, 39 %

No statistically significant obtained models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

21 Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape:NA

Coating:Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine

3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione

(2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

In order to ensure the quality and accuracy of the data, the 3D structure of each compound was generated by converting the SMILES strings of compounds given in Fourches et al., into 3D structures and then manually inspected and compared with the structures provided by Weissleder et al., 2005. 105 NPs were matched successfully.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Specificity = 79 %

Sensitivity = 100 %

Accuracy = 86 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The Model interpretation section ensures that the results are consistent with previous studies Epa et al. 2012 (already reported in this table)

NPs: Nanoparticles

GPTree: Genetic program-based Decision Tree construction tool

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - nN: Number of N atoms (Hydrogen bonding capacity)

- O-058: (atom-centred fragments) =O (Hydrogen bonding capacity)
- SPAM: Average molecular span R
- NCp: Number of terminal primary C(sp³) (Functional group)
- DISPp: Displacement value / weighted by polarizability (Molecular shape and polarizability)
- nHDon: Number of donor atoms for H-bonds (N and O) (Hydrogen bonding capacity)
- ASP: Asphericity (Molecular shape)
- L/Bw Length-to-breadth ratio by WHIM (Molecular shape)
- nSK: Number of non-H atoms (Chemical composition)
- nBT: Number of bonds (Chemical composition)
- nBO: Number of non-H bonds (Degree of unsaturation (hydrogenbonding))
- SCBO: Sum of conventional bond orders (Hdepleted) (Degree of unsaturation (hydrogenbonding))
- G(N...O) Sum of geometrical distances between N..O (Substructure descriptor),GPTree: Genetic Program-based decision Tree construction tool

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cytotoxicity keratinocyte (HaCaT) cell line model of metal oxide
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cytotoxicity keratinocyte (HaCaT) cell line model of metal oxide nanoparticles to human by GPTree

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Xue Z. Wang

x.z.wang@leeds.ac.uk

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Oksel, C., Winkler, D. A., Ma, C. Y., Wilkins, T., & Wang, X. Z. (2016). Accurate and interpretable nanoSAR models from genetic programming-based decision tree construction approaches. *Nanotoxicology*, 10(7), 1001–1012.

(Case study 3)

<http://doi.org/10.3109/17435390.2016.1161857>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human keratinocyte cell line (HaCaT)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as LC50

3.3. Comment on endpoint:

Cell viability was measured using the CytoTox-Glo Cytotoxicity Assay from Promega (Madison, WI). LC50 values for all MeOx were extrapolated using the third order polynomial equation of the log transformed data with the least squares fit in GraphPad (GraphPad Software, Inc., La Jolla, CA). Since GPTree can only work with categorical endpoints, 18 NPs were divided into two homogenous clusters, e.g. low toxicity (nine NPs) and high toxicity (nine NPs), based on a threshold value of 2.4. Activity threshold was chosen based on the natural grouping of NPs with balanced distribution between toxic and non-toxic ENMs. There was no object falling near the decision boundary (between 2.32 and 2.48), hence, there was no need to exclude any compounds from the analysis. To ensure the validity of the data split, k-means clustering method was applied using XLSTAT statistic package (Fahmy, 1993). Results of k-means clustering were identical to the results of data split based on a threshold value of 2.4.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

GPTree: Genetic Program-based decision Tree construction tool

4.3.Descriptors in the model:

- $(\Delta H_f)^c$: Standard enthalpy of formation of metal oxide nanocluster
- X^c : Mulliken's electronegativity
- Chemical Hardness: corresponds to the half the band gap of a chemical compound; 3

4.4.Descriptor selection:

29 descriptors (e.g. 16 quantum-mechanical descriptors, 11 image-based descriptors and 2 experimental measurements).

Within the model building algorithm GPTree. The Genetic algorithm applied with Decision Tree method selects the best trees, then the best descriptors will be chosen till to set conditions are satisfied.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

10/3

Descriptor: Chemical ratio :3:10 ~ 1:3

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal oxide NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

10 Metal Oxide

List:

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): Average particle size: 15-210

Average particle size in media: 189-2029

Other info: To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semiempirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

6.6.Pre-processing of data before modelling:

The dataset was split into training (10 NPs) and test (eight NPs) datasets keeping the same NPs in each set as in the source publication.

6.7.Statistics for goodness-of-fit:

Specificity = 100 %

Sensitivity = 100 %

Accuracy = 100 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Y-scrambling:

Accuracies = 39-54 %

No statistically significant obtained models.

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

8 Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size(nm): Average particle size: 15-210

Average particle size in media: 189-2029

Other properties:

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI

CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semiempirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Specificity = 100 %

Sensitivity = 100 %

Accuracy = 100 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The Model interpretation section ensures that the results are consistent with previous studies source publication

NPs: Nanoparticles

GPTree: Genetic program-based Decision Tree construction tool

9.2.Bibliography:

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Cell, Human keratinocyte cell line (HaCaT), QSAR, - $(\Delta H_f)^c$: Standard enthalpy of formation of metal oxide nanocluster

- X^c : Mulliken's electronegativity

- Chemical Hardness: corresponds to the half the band gap of a chemical compound,GPTree: Genetic Program-based decision Tree construction tool

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictive model of Gold nanoparticles exocytosis on macrophages by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictive model of Gold nanoparticles exocytosis on macrophages by GPTree

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Xue Z. Wang

x.z.wang@leeds.ac.uk

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Oksel, C., Winkler, D. A., Ma, C. Y., Wilkins, T., & Wang, X. Z. (2016). Accurate and interpretable nanoSAR models from genetic programming-based decision tree construction approaches. *Nanotoxicology*, 10(7), 1001–1012.

(Case study 4)

<http://doi.org/10.3109/17435390.2016.1161861>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human macrophage-like

U937 cell

3.2. Endpoint:

In vitro - Exocytosis - measured as classification into low, medium or high exocytosis by count of NPs

3.3.Comment on endpoint:

Undifferentiated (monocytes) or phorbol-12-myristate-13-acetate (PMA)-differentiated U937 cells (macrophages) were treated with each formulation of serum-coated Gold Nanoparticles (GNPs) for 6 h in serum-supplemented media at 37 °C to saturate their endocytosis.

The results of Oh & Park (2014) demonstrated that cationic GNPs exhibited the lowest rate of Exocytosis while PEGylated ones showed the highest rate. They also noted that the remaining ones, anionic and zwitterionic GNPs, exhibited medium Exocytosis rates. Based on these findings, we divided 12 GNPs into three homogenous clusters, e.g. low (three GNPs), medium (six GNPs) and high Exocytosis (three GNPs).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

GPTree: Genetic Program-based decision Tree construction tool

4.3.Descriptors in the model:

- Charge Accumulation
- Zeta Potential_B (Before coating)
- Charge Density (Before coating); 3

4.4.Descriptor selection:

Initial descriptors were obtained from (already reported in this table)

Bigdeli, A., Hormozi-Nezhad, M. R., & Parastar, H. (2015). Using nano-QSAR to determine the most responsible factor(s) in gold nanoparticle Exocytosis. RSC Advances, 5(70), 57030–57037. JOUR.

Within the model building algorithm GPTree. The Genetic algorithm applied with Decision Tree method selects the best trees, then the best descriptors will be chosen till to set conditions are satisfied.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

9/3

Descriptor: Chemical ratio :3:9 ~ 1:3

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of Gold NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

9 Metal

List: Au

Shape: NA

Coating: Uncoated

Opsonisation by serum proteins

Size (nm): 14 - 51

Other info: TEM extracted nano-descriptors including size, surface area, aspect ratio, corner count, curvature, aggregation state, and shape performing image processing on the TEM images shown in Figure 1 in Bigdeli et al. 2015 (already reported in this table).

6.6. Pre-processing of data before modelling:

Randomly 1 NP from each class (three) were used from an external validation set.

6.7. Statistics for goodness-of-fit:

Specificity = 100 %

Sensitivity = 100 %

Accuracy = 100 %

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Y-scrambling:

Accuracies = 1-27 %

No statistically significant obtained models.

7. External validation - OECD Principle 4
--

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

3 M Metal

List

Au

Shape: NA

Coating: Uncoated

Opsonisation by serum proteins

Size(nm): 14 - 51

Other properties:

TEM extracted nano-descriptors including size, surface area, aspect ratio, corner count, curvature, aggregation state, and shape performing image processing on the TEM images shown in Figure 1 in Bigdeli et al. 2015 (already reported in this table).

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Specificity = 100 %

Sensitivity = 100 %

Accuracy = 100 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

The Model interpretation section ensures that the results are consistent with previous studies Bigdeli et al. 2015 and N. Oh and J. H. Park, 2014.

NPs: Nanoparticles

GPTree: Genetic program-based Decision Tree construction tool

9.2.Bibliography:

N. Oh and J. H. Park, Surface chemistry of gold nanoparticles mediates their Exocytosis in macrophages, ACS Nano, 2014, 8, 6232

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Human macrophage-like

U937 cell, QSAR, - Charge Accumulation

- Zeta Potential_B (Before coating)

- Charge Density (Before coating),GPTree: Genetic Program-based decision Tree construction tool

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictivity model ZnO nanoparticles effects on Cellular viability by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictivity model ZnO nanoparticles effects on Cellular viability by MLREM

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

Chunying Chen

dave.winkler@csiro.au

chencchy@nanoctr.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Yin, H., Chen, R., Chen, Y., Zhao, L., Casey, P. S., ...

Winkler, D. A. (2016). An Experimental and Computational

Approach to the Development of ZnO Nanoparticles that are Safe

by Design. Small. CSIRO Manufacturing Bayview Avenue Clayton

3168 Aus

<http://doi.org/10.1002/smll.201600597>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human umbilical vein endothelial cells (HUVEC)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as percentage of cellular viability

3.3.Comment on endpoint:

Cellular viability decreased in a typical sigmoidal dose-response manner as the concentration of nanoparticles increased.

The full numerical Cellular viability data are listed in Table S2 (Supporting Information of the publication)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLREM: Multiple Linear Regression with Expectation Maximization

4.3.Descriptors in the model:

- Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures Tc of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio Tc/350.
- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
- Doped percentage: Dopant metal oxide level%
- PMAA: Indicator variable = 1 if PMMA coated, zero if not
- SiO₂: Indicator variable = 1 if silica coated, zero if not
- OA: Indicator variable = 1 if oleic acid coated, zero if not
- Serum: Indicator variable = 1 if serum protein coated, zero if not
- Volume: Calculated nanoparticle volume [nm^3]
- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
- Aspect ratio: Aspect ratio
- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
- Zeta potential: Zeta potential in water [mV]
- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
- RP: Reduction (redox) potential [eV]
- Ec: Conduction band energy [kcal mol^{-1}]; 0

4.4.Descriptor selection:

The initial descriptors were selected from experimental results also from theoretical field.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :NA

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of ZnO NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Metal Oxide

List: ZnO

Shape: Spherical, grains, rods, or needles

Coating: Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size (nm): NA

Other info: Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

6.6.Pre-processing of data before modelling:

A k-means clustering algorithm was used to divide each data set into a training set (80% of the data), used to generate the model, and a test set (20% of the data), used to assess the predictivity of the model

6.7.Statistics for goodness-of-fit:

$r^2 = 0.68$

SEE(%)= 26

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20 % of the training set

(~11) Metal Oxide

List

ZnO

Shape: Spherical, grains, rods, or needles

Coating: Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size(nm): NA

Other properties:

Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

$r^2 = 0.61$

SEP(%)= 23

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

There is not any robustness procedure. The final descriptors seem to be the same as the initial ones, but in MLREM cases having less than 15 weights indicates that there are some descriptors that were deleted within the modelling process.

NPs: Nanoparticles

MLREM: Multiple Linear Regression with Expectation Maximization

r^2 : Correlation coefficient

SEE: standard error of estimation

SEP: standard error of prediction

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Human umbilical vein endothelial cells (HUVEC), QSAR, - Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
- Doped percentage: Dopant metal oxide level%
- PMAA: Indicator variable = 1 if PMMA coated, zero if not
- SiO₂: Indicator variable = 1 if silica coated, zero if not
- OA: Indicator variable = 1 if oleic acid coated, zero if not
- Serum: Indicator variable = 1 if serum protein coated, zero if not
- Volume: Calculated nanoparticle volume [nm^3]
- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
- Aspect ratio: Aspect ratio
- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
- Zeta potential: Zeta potential in water [mV]
- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
- RP: Reduction (redox) potential [eV]
- Ec: Conduction band energy [kcal mol^{-1}], MLREM: Multiple Linear Regression with Expectation Maximization

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictivity model ZnO nanoparticles effects on Cellular viability by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictivity model ZnO nanoparticles effects on Cellular viability by BRANNLP

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

Chunying Chen

dave.winkler@csiro.au

chencchy@nanoctr.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Yin, H., Chen, R., Chen, Y., Zhao, L., Casey, P. S., ...

Winkler, D. A. (2016). An Experimental and Computational

Approach to the Development of ZnO Nanoparticles that are Safe

by Design. Small. CSIRO Manufacturing Bayview Avenue Clayton

3168 Aus

<http://doi.org/10.1002/smll.201600597>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human umbilical vein endothelial cells (HUVEC)

3.2.Endpoint:

In vitro - Cytotoxicity - measured as percentage of cellular viability

3.3.Comment on endpoint:

Cellular viability decreased in a typical sigmoidal dose-response manner as the concentration of nanoparticles increased.

The full numerical Cellular viability data are listed in Table S2 (Supporting Information of the publication)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

BRANNLP: Bayesian Regularization Artificial Neural Network

using a sparse Laplacian prior

4.3.Descriptors in the model:

- Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]

- Doped percentage: Dopant metal oxide level%

- PMAA: Indicator variable = 1 if PMMA coated, zero if not

- SiO₂: Indicator variable = 1 if silica coated, zero if not

- OA: Indicator variable = 1 if oleic acid coated, zero if not

- Serum: Indicator variable = 1 if serum protein coated, zero if not

- Volume: Calculated nanoparticle volume [nm^3]

- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]

- Aspect ratio: Aspect ratio

- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]

- Zeta potential: Zeta potential in water [mV]

- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]

- RP: Reduction (redox) potential [eV]

- Ec: Conduction band energy [kcal mol^{-1}]; 0

4.4.Descriptor selection:

The initial descriptors were selected from experimental results also from theoretical field.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :NA

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of ZnO NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Metal Oxide

List: ZnOShape: Spherical, grains, rods, or needlesCoating: Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size (nm): NA

Other info: Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

6.6.Pre-processing of data before modelling:

A k-means clustering algorithm was used to divide each data set into a training set (80% of the data), used to generate the model, and a test set (20% of the data), used to assess the predictivity of the model

6.7.Statistics for goodness-of-fit:

$r^2 = 0.99$

SEE(%)= 3

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20 % of the training set

(~11) Metal Oxide

List

ZnO

Shape:Spherical, grains, rods, or needles

Coating:Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size(nm): NA

Other properties:

Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2 = 0.89$

SEP(%)= 12

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1. Comments:

There is not any robustness procedure. The final descriptors seem to be the same as the initial ones, but in MLREM cases having less than 15 weights indicates that there are some descriptors that were deleted within the modelling process.

NPs: Nanoparticles

BRANNLP: Bayesian Regularization Artificial Neural Network

using a sparse Laplacian prior

r^2 : Correlation coefficient

SEE: standard error of estimation

SEP: standard error of prediction

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Cell, Human umbilical vein endothelial cells (HUVEC), QSAR, - Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
 - Doped percentage: Dopant metal oxide level%
 - PMAA: Indicator variable = 1 if PMMA coated, zero if not
 - SiO₂: Indicator variable = 1 if silica coated, zero if not
 - OA: Indicator variable = 1 if oleic acid coated, zero if not
 - Serum: Indicator variable = 1 if serum protein coated, zero if not
 - Volume: Calculated nanoparticle volume [nm^3]
 - Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
 - Aspect ratio: Aspect ratio
 - Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
 - Zeta potential: Zeta potential in water [mV]
 - IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
 - RP: Reduction (redox) potential [eV]
 - Ec: Conduction band energy [kcal mol^{-1}], BRANNLP: Bayesian Regularization Artificial Neural Network
- using a sparse Laplacian prior

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictivity model ZnO nanoparticles effects on cell membrane damage
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictivity model ZnO nanoparticles effects on cell membrane damage by MLREM

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

Chunying Chen

dave.winkler@csiro.au

chencchy@nanoctr.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Yin, H., Chen, R., Chen, Y., Zhao, L., Casey, P. S., ...

Winkler, D. A. (2016). An Experimental and Computational

Approach to the Development of ZnO Nanoparticles that are Safe

by Design. Small. CSIRO Manufacturing Bayview Avenue Clayton

3168 Aus

<http://doi.org/10.1002/smll.201600597>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human hepatocellular liver carcinoma cells (HepG2)

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

The amount of lactate dehydrogenase (LDH) released is proportional to the number of cells damaged or lysed and is a useful index for cytotoxicity based on the loss of membrane integrity.

The full experimental data are summarized in Tables S3–S6 (Supporting Information of the publication)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLREM: Multiple Linear Regression with Expectation Maximization

4.3.Descriptors in the model:

- Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.
- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
- Doped percentage: Dopant metal oxide level%
- PMAA: Indicator variable = 1 if PMMA coated, zero if not
- SiO₂: Indicator variable = 1 if silica coated, zero if not
- OA: Indicator variable = 1 if oleic acid coated, zero if not
- Serum: Indicator variable = 1 if serum protein coated, zero if not
- Volume: Calculated nanoparticle volume [nm^3]
- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
- Aspect ratio: Aspect ratio
- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
- Zeta potential: Zeta potential in water [mV]
- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
- RP: Reduction (redox) potential [eV]
- Ec: Conduction band energy [kcal mol^{-1}]; 0

4.4.Descriptor selection:

The initial descriptors were selected from experimental results also from theoretical field.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :NA

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of ZnO NPs within the range of experimental parameters (descriptors) of the training set.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Metal Oxide

List: ZnOShape: Spherical, grains, rods, or needlesCoating: Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size (nm): NA

Other info: Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

6.6.Pre-processing of data before modelling:

A k-means clustering algorithm was used to divide each data set into a training set (80% of the data), used to generate the model, and a test set (20% of the data), used to assess the predictivity of the model

6.7.Statistics for goodness-of-fit:

$r^2 = 0.72$

SEE= 140

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20 % of the training set

(~54) Metal Oxide

List

ZnO

Shape:Spherical, grains, rods, or needles

Coating:Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size(nm): NA

Other properties:

Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2 = 0.57$

SEP= 160

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1. Comments:

There is not any robustness procedure. The final descriptors seem to be the same as the initial ones, but in MLREM cases having less than 15 weights indicates that there are some descriptors that were deleted within the modelling process.

NPs: Nanoparticles

MLREM: Multiple Linear Regression with Expectation Maximization

r^2 : Correlation coefficient

SEE: standard error of estimation

SEP: standard error of prediction

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Cell, Human hepatocellular liver carcinoma cells (HepG2), QSAR, - Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
- Doped percentage: Dopant metal oxide level%
- PMAA: Indicator variable = 1 if PMMA coated, zero if not
- SiO₂: Indicator variable = 1 if silica coated, zero if not
- OA: Indicator variable = 1 if oleic acid coated, zero if not
- Serum: Indicator variable = 1 if serum protein coated, zero if not
- Volume: Calculated nanoparticle volume [nm^3]
- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
- Aspect ratio: Aspect ratio
- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
- Zeta potential: Zeta potential in water [mV]
- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
- RP: Reduction (redox) potential [eV]
- Ec: Conduction band energy [kcal mol^{-1}], MLREM: Multiple Linear Regression with Expectation Maximization

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictivity model ZnO nanoparticles effects on cell membrane damage
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictivity model ZnO nanoparticles effects on cell membrane damage by BRANNLP

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

Chunying Chen

dave.winkler@csiro.au

chencchy@nanoctr.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Yin, H., Chen, R., Chen, Y., Zhao, L., Casey, P. S., ...

Winkler, D. A. (2016). An Experimental and Computational

Approach to the Development of ZnO Nanoparticles that are Safe

by Design. Small. CSIRO Manufacturing Bayview Avenue Clayton

3168 Aus

<http://doi.org/10.1002/smll.201600597>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human hepatocellular liver carcinoma cells (HepG2)

3.2.Endpoint:

In vitro - Cytotoxicity - membrane damage measured as lactate dehydrogenase (LDH) release [units/L]

3.3.Comment on endpoint:

The amount of lactate dehydrogenase (LDH) released is proportional to the number of cells damaged or lysed and is a useful index for cytotoxicity based on the loss of membrane integrity.

The full experimental data are summarized in Tables S3–S6 (Supporting Information of the publication)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

BRANNLP: Bayesian Regularization Artificial Neural Network

using a sparse Laplacian prior

4.3.Descriptors in the model:

- Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]

- Doped percentage: Dopant metal oxide level%

- PMAA: Indicator variable = 1 if PMMA coated, zero if not

- SiO₂: Indicator variable = 1 if silica coated, zero if not

- OA: Indicator variable = 1 if oleic acid coated, zero if not

- Serum: Indicator variable = 1 if serum protein coated, zero if not

- Volume: Calculated nanoparticle volume [nm^3]

- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]

- Aspect ratio: Aspect ratio

- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]

- Zeta potential: Zeta potential in water [mV]

- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]

- RP: Reduction (redox) potential [eV]

- Ec: Conduction band energy [kcal mol^{-1}]; 0

4.4.Descriptor selection:

The initial descriptors were selected from experimental results also from theoretical field.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of ZnO NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: ZnOShape: Spherical, grains, rods, or needlesCoating: Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size (nm): NA

Other info: Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

6.6.Pre-processing of data before modelling:

A k-means clustering algorithm was used to divide each data set into a training set (80% of the data), used to generate the model, and a test set (20% of the data), used to assess the predictivity of the model

6.7.Statistics for goodness-of-fit:

$r^2 = 0.93$

SEE= 60

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20 % of the training set

(~54) Metal Oxide

List

ZnO

Shape:Spherical, grains, rods, or needles

Coating:Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size(nm): NA

Other properties:

Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2 = 0.86$

SEP= 80

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

There is not any robustness procedure. The final descriptors seem to be the same as the initial ones, but in MLREM cases having less than 15 weights indicates that there are some descriptors that were deleted within the modelling process.

NPs: Nanoparticles

BRANNLP: Bayesian Regularization Artificial Neural Network

using a sparse Laplacian prior

r^2 : Correlation coefficient

SEE: standard error of estimation

SEP: standard error of prediction

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC


10.3. Keywords:

Cell, Human hepatocellular liver carcinoma cells (HepG2), QSAR, - Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
- Doped percentage: Dopant metal oxide level%
- PMAA: Indicator variable = 1 if PMMA coated, zero if not
- SiO₂: Indicator variable = 1 if silica coated, zero if not
- OA: Indicator variable = 1 if oleic acid coated, zero if not
- Serum: Indicator variable = 1 if serum protein coated, zero if not
- Volume: Calculated nanoparticle volume [nm^3]
- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
- Aspect ratio: Aspect ratio
- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
- Zeta potential: Zeta potential in water [mV]
- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
- RP: Reduction (redox) potential [eV]
- Ec: Conduction band energy [kcal mol^{-1}], BRANNLP: Bayesian Regularization Artificial Neural Network

using a sparse Laplacian prior

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictivity model ZnO nanoparticles effects on Oxidative Stress by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictivity model ZnO nanoparticles effects on Oxidative Stress by MLREM

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

Chunying Chen

dave.winkler@csiro.au

chencchy@nanoctr.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Yin, H., Chen, R., Chen, Y., Zhao, L., Casey, P. S., ...

Winkler, D. A. (2016). An Experimental and Computational

Approach to the Development of ZnO Nanoparticles that are Safe

by Design. Small. CSIRO Manufacturing Bayview Avenue Clayton

3168 Aus

<http://doi.org/10.1002/smll.201600597>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human hepatocellular liver carcinoma cells (HepG2)

3.2.Endpoint:

In vitro - ROS - measured as Luciferase fold increase

3.3.Comment on endpoint:

Antioxidant-response element (ARE) reporter cells are one of the most reliable and sensitive in vitro methods to measure the oxidative stress response in cells. HepG2 cells and their transformed versions with luciferase reporter plasmid (HepG2-ARE) were used to assess oxidative stress.

Luciferase assays, used as a read out of the response signal to oxidative stress, were reported as fold induction relative to values obtained from untreated control cells.

The full experimental luciferase fold increase data is provided in Tables S7–S12 (Supporting Information of the publication)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLREM: Multiple Linear Regression with Expectation Maximization

4.3.Descriptors in the model:

- Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.
- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
- Doped percentage: Dopant metal oxide level%
- PMAA: Indicator variable = 1 if PMMA coated, zero if not
- SiO₂: Indicator variable = 1 if silica coated, zero if not
- OA: Indicator variable = 1 if oleic acid coated, zero if not
- Serum: Indicator variable = 1 if serum protein coated, zero if not
- Volume: Calculated nanoparticle volume [nm^3]
- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
- Aspect ratio: Aspect ratio
- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
- Zeta potential: Zeta potential in water [mV]
- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
- RP: Reduction (redox) potential [eV]
- Ec: Conduction band energy [kcal mol^{-1}]; 0

4.4.Descriptor selection:

The initial descriptors were selected from experimental results also from theoretical field.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of ZnO NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: ZnOShape: Spherical, grains, rods, or needlesCoating: Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size (nm): NA

Other info: Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

6.6.Pre-processing of data before modelling:

A k-means clustering algorithm was used to divide each data set into a training set (80% of the data), used to generate the model, and a test set (20% of the data), used to assess the predictivity of the model

6.7.Statistics for goodness-of-fit:

$r^2 = 0.67$

SEE= 2.1

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20 % of the training set

(8) Metal Oxide

List

ZnO

Shape:Spherical, grains, rods, or needles

Coating:Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size(nm): NA

Other properties:

Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2 = 0.50$

SEP= 2.7

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

There is not any robustness procedure. The final descriptors seem to be the same as the initial ones, but in MLREM cases having less than 15 weights indicates that there are some descriptors that were deleted within the modelling process.

NPs: Nanoparticles

ROS: Reactive oxygen species. Despite of being a normal product of biological metabolism, ROS levels can increase dramatically under stress conditions (it is also known as oxidative stress), and damage the cell structures.

MLREM: Mult

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Cell, Human hepatocellular liver carcinoma cells (HepG2), QSAR, - Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures Tc of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio Tc/350.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
- Doped percentage: Dopant metal oxide level%
- PMAA: Indicator variable = 1 if PMMA coated, zero if not
- SiO₂: Indicator variable = 1 if silica coated, zero if not
- OA: Indicator variable = 1 if oleic acid coated, zero if not
- Serum: Indicator variable = 1 if serum protein coated, zero if not
- Volume: Calculated nanoparticle volume [nm^3]
- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
- Aspect ratio: Aspect ratio
- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
- Zeta potential: Zeta potential in water [mV]
- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
- RP: Reduction (redox) potential [eV]
- Ec: Conduction band energy [kcal mol^{-1}], MLREM: Multiple Linear Regression with Expectation Maximization

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predictivity model ZnO nanoparticles effects on Oxidative Stress by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predictivity model ZnO nanoparticles effects on Oxidative Stress by BRANNLP

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

Chunying Chen

dave.winkler@csiro.au

chencchy@nanoctr.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Yin, H., Chen, R., Chen, Y., Zhao, L., Casey, P. S., ...

Winkler, D. A. (2016). An Experimental and Computational

Approach to the Development of ZnO Nanoparticles that are Safe

by Design. Small. CSIRO Manufacturing Bayview Avenue Clayton

3168 Aus

<http://doi.org/10.1002/smll.201600597>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human hepatocellular liver carcinoma cells (HepG2)

3.2.Endpoint:

In vitro - ROS - measured as Luciferase fold increase

3.3.Comment on endpoint:

Antioxidant-response element (ARE) reporter cells are one of the most reliable and sensitive in vitro methods to measure the oxidative stress response in cells. HepG2 cells and their transformed versions with luciferase reporter plasmid (HepG2-ARE) were used to assess oxidative stress.

Luciferase assays, used as a read out of the response signal to oxidative stress, were reported as fold induction relative to values obtained from untreated control cells.

The full experimental luciferase fold increase data is provided in Tables S7–S12 (Supporting Information of the publication)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

BRANNLP: Bayesian Regularization Artificial Neural Network

using a sparse Laplacian prior

4.3.Descriptors in the model:

- Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures T_c of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio $T_c/350$.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]

- Doped percentage: Dopant metal oxide level%

- PMAA: Indicator variable = 1 if PMMA coated, zero if not

- SiO₂: Indicator variable = 1 if silica coated, zero if not

- OA: Indicator variable = 1 if oleic acid coated, zero if not

- Serum: Indicator variable = 1 if serum protein coated, zero if not

- Volume: Calculated nanoparticle volume [nm^3]

- Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]

- Aspect ratio: Aspect ratio

- Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]

- Zeta potential: Zeta potential in water [mV]

- IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]

- RP: Reduction (redox) potential [eV]

- Ec: Conduction band energy [kcal mol^{-1}]; 0

4.4.Descriptor selection:

The initial descriptors were selected from experimental results also from theoretical field.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of ZnO NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: ZnO

Shape: Spherical, grains, rods, or needles

Coating: Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size (nm): NA

Other info: Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

6.6.Pre-processing of data before modelling:

A k-means clustering algorithm was used to divide each data set into a training set (80% of the data), used to generate the model, and a test set (20% of the data), used to assess the predictivity of the model

6.7.Statistics for goodness-of-fit:

$r^2 = 0.57$

SEE= 2.2

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20 % of the training set

(8) Metal Oxide

List

ZnO

Shape:Spherical, grains, rods, or needlesCoating:Uncoated

PMMA: poly methyl methacrylate

silica (SiO₂)

oleic acid (OA)

serum protein (from culture media)

Size(nm): NAOther properties:Volume: 7,600 - 137,064,200 nm³

Doping concentrations were measured using inductively coupled plasma-atomic emission spectroscopy (ICP-AES, Varian Vista AX Simultaneous Axial) after digestion. Morphologies of the particles were determined by transmission electron microscopy (TEM, JEOL, 100CX-II, Japan). X-ray photoelectron spectroscopy (XPS) analysis was performed using an AXIS Ultra DLD spectrometer (Kratos Analytical Inc., Manchester, UK). XPS data files were processed using the application CasaXPS software (version 2.3.13).

For specific details see (in the publication) Table S1 in supplementary material to check ZnO nanoparticles properties used for QSAR modelling

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation: $r^2 = 0.67$

SEP= 2.4

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

There is not any robustness procedure. The final descriptors seem to be the same as the initial ones, but in MLREM cases having less than 15 weights indicates that there are some descriptors that were deleted within the modelling process.

NPs: Nanoparticles

ROS: Reactive oxygen species. Despite of being a normal product of biological metabolism, ROS levels can increase dramatically under stress conditions (it is also known as oxidative stress), and damage the cell structures.

BRANNLP: Ba

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Cell, Human hepatocellular liver carcinoma cells (HepG2), QSAR, - Calc temp indicator: Indicator variable for calcination = 0 if the particles were not calcined, or 1, 1.43, or 1.71 for calcination temperatures Tc of 350 °C, 500 °C, or 600 °C. The indicator values were the ratio Tc/350.

- Concentration: Nanoparticle concentration [$\mu\text{g mL}^{-1}$]
 - Doped percentage: Dopant metal oxide level%
 - PMAA: Indicator variable = 1 if PMMA coated, zero if not
 - SiO₂: Indicator variable = 1 if silica coated, zero if not
 - OA: Indicator variable = 1 if oleic acid coated, zero if not
 - Serum: Indicator variable = 1 if serum protein coated, zero if not
 - Volume: Calculated nanoparticle volume [nm^3]
 - Surface area: Calculated nanoparticle surface area [$\text{m}^2 \text{g}^{-1}$]
 - Aspect ratio: Aspect ratio
 - Solubility: Nanoparticle aqueous solubility [$\mu\text{g mL}^{-1}$]
 - Zeta potential: Zeta potential in water [mV]
 - IP: Ionization potential in the relevant metal oxidation state [kJ mol^{-1}]
 - RP: Reduction (redox) potential [eV]
 - Ec: Conduction band energy [kcal mol^{-1}], BRANNLP: Bayesian Regularization Artificial Neural Network
- using a sparse Laplacian prior

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Protein binding to CNTs by consensus model of kNN, SVM and RF
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Protein binding to CNTs by consensus model of kNN, SVM and RF

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Alexander Tropsha

Dr. Bing Yan

alex_tropsha@unc.edu

drbingyan@yahoo.com

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Fourches, D., Pu, D., Li, L., Zhou, H., Mu, Q., Su, G., ... Tropsha, A. (2016). Computer-aided design of carbon nanotubes with the desired bioactivity and safety profiles. *Nanotoxicology*, 10(3), 374–383.

(Protein binding case)

<http://doi.org/10.3109/17435390.2015.1073397>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2.Endpoint:

Protein binding of nanotube to carbonic anhydrase (CA) : F_0/F

3.3.Comment on endpoint:

F_0 is the protein fluorescence intensity before CNT binding and F is the fluorescence intensity after CNT binding

Steady-state fluorescence spectra were acquired on a Hitachi F-7000 spectrofluorometer (Hitachi Co. Ltd., Tokyo, Japan) at 22°C. Carbonic anhydrase (Sigma-Aldrich, St. Louis, MO) (50 µg/mL) was titrated with f-MWNTs stock solutions (500 µg/mL) at three different concentrations (0.0, 7.5 and 15.0 µg/mL). The excitation wavelength was set at 280 nm. Emission fluorescence spectra were recorded between 300 and 400 nm. The scanning speed was 1200 nm/min and the excitation/emission slit was 5.0 nm. The fluorescence intensity of 340 nm was used for F_0/F calculation.

For classification modelling: CNTs are labelled as "0" (non-binder), if their CA bindings are smaller than 2.00 and "1" (binder), if their CA bindings are greater than 2.00.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSPR

4.2.Explicit algorithm:

Consensus classification model of:

- kNN: k-Nearest Neighbour
- SVM: Support Vector Machine
- RF: Random Forest

Those models were applied to two different set of descriptors: Dragon and MOE

Finally, the Consensus model contains 6 models.

4.3.Descriptors in the model:

Final list has not provided for any of the models. A summary of relevant descriptors were published:

- MATS2m: Moran autocorrelation of lag 2 weighted by mass
- H-051: number of hydrogen attached to alpha carbons
- nCl: number of chlorine groups
- several electrotopological indices (maxDN, GATS2m and JGI2)

; 0

4.4.Descriptor selection:

Each CNT has been represented by a single copy of its surface modifying organic molecules encoded as SMILES strings. Following the guidelines published by us recently (Fourches et al., 2010), structural curation procedures were carried out to obtain standardized two-dimensional representations of surface modifiers. The latter were then represented by molecular descriptors computed with Dragon (Mauri et al., 2006) and MOE (Vilar et al., 2008) software including constitutional, functional group counts, atom-centered fragments, walk and path counts, information indices, topological and electrostatic indices.

Descriptors with zero value or zero variance as well as one of each pair of highly inter-correlated descriptors ($R^2 \geq 0.95$) were removed leaving 180 (Dragon) and 158 (MOE) descriptors for further

cheminformatics analysis.

4.5. Algorithm and descriptor generation:

No information available

4.6. Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

83/0

Descriptor: Chemical ratio :NA

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

The dissimilarity between compounds (training and virtual library) was computed using molecular descriptors (Dragon) and Euclidean metric between molecules represented as vectors in high-dimensional descriptor space. If the pair wise Euclidean distances between a compound from the library and any of the modelling set compounds exceeded a pre-defined distance threshold (here, a very strict $z = 0.5$ threshold, see Tropsha & Golbraikh, 2007), this compound was excluded.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

83 Carbon-based

List: CNT: Carbon nanotubes

Shape: Fiber

Coating: c1cccc(c1)C(O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(NC(=O)OC1c2c(c3c1cccc3)cccc2)C(=O)O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(N)C(=O)O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(C)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(Cc1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1ccc(cc1)Cl)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1cccc1)(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1ccc(cc1)C)(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(C)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(Cc1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1ccc(cc1)Cl)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(Cc1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(c1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(Cc1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(c1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(Cc1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(Cc1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(C)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NS(c1ccc(cc1)C(=O)=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NS(c1ccc(cc1)C(=O)=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O

Size (nm): Diameter: 40 ± 10

Length: 250 ± 120

Other info: The surface modifiers were reported by SMILES notation

The number of walls of the nanotube will have little, if any, impact on the conformational behaviour of the surface attached decorator groups. The diameter(s) of the carbon nanotubes were not reported by Zhou, then two different diameters of 1 nm or 1.3 nm of diameter each 6.5 nm in length were defined.

The nanotube–decorator complex was geometry optimized using the molecular dynamics simulation (MDS) package GROMACS (version 4.5.2 for Linux) with the ffgmx force field (a derivative of the GROMOS87 force field).

6.6.Pre-processing of data before modelling:

The whole data was used in the different kind of classification models, each one was evaluated applying the k-fold (k=5) cross-validation method.

The aim of the project was focused into design new CNTs bindings, for this reason 20 compounds of a pool of 240,000 compounds from a virtual library were selected with the ensembled model. Those selected compound were experimentally tested, and the results were provided.

For clustering compounds by chemical similarity, it was employed the sequential agglomerative hierarchical non-overlapping (SAHN, classical Johnson's algorithm) method implemented in the ISIDA/ Cluster program (Varnek et al., 2008)

6.7.Statistics for goodness-of-fit:

C. CCR= 63-75 %

C. Consensus CCR= 74 %

(Acceptable predictive power models over 70% of C.CCR)

C. kNN-Dragon CCR= 75 %

C. RF-Dragon CCR= 73 %

See publication's supplementary, Table 3 A.

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Y-randomization found no statistical significant models.

5-fold cross-validation applied

(C. = cumulative, was presented as goodness of fit)

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20 MCarbon-based

List

CNT: Carbon nanotubes

Shape: Fiber

Coating: c1cccc(c1)C(O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(NC(=O)OC1c2c(c3c1cccc3)cccc2)C(=O)O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NS(c1ccccc1)(=O)=O)=O

c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCC(C(=O)OCC)CC1)=O)NS(c1ccc(cc1)C)(=O)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCC(C(=O)OCC)CC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(C)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(c1ccccc1)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(Cc1ccccc1)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(c1ccc(cc1)Cl)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NS(c1ccccc1)(=O)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NS(c1ccc(cc1)C)(=O)=O
c1cccc(c1)C(=O)C1CCC(CC1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O

Size(nm): Diameter: 40 ± 10

Length: 250 ± 120

Other properties:

The surface modifiers were reported by SMILES notation

The number of walls of the nanotube will have little, if any, impact on the conformational behaviour of the surface attached decorator groups. The diameter(s) of the carbon nanotubes were not reported by Zhou, then two different diameters of 1 nm or 1.3 nm of diameter each 6.5 nm in length were defined.

The nanotube–decorator complex was geometry optimized using the molecular dynamics simulation (MDS) package GROMACS (version 4.5.2 for Linux) with the ffgmx force field (a derivative of the GROMOS87 force field).

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

16/20 *100= 80 %

of accuracy for external CNTs experimentally tested.

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

The most important objective of this study was the prioritization of CNTs for experimental testing and their confirmation to demonstrate how computational screening incorporating similarity search and QSAR models can be used to facilitate the design of novel CNTs with the desired biological properties. To this end, they virtually screen an external compound collection consisting of 240 000 small molecules considered to be synthetically feasible and potentially attachable to the surface of CNTs by the developed QSAR models. Publication's Figure 2 illustrates the workflow of the process.

Although it was reported as QSAR model, the endpoint suggests to label it as QSPR

Despite of 5-fold cross validation is presented as external validation, actually all the data was used to develop the model, then, the statistical results have to be considered as goodness of fit, and robustness of the model. The real external validation was obtained with the final screened molecules.

CNTs: Carbon Nanotubes

kNN: k-Nearest Neighbour

RF: Random Forest

SVM: Support Vector Machine

C. CCR = Cumulative Correct Classification Rate

9.2. Bibliography:

Zhou, H., Mu, Q., Gao, N., Liu, A., Xing, Y., Gao, S., ... Yan, B. (2008). A nano-combinatorial library strategy for the discovery of nanotubes with reduced protein-binding, cytotoxicity, and immune response. *Nano Letters*, 8(3), 859–865.
<http://doi.org/10.1021/nl0730155>

10. Summary (JRC QSAR Model Database)**10.1. QMRF number:**

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, Final list has not provided for any of the models. A summary of relevant descriptors were published:

- MATS2m: Moran autocorrelation of lag 2 weighted by mass
- H-051: number of hydrogen attached to alpha carbons
- nCl: number of chlorine groups

- several electrotopological indices (maxDN, GATS2m and JGI2)

,


Consensus classification model of:

- kNN: k-Nearest Neighbour
- SVM: Support Vector Machine
- RF: Random Forest

Those models were applied to two different set of descriptors: Dragon and MOE

Finally, the Consensus model contains 6 models.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cytotoxicity model of decorated CNTs by consensus model of kNN,
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cytotoxicity model of decorated CNTs by consensus model of kNN, SVM and RF

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Alexander Tropsha

Dr. Bing Yan

alex_tropsha@unc.edu

drbingyan@yahoo.com

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Fourches, D., Pu, D., Li, L., Zhou, H., Mu, Q., Su, G., ... Tropsha, A. (2016). Computer-aided design of carbon nanotubes with the desired bioactivity and safety profiles. *Nanotoxicology*, 10(3), 374–383.

(Cytotoxicity case)

<http://doi.org/10.3109/17435390.2015.1073397>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human Macrophages

3.2.Endpoint:

In vitro - Cytotoxicity - measured as percentage of cellular viability

3.3.Comment on endpoint:

THP-1 (human monocyte) cell lines were cultivated in RPMI 1640 (Gibco) with 10% heat-inactivated fetal bovine serum, 2mM Lglutamine, 100 µg/mL penicillin and 100 U/mL streptomycin and grown in a humidified incubator at 37°C (95% room air, 5% CO₂). Cell differentiation into macrophages was triggered by adding Phorbol 12-myristate 13-acetate (PMA; Promega) at a concentration of 50 ng/mL and incubating for 48 h. Differentiated cells were characterized by allowing them to adhere to the plastic well surface in 96-well plates. The nonadherent monocytes were removed, and the adherent macrophages were washed twice in RPMI 1640. Cells were treated with f-MWNT suspensions (200 µg/mL) in complete culture medium. After 24 h of incubation, a cell proliferation (WST-1) assay was used to determine the Cellular viability. For the whole set of 84 CNTs, the average of measurement variations is as low as 3.6% at (200 µg/mL) and 3.8% at (50 µg/mL).

For classification modelling: CNTs are labelled as "0" (non-toxic) if their associated Cellular viability is greater than 50% and "1" (toxic) if their associated Cellular viability is smaller than 50%

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

Consensus classification model of:

- kNN: k-Nearest Neighbour
- SVM: Support Vector Machine
- RF: Random Forest

Those models were applied to two different set of descriptors: Dragon and MOE

Finally, the Consensus model contains 6 models.

4.3.Descriptors in the model:

Final list has not provided for any of the models. A summary of relevant descriptors were published:

- nCp: number of terminal sp³ carbons
- BAC: Balaban centric index
- nPyrrolidines: number of pyrrolidine substructure

; 0

4.4.Descriptor selection:

Each CNT has been represented by a single copy of its surface modifying organic molecules encoded as SMILES strings. Following the guidelines published by us recently (Fourches et al., 2010), structural curation procedures were carried out to obtain standardized two-dimensional representations of surface modifiers. The latter were then represented by molecular descriptors computed with Dragon (Mauri et al., 2006) and MOE (Vilar et al., 2008) software including constitutional, functional group counts, atom-centered fragments, walk and path counts, information indices, topological and electrostatic indices.

Descriptors with zero value or zero variance as well as one of each pair of highly inter-correlated descriptors ($R^2 \geq 0.95$) were removed leaving 180 (Dragon) and 158 (MOE) descriptors for further

cheminformatics analysis.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

73/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The dissimilarity between compounds (training and virtual library) was computed using molecular descriptors (Dragon) and Euclidean metric between molecules represented as vectors in high-dimensional descriptor space. If the pair wise Euclidean distances between a compound from the library and any of the modelling set compounds exceeded a pre-defined distance threshold (here, a very strict $z = 0.5$ threshold, see Tropsha & Golbraikh, 2007), this compound was excluded.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

73 Carbon-based

List: CNT: Carbon nanotubes

Shape: Fiber

Coating: c1cccc(c1)C(O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(NC(=O)OC1c2c(c3c1cccc3)cccc2)C(=O)O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(N)C(=O)O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(C)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(Cc1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1ccc(cc1)Cl)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1cccc1)(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1ccc(cc1)C)(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(C)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(Cc1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1ccc(cc1)Cl)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(Cc1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(c1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(Cc1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(c1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(Cc1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1cccc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(Cc1cccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1cccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(C)=O)=O


```

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O

```

Size (nm): Diameter: 40 ± 10

Length: 250 ± 120

Other info: The surface modifiers were reported by SMILES notation

The number of walls of the nanotube will have little, if any, impact on the conformational behaviour of the surface attached decorator groups. The diameter(s) of the carbon nanotubes were not reported by Zhou, then two different diameters of 1 nm or 1.3 nm of diameter each 6.5 nm in length were defined.

The nanotube–decorator complex was geometry optimized using the molecular dynamics simulation (MDS) package GROMACS (version 4.5.2 for Linux) with the ffgmx force field (a derivative of the GROMOS87 force field).

6.6.Pre-processing of data before modelling:

From the analysis of cytotoxicity values following a normal distribution, the CNTs around the median cell survive range (37-43%) were excluded, keeping 73.

The data was used in the different kind of classification models, each one was evaluated applying the k-fold (k=5) cross-validation method.

The aim of the project was focused into design new CNTs bindings, for this reason 20 compounds of a pool of 240,000 compounds from a virtual library were selected with the ensembled model. Those selected compound were experimentally tested, and the results were provided.

For clustering compounds by chemical similarity, it was employed the sequential agglomerative hierarchical non-overlapping (SAHN, classical Johnson's algorithm) method implemented in the ISIDA/ Cluster program (Varnek et al., 2008)

6.7. Statistics for goodness-of-fit:

C. CCR= 63-77 %

C. Consensus CCR= 77 %

(Acceptable predictive power models over 70% of C.CCR)

C. kNN-Dragon CCR= 70 %

C. SVM-Dragon CCR= 77 %

C. RF-Dragon CCR= 73 %

See publication's supplementary, Table 3 B.

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Y-randomization found no statistical significant models.

5-fold cross-validation applied (C. =cumulative, was presented as goodness of fit)

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20 MCarbon-based

List

CNT: Carbon nanotubes

Shape:Fiber

Coating:c1cccc(c1)C(O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(NC(=O)OC1c2c(c3c1cccc3)cccc2)C(=O)O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(N)C(=O)O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCC1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1cccc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)NC(OC1c2c(cccc2)c2c1cccc2)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCC1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1cccc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1cccc(c1)C(F)(F)F)=O)N[H])=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(C)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(Cc1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NC(c1ccc(cc1)Cl)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1cccc1)(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1ccc(cc1)C)(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCCCC)=O)NS(c1cccc1)[N+](=O)[O-])(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(C)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(Cc1cccc1)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1cc(ccc1)[N+](=O)[O-])=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NC(c1ccc(cc1)Cl)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1cccc1)(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1ccc(cc1)C)(=O)=O)=O

c1cccc(c1)C(Oc1ccc(cc1)CC(C(N(CCCC)CCCC)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NC1CCCCC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccccc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(NCc1ccccc1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(Cc1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1cc(ccc1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NC(c1ccc(cc1)Cl)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1ccccc1)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1ccc(cc1)C)(=O)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(N1CCCC1)=O)NS(c1cccc(c1)[N+](=O)[O-])(=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(C)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(c1ccccc1)=O)=O
c1cccc(c1)C(Oc1ccc(cc1)CC(C(Nc1ccc(C(=O)OCC)cc1)=O)NC(Cc1ccccc1)=O)=O

c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCC(C(=O)OCC)CC1)=O)NC(c1cc(c1)[N+](=O)[O-])=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCC(C(=O)OCC)CC1)=O)NC(c1ccc(cc1)Cl)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCC(C(=O)OCC)CC1)=O)NS(c1cccc1)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCC(C(=O)OCC)CC1)=O)NS(c1ccc(cc1)C)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCC(C(=O)OCC)CC1)=O)NS(c1cccc(c1)[N+](=O)[O-])=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(C)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(c1cccc1)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(Cc1cccc1)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(c1cc(c1)[N+](=O)[O-])=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NC(c1ccc(cc1)Cl)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NS(c1cccc1)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NS(c1ccc(cc1)C)=O
c1cccc(c1)C(=O)C1CCC(C1)CC(C(NC1CCCC(c1)C(F)(F)F)=O)NS(c1cccc(c1)[N+](=O)[O-])=O

Size(nm): Diameter: 40 ± 10

Length: 250 ± 120

Other properties:

The surface modifiers were reported by SMILES notation

The number of walls of the nanotube will have little, if any, impact on the conformational behaviour of the surface attached decorator groups. The diameter(s) of the carbon nanotubes were not reported by Zhou, then two different diameters of 1 nm or 1.3 nm of diameter each 6.5 nm in length were defined.

The nanotube–decorator complex was geometry optimized using the molecular dynamics simulation (MDS) package GROMACS (version 4.5.2 for Linux) with the ffgmx force field (a derivative of the GROMOS87 force field).

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

17/20 *100= 85 %

of accuracy for external CNTs experimentally tested.

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The most important objective of this study was the prioritization of CNTs for experimental testing and their confirmation to demonstrate how computational screening incorporating similarity search and QSAR models can be used to facilitate the design of novel CNTs with the desired biological properties. To this end, they virtually screen an external compound collection consisting of 240 000 small molecules considered to be synthetically feasible and potentially attachable to the surface of CNTs by the developed QSAR models. Publication's Figure 2 illustrates the workflow of the process.

Despite of 5-fold cross validation is presented as external validation, actually all the data was used to develop the model, then, the statistical results have to be considered as goodness of fit, and robustness of the model. The real external validation was obtained with the final screened molecules.

CNTs: Carbon Nanotubes

kNN: k-Nearest Neighbour

RF: Random Forest

SVM: Support Vector Machine

C. CCR = Cumulative Correct Classification Rate

9.2. Bibliography:

Zhou, H., Mu, Q., Gao, N., Liu, A., Xing, Y., Gao, S., ... Yan, B. (2008). A nano-combinatorial library strategy for the discovery of nanotubes with reduced protein-binding, cytotoxicity, and immune response. *Nano Letters*, 8(3), 859–865.
<http://doi.org/10.1021/nl0730155>

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Human Macrophages, QSAR, Final list has not provided for any of the models. A summary of

relevant descriptors were published:

- nCp: number of terminal sp³ carbons
- BAC: Balaban centric index
- nPyrrolidines: number of pyrrolidine substructure

,


Consensus classification model of:

- kNN: k-Nearest Neighbour
- SVM: Support Vector Machine
- RF: Random Forest

Those models were applied to two different set of descriptors: Dragon and MOE

Finally, the Consensus model contains 6 models.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Size and cytotoxicity of chitosan/streptokinase prediction by ANN
	Printing Date: 30/03/2017

1.QSAR identifier

1.1.QSAR identifier (title):

Size and cytotoxicity of chitosan/streptokinase prediction by ANN

1.2.Other related models:

NA

1.3.Software coding the model:

NA

2.General information

2.1.Date of QMRF:

30/03/2017

2.2.QMRF author(s) and contact details:

LEITAT

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

A. Amani

aamani@sina.tums.ac.ir

2.6.Date of model development and/or publication:

2016

2.7.Reference(s) to main scientific papers and/or software package:

Baharifar, H., & Amani, A. (2016). Cytotoxicity of chitosan/streptokinase nanoparticles as a function of size: An Artificial Neural Networks study. *Nanomedicine: Nanotechnology, Biology, and Medicine*, 12(1), 171–180.

<http://doi.org/10.1016/j.nano.2015.09.002>

2.8.Availability of information about the model:

No information available

2.9.Availability of another QMRF for exactly the same model:

No information available

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Cell

Mrc-5 cell line

3.2.Endpoint:

In vitro - Cytotoxicity - measured as percentage of cellular viability

at a time of NP Size

3.3.Comment on endpoint:

Mrc-5: Diploid human cell culture line composed of fibroblasts derived from lung tissue of an 14 week old aborted caucasian male fetus.

30 samples having different Chitosan (Cs) concentrations (i.e. 0.5-2.0 mg/mL) were prepared in 1% acetic acid solution (pH values 4.5-6.0). To dissolve the polymer, stirring time was set between 0.5 and 2.0 h and stir rate was fixed at 1000 rpm. Then, SK solution (0.1 mg/ml) was added to the solution drop wise and stirred.

To determine size and zeta potential of nanoparticles, photon correlation spectroscopy (PCS, Zeta sizer Nano, Malvern, UK) was used at 25 °C

MTT assay was performed and Cellular viability was obtained following the equation 1

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

ANN: Artificial Neural Network

by INForm v4.02 software

4.3.Descriptors in the model:

- pH
- Cs concentration
- Stirring time; 3

4.4.Descriptor selection:

The initial descriptors were selected from theoretical field and previous studies.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

21/3

Descriptor: Chemical ratio :3:21 ~ 1:7

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of chitosan/streptokinase NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

21 Polymeric

List: Streptokinase (SK) and ChitosanShape: NACoating: NASize (nm): 200-800Other info: To determine size and zeta potential of nanoparticles, photon correlation spectroscopy (PCS, Zeta sizer Nano, Malvern, UK) was used at 25 °C**6.6. Pre-processing of data before modelling:**

NA

6.7. Statistics for goodness-of-fit: $R^2_{\text{train}} = 0.93$ $R^2_{\text{test}} = 0.91$ **6.8. Robustness - Statistics obtained by leave-one-out cross-validation:**

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

7 MPolymeric

List

Streptokinase (SK) and Chitosan

Shape: NA

Coating: NA

Size(nm): 200-800

Other properties:

To determine size and zeta potential of nanoparticles, photon correlation spectroscopy (PCS, Zeta sizer Nano, Malvern, UK) was used at 25 °C

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

$R^2 = 0.94$

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Good procedure, and the parameters of the Neural Network were provided, which allows to reproduce the model. The weakness of the model could come from the lack of a robustness evaluation.

Mechanistic Interpretation. Despite of an used ANN and the abstract of the internally structure of the model, a good interpretation of the results due to fixing parameters to see the response of the other ones was done. Also the study for different descriptors around the size parameter was performed (Predictivity of cytotoxicity and size; Cytotoxicity with size as one of the descriptors; and finally with size as the only parameter. The following classifications in the table include the just mentioned cases)

NPs: Nanoparticles

ANN: Artificial Neural Network

R²: correlation coefficient

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Mrc-5 cell line


, QSAR, - pH

- Cs concentration

- Stirring time, ANN: Artificial Neural Network

by INForm v4.02 software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cytotoxicity of chitosan/streptokinase prediction by ANN
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cytotoxicity of chitosan/streptokinase prediction by ANN

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A. Amani

aamani@sina.tums.ac.ir

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Baharifar, H., & Amani, A. (2016). Cytotoxicity of chitosan/streptokinase nanoparticles as a function of size: An Artificial Neural Networks study. *Nanomedicine: Nanotechnology, Biology, and Medicine*, 12(1), 171–180.

<http://doi.org/10.1016/j.nano.2015.09.002>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Mrc-5 cell line

3.2. Endpoint:

In vitro - Cytotoxicity - measured as percentage of cellular viability

3.3. Comment on endpoint:

Mrc-5: Diploid human cell culture line composed of fibroblasts derived from lung tissue of an 14 week old aborted caucasian male fetus.

30 samples having different Chitosan (Cs) concentrations (i.e. 0.5-2.0 mg/mL) were prepared in 1% acetic acid solution (pH values 4.5-6.0). To dissolve the polymer, stirring time was set between 0.5 and 2.0 h and stir rate was fixed at 1000 rpm. Then, SK solution (0.1 mg/ml) was added to the solution drop wise and stirred.

MTT assay was performed and Cellular viability was obtained following the equation 1

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

ANN: Artificial Neural Network

by INForm v4.02 software

4.3.Descriptors in the model:

- pH
- Cs concentration
- Stirring time
- Size; 4

4.4.Descriptor selection:

The initial descriptors were selected from theoretical field and previous studies, plus the size descriptor, used in the previous model

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

21/4

Descriptor: Chemical ratio :4:21 ~ 1:5

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of chitosan/streptokinase NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

21 Polymeric

List: Streptokinase (SK) and Chitosan

Shape: NA

Coating: NA

Size (nm): 200-800

Other info: To determine size and zeta potential of nanoparticles, photon correlation spectroscopy (PCS, Zeta sizer Nano, Malvern, UK) was used at 25 °C

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

$R^2_{\text{train}} = 0.90$

$R^2_{\text{test}} = 0.96$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

7 MPolymeric

List

Streptokinase (SK) and Chitosan

Shape: NA

Coating: NA

Size(nm): 200-800

Other properties:

To determine size and zeta potential of nanoparticles, photon correlation spectroscopy (PCS, Zeta sizer Nano, Malvern, UK) was used at 25 °C

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

$R^2 = 0.79$

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Good procedure, and the parameters of the Neural Network were provided, which allows to reproduce the model. The weakness of the model could come from the lack of a robustness evaluation.

Mechanistic Interpretation. Despite of an used ANN and the abstract of the internally structure of the model, a good interpretation of the results due to fixing parameters to see the response of the other ones was done. Also the study for different descriptors around the size parameter was performed (Predictivity of cytotoxicity and size; Cytotoxicity with size as one of the descriptors; and finally with size as the only parameter. The following and the previous classifications in the table include the just mentioned cases)

NPs: Nanoparticles

Cs: Chitosan

ANN: Artificial Neural Network

R²: correlation coefficient

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, Mrc-5 cell line

, QSAR, - pH


- Cs concentration

- Stirring time

- Size, ANN: Artificial Neural Network

by INForm v4.02 software

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Cytotoxicity of chitosan/streptokinase prediction by ANN with size as
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Cytotoxicity of chitosan/streptokinase prediction by ANN with size as solely dependent variable

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A. Amani

aamani@sina.tums.ac.ir

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Baharifar, H., & Amani, A. (2016). Cytotoxicity of chitosan/streptokinase nanoparticles as a function of size: An Artificial Neural Networks study. *Nanomedicine: Nanotechnology, Biology, and Medicine*, 12(1), 171–180.

<http://doi.org/10.1016/j.nano.2015.09.002>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Mrc-5 cell line

3.2. Endpoint:

In vitro - Cytotoxicity - measured as percentage of cellular viability

3.3. Comment on endpoint:

Mrc-5: Diploid human cell culture line composed of fibroblasts derived from lung tissue of an 14 week old aborted caucasian male fetus.

30 samples having different Chitosan (Cs) concentrations (i.e. 0.5-2.0 mg/mL) were prepared in 1% acetic acid solution (pH values 4.5-6.0). To dissolve the polymer, stirring time was set between 0.5 and 2.0 h and stir rate was fixed at 1000 rpm. Then, SK solution (0.1 mg/ml) was added to the solution drop wise and stirred.

MTT assay was performed and Cellular viability was obtained following the equation 1

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

ANN: Artificial Neural Network

by INForm v4.02 software

4.3.Descriptors in the model:

- Size; 1

4.4.Descriptor selection:

To test the effect of size in the Cellular viability, they decide to develop a model with only the size as independent parameter.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

21/1

Descriptor: Chemical ratio :01:21

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of chitosan/streptokinase NPs within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

21 Polymeric

List: Streptokinase (SK) and Chitosan

Shape: NA

Coating: NA

Size (nm): 200-800

Other info: To determine size and zeta potential of nanoparticles, photon correlation spectroscopy (PCS, Zeta sizer Nano, Malvern, UK) was used at 25 °C

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

$R^2_{\text{train}} = 0.91$

$R^2_{\text{test}} = 0.87$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

7 MPolymeric

List

Streptokinase (SK) and Chitosan

Shape: NACoating: NASize(nm): 200-800Other properties:

To determine size and zeta potential of nanoparticles, photon correlation spectroscopy (PCS, Zeta sizer Nano, Malvern, UK) was used at 25 °C

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation: $R^2 = 0.80$ **7.8. Predictivity - Assessment of the external validation set:**

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Good procedure, and the parameters of the Neural Network were provided, which allows to reproduce the model. The weakness of the model could come from the lack of a robustness evaluation.

Mechanistic Interpretation. Despite of an used ANN and the abstract of the internally structure of the model, a good interpretation of the results due to fixing parameters to see the response of the other ones was done. Also the study for different descriptors around the size parameter was performed (Predictivity of cytotoxicity and size; Cytotoxicity with size as one of the descriptors; and finally with size as the only parameter. The previous classifications in the table include the just mentioned cases)

NPs: Nanoparticles

Cs: Chitosan

ANN: Artificial Neural Network

R²: correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Cell, Mrc-5 cell line

, QSAR, - Size, ANN: Artificial Neural Network

by INForm v4.02 software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by FT
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by FT
(Data set I)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Nitrifying bacteria

Brachionus plicatilis

Brachionus Calyciflorus

Caenorhabditis elegans

Ceriodaphnia dubia

Daphnia pulex

Fathead

3.2.Endpoint:

In vivo - Ecotoxicological endpoint - measured as LC50

3.3.Comment on endpoint:

400 ENMs from 90 publications or reports provided with experimental data on LC50.

Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.

In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

FT: Functional Tree

by WEKA v3.6

4.3.Descriptors in the model:

- tholepolarizability_a_zz
- volume: Calculate the van der Waals volume of the molecule
- polarsurfacearea: Topological Polar Surface Area calculation (2D)
- SddTi
- SsAg
- SdAg
- Se1Al1Al1
- SsCo
- SdCa
- SsSn
- SsNi
- SsSe
- ALogPS_logP: octanol/water partition coefficient.; 13

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom

and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

320/13

Descriptor: Chemical ratio :13:320 ~ 1:25

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

320 Metal

Metal Oxide

List: Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6. Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7. Statistics for goodness-of-fit:

Sensitivity = 0.750

Specificity = 0.678

Accuracy = 0.709

CCR = 0.714

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

80 Metal

Metal Oxide

List

Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.686

Specificity = 0.733

Accuracy = 0.713

CCR = 0.710

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

FT: Functional Tree

OCHEM: Online Chemical

9.2. Bibliography:

See Supplementary material Table S4 in the publication

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Organism, Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Nitrifying bacteria

Brachionus plicatilis

Brachionus Calyciflorus

Caenorhabditis elegans

Ceriodaphnia dubia

Daphnia pulex

Fathead, QSAR, - tholepolarizability_a_zz


- volume: Calculate the van der Waals volume of the molecule

- polarsurfacearea: Topological Polar Surface Area calculation (2D)

- SddTi

- SsAg
- SdAg
- Se1Al1Al1
- SsCo
- SdCa
- SsSn
- SsNi
- SsSe
- ALogPS_logP: octanol/water partition coefficient.,FT: Functional Tree
by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by C4.5 Decision tree
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by C4.5 Decision tree
(Data set I)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)
Pseudokirchneriella subcapitata (algae)
Nitrifying bacteria
Brachionus plicatilis
Brachionus Calyciflorus
Caenorhabditis elegans
Ceriodaphnia dubia
Daphnia pulex
Fathead

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as LC50

3.3.Comment on endpoint:

400 ENMs from 90 publications or reports provided with experimental data on LC50.
Units of the toxicity values were presented into mg/L. For classification models a threshold must be set to label with Active or Inactive classes the different individuals into the data sets.
In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

C4.5 Decision Tree

by WEKA v3.6

4.3.Descriptors in the model:

- maximalprojectionsize: relates to the size of the molecule perpendicular to the minimal projection area surface (based on the van der Waals radius)
- molecularpolarizability: associates with the polarizability of the molecule; 2

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

320/2

Descriptor: Chemical ratio :2:320 ~ 1:160

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

320 Metal

Metal Oxide

List: Ag

Al

Cu

Co

Fe

Ni

Sn

Ti
 AgO
 Al₂O₃
 CaO
 CeO₂
 CuO
 Cu₂O
 CoO
 Fe₂O₃
 Fe₃O₄
 La₂O₃
 MgO
 Ni₂O₃
 SiO₂
 SnO₂
 ZnO
 TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.671

Specificity = 0.750

Accuracy = 0.716

CCR = 0.711

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

80 MMetal

Metal Oxide

List

Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.686

Specificity = 0.733

Accuracy = 0.713

CCR = 0.710

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

OCHEM: Online Chemical modelling Environment

9.2.Bibliography:

See Supplementary material Table S4 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cells and Organism

, Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Nitrifying bacteria

Brachionus plicatilis

Brachionus Calyciflorus

Caenorhabditis elegans

Ceriodaphnia dubia


Daphnia pulex

Fathead, QSAR, - maximalprojectionsize: relates to the size of the molecule perpendicular to the minimal projection area surface (based on the van der Waals radius)

- molecularpolarizability: associates with the polarizability of the molecule,C4.5 Decision Tree

by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by RT
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by RT
(Data set I)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)
Pseudokirchneriella subcapitata (algae)
Nitrifying bacteria
Brachionus plicatilis
Brachionus Calyciflorus
Caenorhabditis elegans
Ceriodaphnia dubia
Daphnia pulex
Fathead

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as LC50

3.3.Comment on endpoint:

400 ENMs from 90 publications or reports provided with experimental data on LC50.
Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.
In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

RT: Random Tree

by WEKA v3.6

4.3.Descriptors in the model:

- asa_ASA- : solvent accessible surface areas of all atoms with negative partial charge
- asa_ASA_P: solvent accessible surface area of all polar atoms
- exactmass: Exact molecule mass calculation based on the most frequent natural isotopes of the elements
- logd: is the logarithm of the distribution coefficient
- ALogPS_logS: solubility in water
- ALogPS_logP: octanol/water partition coefficient.
- minimalprojectionarea: Calculates the minimal projection area
- maximalprojectionradius: Calculates the maximal projection radius
- wienerindex: Wiener index is a topological index of a molecule, defined as the sum of the lengths of the shortest paths between all pairs of vertices in the chemical graph representing the non-hydrogen atoms in the molecule
- tholepolarizability_a_xx
- dreidingenergy: Calculates the dreiding energy of a conformer of the molecule in kcal/mol
- chainatomcount: Chain atom count

- molecular polarizability: associates with the polarizability of the molecule; 13

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

320/13

Descriptor: Chemical ratio :13:320 ~ 1:25

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

320 Metal

Metal Oxide

List: Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.679

Specificity = 0.728

Accuracy = 0.706

CCR = 0.704

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

80 M Metal

Metal Oxide

List

Ag

Al

Cu

Co

Fe

Ni

Sn
 Ti
 AgO
 Al₂O₃
 CaO
 CeO₂
 CuO
 Cu₂O
 CoO
 Fe₂O₃
 Fe₃O₄
 La₂O₃
 MgO
 Ni₂O₃
 SiO₂
 SnO₂
 ZnO
 TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.629

Specificity = 0.778

Accuracy = 0.713

CCR = 0.704

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

RT: Random Tree

OCHEM: Online Chemical mode

9.2. Bibliography:

See Supplementary material Table S4 in the publication

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cells and Organism

, Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Nitrifying bacteria

Brachionus plicatilis

Brachionus Calyciflorus

Caenorhabditis elegans


Ceriodaphnia dubia

Daphnia pulex

Fathead, QSAR, - asa_ASA- : solvent accessible surface areas of all atoms with negative partial charge

- asa_ASA_P: solvent accessible surface area of all polar atoms
- exactmass: Exact molecule mass calculation based on the most frequent natural isotopes of the elements
- logd: is the logarithm of the distribution coefficient
- ALogPS_logS: solubility in water
- ALogPS_logP: octanol/water partition coefficient.
- minimalprojectionarea: Calculates the minimal projection area
- maximalprojectionradius: Calculates the maximal projection radius
- wienerindex: Wiener index is a topological index of a molecule, defined as the sum of the lengths of the shortest paths between all pairs of vertices in the chemical graph representing the non-hydrogen atoms in the molecule
- tholepolarizability_a_xx
- dreidingenergy: Calculates the dreiding energy of a conformer of the molecule in kcal/mol
- chainatomcount: Chain atom count
- molecularpolarizability: associates with the polarizability of the molecule, RT: Random Tree by WEKA v3.6

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by Simple CART
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by Simple CART
(Data set I)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)
Pseudokirchneriella subcapitata (algae)
Nitrifying bacteria
Brachionus plicatilis
Brachionus Calyciflorus
Caenorhabditis elegans
Ceriodaphnia dubia
Daphnia pulex
Fathead

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as LC50

3.3.Comment on endpoint:

400 ENMs from 90 publications or reports provided with experimental data on LC50.
Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.
In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Simple CART: Classification and Regression Tree
by WEKA v3.6

4.3.Descriptors in the model:

- asa_ASA- : solvent accessible surface areas of all atoms with negative partial charge
- minimalprojectionsize: Calculates the minimal projection size
- maximalprojectionarea: Calculates the minimal projection area
- minimalprojectionradius: Calculates the maximal projection radius
- averagemolecularpolarizability: Average molecular polarizability calculation; 5

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

320/5

Descriptor: Chemical ratio :5:320 ~ 1:64

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

320 Metal

Metal Oxide

List: Ag

Al

Cu

Co

Fe

Ni
 Sn
 Ti
 AgO
 Al₂O₃
 CaO
 CeO₂
 CuO
 Cu₂O
 CoO
 Fe₂O₃
 Fe₃O₄
 La₂O₃
 MgO
 NiO
 SiO₂
 SnO₂
 ZnO
 TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.707

Specificity = 0.678

Accuracy = 0.691

CCR = 0.693

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

80 M Metal

Metal Oxide

List

Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La2O3

MgO

NiO

SiO2

SnO2

ZnO

TiO2

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.686

Specificity = 0.689

Accuracy = 0.688

CCR = 0.688

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)")

should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

CART: Correlation and regression tree

OCHEM

9.2.Bibliography:

See Supplementary material Table S4 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cells and Organism

, Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Nitrifying bacteria

Brachionus plicatilis

Brachionus Calyciflorus

Caenorhabditis elegans

Ceriodaphnia dubia

Daphnia pulex

Fathead, QSAR, - asa_ASA- : solvent accessible surface areas of all atoms with negative partial charge

- minimalprojectionsize: Calculates the minimal projection size


- maximalprojectionarea: Calculates the minimal projection area

- minimalprojectionradius: Calculates the maximal projection radius

- averagemolecularpolarizability: Average molecular polarizability calculation, Simple CART: Classification and Regression Tree

by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by FT
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by FT
(Data set II)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)
Pseudokirchneriella subcapitata (algae)
Anabaena
Bacillus subtilis
Bacillus licheniformis
Brachionus plicatilis
Caenorhabditis elegans
Chlorella
Chlamydomonas reinhardtii

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as EC50

3.3.Comment on endpoint:

450 ENM records from 79 publications or reports with quantitative information on EC50 values. Units of the toxicity values were presented into mg/L. For classification models a threshold must be set to label with Active or Inactive classes the different individuals into the data sets. In that case the threshold was set at 10.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

FT: Functional Tree
by WEKA v3.6

4.3.Descriptors in the model:

- rotatablebondcount: Rotatable bond count
- SsAg
- Se2Ni1O1
- Se1Au1Au1
- SdsDy
- Se1Er2O2ds
- SsFe
- SsAl
- SdsSb; 9

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

360/9

Descriptor: Chemical ratio :9:360 ~1:40

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

360 Metal

Metal Oxide

List: Ag

Al
 Au
 Cu
 Fe
 Ni
 Ti
 Al₂O₃
 CeO₂
 CuO
 CuO/ZnO
 Dy₂O₃
 Fe₂O₃
 Fe₃O₄
 NiO
 Sb₂O₃
 Sm₂O₃
 TiO₂
 ZnO

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.741

Specificity = 0.503

Accuracy = 0.633

CCR = 0.622

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

90 MMetal

Metal Oxide

List

Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃

CeO₂

CuO

CuO/ZnO

Dy₂O₃

Fe₂O₃

Fe₃O₄

NiO

Sb₂O₃

Sm₂O₃

TiO₂

ZnO

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.796

Specificity = 0.415

Accuracy = 0.622

CCR = 0.606

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

FT: Functional

9.2.Bibliography:

See Supplementary material Table S5 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cells and Organism

, Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Anabaena

Bacillus subtilis

Bacillus licheniformis

Brachionus plicatilis

Caenorhabditis elegans

Chlorella

Chlamydomonas reinhardtii, QSAR, - rotatablebondcount: Rotatable bond count

- SsAg

- Se2Ni1O1

- Se1Au1Au1

- SdsDy

- Se1Er2O2ds


- SsFe

- SsAl

- SdsSb,FT: Functional Tree

by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by C4.5 Decision tree
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by C4.5 Decision tree
(Data set II)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)
 Pseudokirchneriella subcapitata (algae)
 Anabaena
 Bacillus subtilis
 Bacillus licheniformis
 Brachionus plicatilis
 Caenorhabditis elegans
 Chlorella
 Chlamydomonas reinhardtii

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as EC50

3.3.Comment on endpoint:

450 ENM records from 79 publications or reports with quantitative information on EC50 values. Units of the toxicity values were presented into mg/L. For classification models a threshold must be set to label with Active or Inactive classes the different individuals into the data sets. In that case the threshold was set at 10.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

C4.5 Decision Tree

by WEKA v3.6

4.3.Descriptors in the model:

- atomgroupcount
- ALogPS_logS: solubility in water
- atomcount: Number of atoms in the molecule
- minimalprojectionarea: Calculates the minimal projection area; 4

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

360/4

Descriptor: Chemical ratio :4:360 ~ 1:90

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

360 Metal

Metal Oxide

List: Ag

Al

Au

Cu

Fe

Ni

Ti
 Al₂O₃
 CeO₂
 CuO
 CuO/ZnO
 Dy₂O₃
 Fe₂O₃
 Fe₃O₄
 NiO
 Sb₂O₃
 Sm₂O₃
 TiO₂
 ZnO

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.695

Specificity = 0.546

Accuracy = 0.628

CCR = 0.621

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

90 MMetal

Metal Oxide

List

Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃CeO₂

CuO

CuO/ZnO

Dy₂O₃Fe₂O₃Fe₃O₄

NiO

Sb₂O₃Sm₂O₃TiO₂

ZnO

Shape:NACoating:NASize(nm): NAOther properties:

Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.816

Specificity = 0.415

Accuracy = 0.633

CCR = 0.616

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a

specified exposure time.

OCHEM: Online C

9.2.Bibliography:

See Supplementary material Table S5 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cells and Organism

, *Danio rerio* (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Anabaena

Bacillus subtilis

Bacillus licheniformis

Brachionus plicatilis

Caenorhabditis elegans

Chlorella

Chlamydomonas reinhardtii, QSAR, - atomgroupcount


- ALogPS_logS: solubility in water

- atomcount: Number of atoms in the molecule

- minimalprojectionarea: Calculates the minimal projection area,C4.5 Decision Tree

by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by RT
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by RT
(Data set II)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)
Pseudokirchneriella subcapitata (algae)
Anabaena
Bacillus subtilis
Bacillus licheniformis
Brachionus plicatilis
Caenorhabditis elegans
Chlorella
Chlamydomonas reinhardtii

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as EC50

3.3.Comment on endpoint:

450 ENM records from 79 publications or reports with quantitative information on EC50 values. Units of the toxicity values were presented into mg/L. For classification models a threshold must be set to label with Active or Inactive classes the different individuals into the data sets. In that case the threshold was set at 10.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

RT: Random Tree

by WEKA v3.6

4.3.Descriptors in the model:

- atomcount: Number of atoms in the molecule
- balabanindex: The Balaban Index, graph index defined for a graph on n nodes and m edges
- molecularsurfacearea: Molecular Surface Area calculation (3D)
- molecularpolarizability: associates with the polarizability of the molecule
- maximalprojectionarea: Calculates the minimal projection area
- minimalprojectionradius: Calculates the minimal projection radius
- maximalprojectionradius: Calculates the maximal projection radius
- minimalprojectionsize: Calculates the minimal projection size
- maximalprojectionsize: relates to the size of the molecule perpendicular to the minimal projection area surface (based on the van der Waals radius)
- ALogPS_logS: solubility in water
- ALogPS_logP: octanol/water partition coefficient.
- logp: the logarithm of the partition coefficient of a molecule observed in a water–n-octanol system which has been adopted as the standard measure of lipophilicity

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

360/12

Descriptor: Chemical ratio :12:95 ~ 1:8

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

360 Metal

Metal Oxide

List: Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃

CeO₂

CuO

CuO/ZnO

Dy₂O₃

Fe₂O₃

Fe₃O₄

NiO

Sb₂O₃

Sm₂O₃

TiO₂

ZnO

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.741

Specificity = 0.479

Accuracy = 0.622

CCR = 0.610

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

90 Metal

Metal Oxide

List

Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃

CeO₂

CuO

CuO/ZnO

Dy₂O₃

Fe₂O₃

Fe₃O₄

NiO

Sb₂O₃

Sm₂O₃

TiO₂

ZnO

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.816

Specificity = 0.439

Accuracy = 0.644

CCR = 0.628

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)")

should be moved to the test set and the external validation data could be interpreted as internal test set with robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

RT: Random Tree

9.2.Bibliography:

See Supplementary material Table S5 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cells and Organism

, Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Anabaena

Bacillus subtilis

Bacillus licheniformis

Brachionus plicatilis

Caenorhabditis elegans

Chlorella

Chlamydomonas reinhardtii, QSAR, - atomcount: Number of atoms in the molecule

- balabanindex: The Balaban Index, graph index defined for a graph on n nodes and m edges

- molecularsurfacearea: Molecular Surface Area calculation (3D)

- molecularpolarizability: associates with the polarizability of the molecule

- maximalprojectionarea: Calculates the minimal projection area

- minimalprojectionradius: Calculates the minimal projection radius

- maximalprojectionradius: Calculates the maximal projection radius

- minimalprojectionsize: Calculates the minimal projection size

- maximalprojectionsize: relates to the size of the molecule perpendicular to the minimal projection area surface (based on the van der Waals radius)

- ALogPS_logS: solubility in water


- ALogPS_logP: octanol/water partition coefficient.

- logp: the logarithm of the partition coefficient of a molecule observed in a water–n-octanol system which has been adopted as the standard measure of lipophilicity

,RT: Random Tree

by WEKA v3.6

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by Simple CART
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by Simple CART
(Data set II)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)
Pseudokirchneriella subcapitata (algae)
Anabaena
Bacillus subtilis
Bacillus licheniformis
Brachionus plicatilis
Caenorhabditis elegans
Chlorella
Chlamydomonas reinhardtii

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as EC50

3.3.Comment on endpoint:

450 ENM records from 79 publications or reports with quantitative information on EC50 values. Units of the toxicity values were presented into mg/L. For classification models a threshold must be set to label with Active or Inactive classes the different individuals into the data sets. In that case the threshold was set at 10.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Simple CART: Classification and Regression Tree
by WEKA v3.6

4.3.Descriptors in the model:

- asa_ASA_H
- minimalprojectionsize: Calculates the minimal projection size
- minimalprojectionarea: Calculates the minimal projection area
- tholepolarizability_a_xx
- tholepolarizability_a_yy
- averagemolecularpolarizability: Average molecular polarizability calculation; 7

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

360/7

Descriptor: Chemical ratio :7:95 ~ 1:14

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

360 Metal

Metal Oxide

List: Ag

Al

Au

Cu

Fe
 Ni
 Ti
 Al₂O₃
 CeO₂
 CuO
 CuO/ZnO
 Dy₂O₃
 Fe₂O₃
 Fe₃O₄
 NiO
 Sb₂O₃
 Sm₂O₃
 TiO₂
 ZnO

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.650

Specificity = 0.564

Accuracy = 0.611

CCR = 0.607

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

90 MMetal

Metal Oxide

List

Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃

CeO₂

CuO

CuO/ZnO

Dy₂O₃

Fe₂O₃

Fe₃O₄

NiO

Sb₂O₃

Sm₂O₃

TiO₂

ZnO

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.796

Specificity = 0.439

Accuracy = 0.633

CCR =0.618

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

CART: Correlati

9.2.Bibliography:

See Supplementary material Table S5 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cells and Organism

, Danio rerio (fish) - Zebrafish

Daphnia magna (crustacean)

Escherichia coli (bacteria)

Pseudokirchneriella subcapitata (algae)

Anabaena

Bacillus subtilis

Bacillus licheniformis

Brachionus plicatilis

Caenorhabditis elegans

Chlorella

Chlamydomonas reinhardtii, QSAR, - asa_ASA_H

- minimalprojectionsize: Calculates the minimal projection size

- minimalprojectionarea: Calculates the minimal projection area


- tholepolarizability_a_xx

- tholepolarizability_a_yy

- averagemolecularpolarizability: Average molecular polarizability calculation, Simple CART: Classification and Regression Tree

by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by FT
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by FT
(Data set III)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Escherichia coli (bacteria)

Staphylococcus aureus (bacteria)

Azotobacter vinelandii
 Enterococcus faecalis
 Nitrosomonas europaea
 Neisseria gonorrhoeae
 Pseudomonas aeruginosa
 Klebsiella pneumoniae ESBL
 Salmonella
 Staphylococcus epidermidis methicillinsuscep

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as MIC

3.3.Comment on endpoint:

166 ENMs obtained from 13 publications with experimental values of the MIC. MIC characterizes the antimicrobial properties of ENMs.

Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.

In that case the threshold was set at 10.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

FT: Functional Tree

by WEKA v3.6

4.3.Descriptors in the model:

- minimalprojectionarea: Calculates the minimal projection area
- asa_ASA_H
- balabanindex: The Balaban Index, graph index defined for a graph on n nodes and m edges
- hararyindex: The Harary index of a graph G on n vertices was defined by Plavšić et al. (1993)
- asa_ASA+: solvent accessible surface areas of all atoms with positive partial charge
- SsCu
- ALogPS_logP: octanol/water partition coefficient.
- ALogPS_logS: solubility in water
- averagemolecularpolarizability: Average molecular polarizability calculation; 9

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5. Algorithm and descriptor generation:

No information available

4.6. Software name and version for descriptor generation:

No information available

4.7. Chemicals/Descriptors ratio:

133/9

Descriptor: Chemical ratio :9:133 ~ 1:15

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

133 Metal

Metal Oxide

List: Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.743

Specificity = 0.762

Accuracy = 0.752

CCR = 0.753

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable
MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

33 MMetal

Metal Oxide

List

Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.706

Specificity = 0.688

Accuracy = 0.697

CCR = 0.697

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

MIC: Minimum inhibitory concentration

FT: Functional Tree

OCHEM: Online Chemical modelling Environment

9.2. Bibliography:

See Supplementary material Table S6 in the publication

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cells and Organism

, Escherichia coli (bacteria)

Staphylococcus aureus (bacteria)

Azotobacter vinelandii

Enterococcus faecalis

Nitrosomonas europaea

Neisseria gonorrhoeae

Pseudomonas aeruginosa


Klebsiella pneumoniae ESBL

Salmonella

Staphylococcus epidermidis methicillinsuscep, QSAR, - minimalprojectionarea: Calculates the minimal projection area

- asa_ASA_H
- balabanindex: The Balaban Index, graph index defined for a graph on n nodes and m edges
- hararyindex: The Harary index of a graph G on n vertices was defined by Plavšić et al. (1993)
- asa_ASA+: solvent accessible surface areas of all atoms with positive partial charge
- SsCu
- ALogPS_logP: octanol/water partition coefficient.
- ALogPS_logS: solubility in water
- averagemolecularpolarizability: Average molecular polarizability calculation, FT: Functional Tree by WEKA v3.6

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by C4.5 Decision tree
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by C4.5 Decision tree
(Data set III)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Escherichia coli (bacteria)

Staphylococcus aureus (bacteria)

Azotobacter vinelandii
 Enterococcus faecalis
 Nitrosomonas europaea
 Neisseria gonorrhoeae
 Pseudomonas aeruginosa
 Klebsiella pneumoniae ESBL
 Salmonella
 Staphylococcus epidermidis methicillinsuscep

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as MIC

3.3.Comment on endpoint:

166 ENMs obtained from 13 publications with experimental values of the MIC. MIC characterizes the antimicrobial properties of ENMs.

Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.

In that case the threshold was set at 10.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

C4.5 Decision Tree

by WEKA v3.6

4.3.Descriptors in the model:

- ALogPS_logS: solubility in water; 1

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

133/1

Descriptor: Chemical ratio :1:133

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

133 Metal

Metal Oxide

List: Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.743

Specificity = 0.778

Accuracy = 0.759

CCR = 0.761

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

33 Metal

Metal Oxide

List

Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂Shape:NACoating:NASize(nm): NAOther properties:

Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.706

Specificity = 0.688

Accuracy = 0.697

CCR = 0.697

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be moved to the test set and the external validation data could be interpreted as test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

MIC: Minimum inhibitory concentration

OCHEM

9.2. Bibliography:

See Supplementary material Table S6 in the publication

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cells and Organism

, Escherichia coli (bacteria)

Staphylococcus aureus (bacteria)

Azotobacter vinelandii

Enterococcus faecalis

Nitrosomonas europaea

Neisseria gonorrhoeae

Pseudomonas aeruginosa


Klebsiella pneumoniae ESBL

Salmonella

Staphylococcus epidermidis methicillinsuscep, QSAR, - ALogPS_logS: solubility in water, C4.5 Decision Tree

by WEKA v3.6

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by RT
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by RT
(Data set III)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Escherichia coli (bacteria)

Staphylococcus aureus (bacteria)

Azotobacter vinelandii
 Enterococcus faecalis
 Nitrosomonas europaea
 Neisseria gonorrhoeae
 Pseudomonas aeruginosa
 Klebsiella pneumoniae ESBL
 Salmonella
 Staphylococcus epidermidis methicillinsuscep

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as MIC

3.3.Comment on endpoint:

166 ENMs obtained from 13 publications with experimental values of the MIC. MIC characterizes the antimicrobial properties of ENMs.

Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.

In that case the threshold was set at 10.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

RT: Random Tree

by WEKA v3.6

4.3.Descriptors in the model:

- tholepolarizability_a_yy
 - ALogPS_logP: solubility in water
 - maximalprojectionarea: Calculates the maximal projection area
 - chainatomcount: Chain atom count
- ; 4

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

133/4

Descriptor: Chemical ratio :4:133 ~ 1:33

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

133 Metal

Metal Oxide

List: Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.814

Specificity = 0.587

Accuracy = 0.707

CCR = 0.701

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA
 NP size: NA
 NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

33 Metal
 Metal Oxide
List
 Ag
 Cu
 Fe
 CuO
 Fe₃O₄
 ZnO
 TiO₂
Shape: NA
Coating: NA
Size(nm): NA
Other properties:

Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.706
 Specificity = 0.688
 Accuracy = 0.697
 CCR = 0.697

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

RT: Random Tree

OCHEM: Online Chemical mode

9.2.Bibliography:

See Supplementary material Table S6 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cells and Organism

, Escherichia coli (bacteria)

Staphylococcus aureus (bacteria)

Azotobacter vinelandii

Enterococcus faecalis

Nitrosomonas europaea

Neisseria gonorrhoeae

Pseudomonas aeruginosa

Klebsiella pneumoniae ESBL

Salmonella


Staphylococcus epidermidis methicillinsuscep, QSAR, - tholepolarizability_a_yy

- ALogPS_logP: solubility in water
- maximalprojectionarea: Calculates the maximal projection area
- chainatomcount: Chain atom count

,RT: Random Tree

by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification by Simple CART
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification by Simple CART
(Data set III)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cells and Organism

Escherichia coli (bacteria)

Staphylococcus aureus (bacteria)

Azotobacter vinelandii
 Enterococcus faecalis
 Nitrosomonas europaea
 Neisseria gonorrhoeae
 Pseudomonas aeruginosa
 Klebsiella pneumoniae ESBL
 Salmonella
 Staphylococcus epidermidis methicillinsuscep

3.2.Endpoint:

In vivo and In vitro - Ecotoxicological endpoint - measured as MIC

3.3.Comment on endpoint:

166 ENMs obtained from 13 publications with experimental values of the MIC. MIC characterizes the antimicrobial properties of ENMs.

Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.

In that case the threshold was set at 10.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Simple CART: Classification and Regression Tree

by WEKA v3.6

4.3.Descriptors in the model:

- ALogPS_logS: solubility in water; 1

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, three types of descriptors were calculated and acquired, the E-state, ALogPS, and Chemaxon descriptors. For the E-state, both atom and bond types were considered for the indices and counts descriptors during calculation. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

133/1

Descriptor: Chemical ratio :1:133

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

133 Metal

Metal Oxide

List: Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.743

Specificity = 0.778

Accuracy = 0.759

CCR = 0.761

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

33 Metal

Metal Oxide

List

Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂Shape:NACoating:NASize(nm): NAOther properties:

Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.706

Specificity = 0.688

Accuracy = 0.697

CCR = 0.697

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be moved to the test set and the external validation data could be interpreted as test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

CART: Correlation and regression tree

OCHEM

9.2. Bibliography:

See Supplementary material Table S6 in the publication

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cells and Organism

, Escherichia coli (bacteria)

Staphylococcus aureus (bacteria)

Azotobacter vinelandii

Enterococcus faecalis

Nitrosomonas europaea

Neisseria gonorrhoeae

Pseudomonas aeruginosa

Klebsiella pneumoniae ESBL


Salmonella

Staphylococcus epidermidis methicillin-suscep, QSAR, - ALogPS_logS: solubility in water, Simple

CART: Classification and Regression Tree

by WEKA v3.6

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification on Danio rerio by FT
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification on Danio rerio by FT

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Danio rerio (fish) - Zebrafish

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as LC50

3.3. Comment on endpoint:

Danio rerio (94 records) cases were selected from dataset I.

Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.

In that case the threshold was set at 100.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

FT: Functional Tree

by WEKA v3.6

4.3.Descriptors in the model:

- Averagemolecularpolarizability: Average molecular polarizability calculation
- molecularpolarizability: associates with the polarizability of the molecule
- mass: Molecule mass calculation
- volume: Calculate the van der Waals volume of the molecule
- plattindex: The Platt index
- apKb1: dissociation constant
- ALogPS_logS: solubility in water; 7

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, two types of descriptors were calculated and acquired ALogPS, and Chemaxon descriptors. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

76/7

Descriptor: Chemical ratio :7:76 ~ 1:11

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

76 Metal

Metal Oxide

List: Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material.

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.943

Specificity = 0.913

Accuracy = 0.934

CCR = 0.928

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No
Chemical Name: not applicable
SMILES: not applicable
Formula: not applicable
INChI: not applicable
MOL file: not applicable

Part extended for NPs.

NP composition: NA
NP size:NA
NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

18 MMetal
Metal Oxide

List

Ag
Al
Cu
Co
Fe
Ni
Sn
Ti
AgO
Al₂O₃
CaO
CeO₂
CuO
Cu₂O
CoO
Fe₂O₃
Fe₃O₄
La₂O₃
MgO
Ni₂O₃
SiO₂
SnO₂
ZnO
TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material.

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 1.000

Specificity = 1.000

Accuracy = 1.000

CCR = 1.000

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

FT: Functional Tree

OCHEM: Online Chemical

9.2. Bibliography:

See Supplementary material Table 4, 5 and 6 in the publication

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Organism, *Danio rerio* (fish) - Zebrafish, QSAR, - Averagemolecularpolarizability: Average molecular polarizability calculation

- molecularpolarizability: associates with the polarizability of the molecule
 - mass: Molecule mass calculation
 - volume: Calculate the van der Waals volume of the molecule
 - plattindex: The Platt index
 - apKb1: dissociation constant
 - ALogPS_logS: solubility in water, FT: Functional Tree
- by WEKA v3.6

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification on Danio rerio by C4.5
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification on Danio rerio by C4.5 Decision tree

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Danio rerio (fish) - Zebrafish

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as LC50

3.3. Comment on endpoint:

Danio rerio (94 records) cases were selected from dataset I.

Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.

In that case the threshold was set at 100.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

C4.5 Decision Tree

by WEKA v3.6

4.3.Descriptors in the model:

- Exactmass: Exact molecule mass calculation based on the most frequent natural isotopes of the elements

- asa_ASA: solvent accessible surface areas of all atoms; 2

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, two types of descriptors were calculated and acquired ALogPS, and Chemaxon descriptors. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

76/2

Descriptor: Chemical ratio :2:76 ~ 1:38

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

76 Metal

Metal Oxide

List: Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni2O3

SiO2

SnO2

ZnO

TiO2

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.906

Specificity = 0.913

Accuracy = 0.908

CCR = 0.910

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

18 MMetal

Metal Oxide

List

Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary

material

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 1.000

Specificity = 1.000

Accuracy = 1.000

CCR = 1.000

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

OCHEM: Online Chemical Modelling Environment

9.2.Bibliography:

See Supplementary material Table 4, 5 and 6 in the publication

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Organism, Danio rerio (fish) - Zebrafish, QSAR, - Exactmass: Exact molecule mass calculation based on the most frequent natural isotopes of the elements

- asa_ASA: solvent accessible surface areas of all atoms,C4.5 Decision Tree
by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification on Daphnia Magna by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification on Daphnia Magna by FT

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Daphnia magna (crustacean)

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as LC50

3.3. Comment on endpoint:

Daphnia magna (102 records) cases were selected from dataset I.
 Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.
 In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

FT: Functional tree

by WEKA v3.6

4.3.Descriptors in the model:

- Molecularpolarizability: associates with the polarizability of the molecule
- tholepolarizability_a_xx
- tholepolarizability_a_zz
- exactmass: Exact molecule mass calculation based on the most frequent natural isotopes of the elements
- volume: Calculate the van der Waals volume of the molecule
- logp: the logarithm of the partition coefficient of a molecule observed in a water–n-octanol system which has been adopted as the standard measure of lipophilicity
- asa_ASA+ : solvent accessible surface areas of all atoms with positive partial charge
- asa_ASA_P; 8

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, two types of descriptors were calculated and acquired ALogPS, and Chemaxon descriptors. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

82/8

Descriptor: Chemical ratio :8:82 ~ 1:10

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

82 Metal

Metal Oxide

List: Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO
 Cu₂O
 CoO
 Fe₂O₃
 Fe₃O₄
 La₂O₃
 MgO
 Ni₂O₃
 SiO₂
 SnO₂
 ZnO
 TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.843

Specificity = 0.968

Accuracy = 0.890

CCR = 0.906

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20 MMetal

Metal Oxide

List

Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.750

Specificity = 1.000

Accuracy = 0.850

CCR = 0.875

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

FT: Functional Tree

OCHEM: Online Chemical

9.2.Bibliography:

See Supplementary material Table 4, 5 and 6 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Organism, Daphnia magna (crustacean), QSAR, - Molecularpolarizability: associates with the polarizability of the molecule

- tholepolarizability_a_xx

- tholepolarizability_a_zz

- exactmass: Exact molecule mass calculation based on the most frequent natural isotopes of the elements

- volume: Calculate the van der Waals volume of the molecule


- logp: the logarithm of the partition coefficient of a molecule observed in a water–n-octanol system which has been adopted as the standard measure of lipophilicity

- asa_ASA+ : solvent accessible surface areas of all atoms with positive partial charge

- asa_ASA_P,FT: Functional tree

by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification on Daphnia Magna by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification on Daphnia Magna by C4.5 Decision tree

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Daphnia magna (crustacean)

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as LC50

3.3. Comment on endpoint:

Daphnia magna (102 records) cases were selected from dataset I.
 Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.
 In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

C4.5 Decision Tree

by WEKA v3.6

4.3.Descriptors in the model:

- asa_ASA- :solvent accessible surface areas of all atoms with negative partial charge; 1

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, two types of descriptors were calculated and acquired ALogPS, and Chemaxon descriptors. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

82/1

Descriptor: Chemical ratio :1:82

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

82 Metal

Metal Oxide

List: Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.843

Specificity = 0.968

Accuracy = 0.890

CCR = 0.906

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

20 MMetal

Metal Oxide

List

Ag

Al

Cu

Co

Fe

Ni

Sn

Ti

AgO

Al₂O₃

CaO

CeO₂

CuO

Cu₂O

CoO

Fe₂O₃

Fe₃O₄

La₂O₃

MgO

Ni₂O₃

SiO₂

SnO₂

ZnO

TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 90 different publications.

For specific details see (in the publication) Table S4 from Supplementary material

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.750

Specificity = 1.000

Accuracy = 0.850

CCR = 0.875

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

LC50: for a substance is the dose required to kill half the members of a tested population after a specified test duration.

OCHEM: Online Chemical Modelling Environment

9.2.Bibliography:

See Supplementary material Table 4, 5 and 6 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Organism, Daphnia magna (crustacean), QSAR, - asa_ASA- :solvent accessible surface areas of all atoms with negative partial charge,C4.5 Decision Tree

by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification on Pseudokirchneriella
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification on Pseudokirchneriella subcapitata by FT

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Pseudokirchneriella subcapitata (algae)

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as EC50

3.3. Comment on endpoint:

Pseudokirchneriella subcapitata (66 records) cases were selected from dataset II. Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets. In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

FT: Functional tree

by WEKA v3.6

4.3.Descriptors in the model:

- Molecularpolarizability: associates with the polarizability of the molecule
- tholepolarizability_a_yy
- mass: Molecule mass calculation
- minimalprojectionarea: Calculates the minimal projection area
- volume: Calculate the van der Waals volume of the molecule
- dreidingenergy: Calculates the dreiding energy of a conformer of the molecule in kcal/mol
- hyperwienerindex: Hyper Wiener index
- ALogPS_logS: solubility in water; 8

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, two types of descriptors were calculated and acquired ALogPS, and Chemaxon descriptors. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

53/8

Descriptor: Chemical ratio :8:53 ~ 1:7

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the

same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

53 Metal

Metal Oxide

List: Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃

CeO₂

CuO

CuO/ZnO

Dy₂O₃

Fe₂O₃

Fe₃O₄

NiO
Sb₂O₃
Sm₂O₃
TiO₂
ZnO

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material.

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.944

Specificity = 0.914

Accuracy = 0.925

CCR = 0.929

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

13 MMetal

Metal Oxide

List

Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃

CeO₂

CuO

CuO/ZnO

Dy₂O₃

Fe₂O₃

Fe₃O₄

NiO

Sb₂O₃

Sm₂O₃

TiO₂

ZnO

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material.

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of

the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.750

Specificity = 1.000

Accuracy = 0.923

CCR = 0.875

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

FT: Functional

9.2.Bibliography:

See Supplementary material Table 4, 5 and 6 in the publication

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Organism, *Pseudokirchneriella subcapitata* (algae), QSAR, - Molecularpolarizability: associates with the polarizability of the molecule

- tholepolarizability_a_yy
- mass: Molecule mass calculation
- minimalprojectionarea: Calculates the minimal projection area
- volume: Calculate the van der Waals volume of the molecule
- dreidingenergy: Calculates the dreiding energy of a conformer of the molecule in kcal/mol
- hyperwienerindex: Hyper Wiener index
- ALogPS_logS: solubility in water,FT: Functional tree
by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification on Pseudokirchneriella
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification on Pseudokirchneriella subcapitata by C4.5 Decision tree

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software

package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Organism

Pseudokirchneriella subcapitata (algae)

3.2. Endpoint:

In vivo - Ecotoxicological endpoint - measured as EC50

3.3.Comment on endpoint:

Pseudokirchneriella subcapitata (66 records) cases were selected from dataset II.
Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.
In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

C4.5 Decision Tree

by WEKA v3.6

4.3.Descriptors in the model:

- Minimalprojectionarea: Calculates the minimal projection area; 1

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, two types of descriptors were calculated and acquired ALogPS, and Chemaxon descriptors. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

53/1

Descriptor: Chemical ratio :1:53

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

53 Metal

Metal Oxide

List: Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃

CeO₂

CuO

CuO/ZnO

Dy₂O₃

Fe₂O₃

Fe₃O₄

NiO

Sb₂O₃

Sm₂O₃

TiO₂

ZnO

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material.

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.944

Specificity = 0.914

Accuracy = 0.925

CCR = 0.929

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

13 MMetal

Metal Oxide

List

Ag

Al

Au

Cu

Fe

Ni

Ti

Al₂O₃

CeO₂

CuO

CuO/ZnO

Dy₂O₃

Fe₂O₃

Fe₃O₄

NiO

Sb₂O₃

Sm₂O₃

TiO₂

ZnO

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 79 different publications.

For specific details see (in the publication) Table S5 from Supplementary material.

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.750

Specificity = 1.000

Accuracy = 0.923

CCR = 0.875

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

EC50 : concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after a specified exposure time.

OCHEM: Online C

9.2.Bibliography:

See Supplementary material Table 4, 5 and 6 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Organism, *Pseudokirchneriella subcapitata* (algae), QSAR, - Minimalprojectionarea: Calculates the minimal projection area, C4.5 Decision Tree
by WEKA v3.6

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Metal-based nanoparticle hazard classification on Staphylococcus
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Metal-based nanoparticle hazard classification on Staphylococcus aureus by C4.5 Decision tree

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Guangchao Chen

chen@cml.leidenuniv.nl

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Chen, G., Peijnenburg, W. J. G. M., Kovalishyn, V., & Vijver, M. G. (2016). Development of nanostructure-activity relationships assisting the nanomaterial hazard categorization for risk assessment and regulatory decision-making. RSC Advances, 6(57)

<http://doi.org/10.1039/c6ra06159a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Staphylococcus aureus (bacteria)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as MIC

3.3. Comment on endpoint:

Staphylococcus aureus (39 records) cases were selected from dataset III.

Units of the toxicity values were presented into mg/L. For classification models a threshold must to be set to label with Active or Inactive classes the different individuals into the data sets.

In that case the threshold was set at 1.0 mg/L

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

C4.5 Decision Tree

by WEKA v3.6

4.3.Descriptors in the model:

- ALogPS_logS: solubility in water; 1

4.4.Descriptor selection:

Using the 'Calculate descriptors' function implemented in OCHEM, two types of descriptors were calculated and acquired ALogPS, and Chemaxon descriptors. The selected subgroups of Chemaxon descriptors are elemental analysis, charge, geometry, partitioning, protonation and isomers that are generated at the specified pH value 7.4

Within the development of the model the best descriptors are set in the final trees.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

32/1

Descriptor: Chemical ratio :1:32

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-based NPs within the range of experimental parameters (descriptors) of the training set with the same applied organism in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

32 Metal

Metal Oxide

List: Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂

Shape: NA

Coating: NA

Size (nm): NA

Other info: Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material.

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

6.6.Pre-processing of data before modelling:

To estimate the predictive power of generated models, each dataset was randomly split into a training set (80%) and a test set (20%) before model construction. The learning process on the training set was executed in 10-fold cross validation to ensure the model stability

6.7.Statistics for goodness-of-fit:

Sensitivity = 0.833

Specificity = 0.875

Accuracy = 0.844

CCR = 0.854

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-fold cross validation applied (no data presented)

Y-randomization technique was applied, no significance models were obtained (close to 50% of accuracy, hence prediction is due by chance)

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

7 MMetal

Metal Oxide

List

Ag

Cu

Fe

CuO

Fe₃O₄

ZnO

TiO₂

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

Data obtained from 13 different publications.

For specific details see (in the publication) Table S6 from Supplementary material.

The species with more data record were chosen for the model development.

Due to the limited information characterizing the coating and functional groups of ENMs, descriptors were generated by the OCHEM to represent the core of the metal-based ENMs

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity = 0.800

Specificity = 1.000

Accuracy = 0.857

CCR = 0.900

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

It is not clear if the 10-fold cross validation was applied during the development of the model or after of it, only in order to evaluate the "predictivity" and the robustness of the model. If it was done to obtain the final model, the data filling the external set validation ("NPs used as test set" and also the "Predictivity (External validation statistics)") should be move to the test set and the external validation data could be interpreted as test test set and robustness properties.

Specific-species give a better accuracy results than the global ones.

There is a mechanistic Interpretation for the most relevant descriptors.

NPs: Nanoparticles

ENMs: Engineered nanomaterials

CCR: Correct Classification Rate

MIC: Minimum inhibitory concentration

OCHEM: Online Chemical Modelling Environment

9.2.Bibliography:

See Supplementary material Table 4, 5 and 6 in the publication

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Staphylococcus aureus (bacteria), QSAR, - ALogPS_logS: solubility in water,C4.5 Decision Tree
by WEKA v3.6

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting metal oxide Nps toxicity to E. Coli cell line by PLS and
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting metal oxide Nps toxicity to E. Coli cell line by PLS and pEC50_HaCaT toxicity as descriptor (Nano-QTTR)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tomasz Puzyn

t.puzyn@qsar.eu.org

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software

package:

Kar, S., Gajewicz, A., Roy, K., Leszczynski, J., & Puzyn, T. (2016). Extrapolating between toxicity endpoints of metal oxide nanoparticles: Predicting toxicity to Escherichia coli and human keratinocyte cell line (HaCaT) with Nano-QTTR. Ecotoxicology and

<http://doi.org/10.1016/j.ecoenv.2015.12.033>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/EC50)

3.3.Comment on endpoint:

Combining two datasets, common 16 metal oxides were used to develop nano-QTTR models. As CuO has only cytotoxicity value for E. coli, and, WO₃ and Mn₂O₃ had cytotoxicity values known for HaCaT cell line, we have used them as true external compounds for prediction of cytotoxicity value to HaCaT cell line and E. coli, respectively.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- Xox: charge of metal cation corresponding to a given oxide
- (ΔH_f)^c: Standard enthalpy of formation of metal oxide nanocluster
- pEC₅₀HaCaT : HaCaT cytotoxicity; 3

4.4.Descriptor selection:

Calculated as well as collected a set of 34 descriptors quantitatively describing variability of the nanoparticles' structure. These included: 17 quantum-mechanical descriptors, 11 image descriptors (derived from Transmission Electron Microscopy images) and 6 descriptors were taken directly from the publically available periodic table (Supplementary section of publication, Table S1).

A stepwise-MLR was performed to identify the most important descriptors. Finally, the cytotoxicity from HaCaT studies was added as a variable.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

16/3

Descriptor: Chemical ratio :3:16 ~ 1:5

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

The developed model is restricted to gram-negative bacteria only and specifically to E. coli. And, despite of is not specified in the paper, the parameters of the new NPs should belong to the range of the applied descriptors in the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

16 Metal Oxide

List:Al₂O₃Bi₂O₃

CoO

CuO

Cr₂O₃Fe₂O₃In₂O₃La₂O₃

NiO

Sb₂O₃SiO₂SnO₂TiO₂V₂O₃Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

6.6.Pre-processing of data before modelling:

As CuO, Mn₂O₃ and WO₃ are not employed in the development of any model, they are considered as true external data points in the present study.

Only NPs which fit between both sets were used in the training set. Due the size of the data, an extent internal validation was computed with several cross-validation procedures.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.91$$

$$RMSE_C = 0.19$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$$Q^2_LOO = 0.88$$

$$RMSE_CV = 0.17$$

$Q^2_{L10O} = 0.88$

$Q^2_{L20O} = 0.87$

$Q^2_{L25O} = 0.85$

$Q^2_{L50O} = 0.88$

$(c)R^2_p = 0.80$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

1 M Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

CuO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

Y₂O₃

ZnO

ZrO₂

Shape:NA

Coating:NA

Size(nm): 15-90

Other properties:

Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Mn₂O₃ and WO₃ were predicted with the obtained model and compared with other studies. Comparable results were obtained.

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The nano-QTTR is not clearly well fitted with the meaning of QSAR, but it is interesting the idea of the model, as they explain, where will be possible to use in vitro data to develop a model of in vivo toxicity, hence reducing the number of experimental work with animals.

There is a widespread Mechanistic Interpretation.

More statistical results were provided related with Roy et al., 2012, as in a few previous classified papers where we decided to avoid that data since, there is not a good parameter to compare with the rest of classified publication in the table

nano-QTTR: nano quantitative toxicity-toxicity relationship

R^2 : Correlation coefficient

RMSE_CV: Cross-validation Root-mean-square-error

Q^2_{LOO} : Leave-one-out cross-validation coefficient

Q^2_{L100} : Leave-10%-out cross-validation coefficient

Q^2_{L20}

9.2.Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ...

Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, 6(3), 175–178.

and

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR,

- Xox: charge of metal cation corresponding to a given oxide
- $(\Delta H_f)^c$: Standard enthalpy of formation of metal oxide nanocluster
- pEC₅₀HaCaT : HaCaT cytotoxicity,PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting metal oxide Nps toxicity to HaCaT cell line by PLS and
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting metal oxide Nps toxicity to HaCaT cell line by PLS and pEC50_E.Coli toxicity as descriptor (Nano-QTTR)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tomasz Puzyn

t.puzyn@qsar.eu.org

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software

package:

Kar, S., Gajewicz, A., Roy, K., Leszczynski, J., & Puzyn, T. (2016). Extrapolating between toxicity endpoints of metal oxide nanoparticles: Predicting toxicity to Escherichia coli and human keratinocyte cell line (HaCaT) with Nano-QTTR. Ecotoxicology and

<http://doi.org/10.1016/j.ecoenv.2015.12.033>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human keratinocyte cell line (HaCaT)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/EC50)

3.3.Comment on endpoint:

Combining two datasets, common 16 metal oxides were used to develop nano-QTTR models. As CuO has only cytotoxicity value for E. coli, and, WO₃ and Mn₂O₃ had cytotoxicity values known for HaCaT cell line, we have used them as true external compounds for prediction of cytotoxicity value to HaCaT cell line and E. coli, respectively.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- μ : electronic chemical potential
- $(\Delta H_f)^c$: enthalpy of formation of metal oxide nano-cluster representing fragment of surface
- pEC₅₀E.Coli : E. Coli cytotoxicity; 3

4.4.Descriptor selection:

Calculated as well as collected a set of 34 descriptors quantitatively describing variability of the nanoparticles' structure. These included: 17 quantum-mechanical descriptors, 11 image descriptors (derived from Transmission Electron Microscopy images) and 6 descriptors were taken directly from the publically available periodic table (Supplementary section of publication, Table S1).

A stepwise-MLR was performed to identify the most important descriptors. Finally, the cytotoxicity from E.Coli studies was added as a variable.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

16/3

Descriptor: Chemical ratio :3:16 ~ 1:5

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

The developed model is restricted to gram-negative bacteria only and specifically to E. coli (as a descriptor). And, despite of is not specified in the paper, the parameters of the new NPs should belong to the range of the applied descriptors in the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

16 Metal Oxide

List:Al₂O₃Bi₂O₃

CoO

Cr₂O₃Fe₂O₃In₂O₃La₂O₃Mn₂O₃

NiO

Sb₂O₃SiO₂SnO₂TiO₂V₂O₃WO₃Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

6.6.Pre-processing of data before modelling:

As CuO, Mn₂O₃ and WO₃ are not employed in the development of any model, they are considered as true external data points in the present study.

Only NPs which fit between both sets were used in the training set. Due the size of the data, an extent internal validation was computed with several cross-validation procedures.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.88$$

$$RMSE_C = 0.14$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$$Q^2_LOO = 0.80$$

RMSE_CV = 0.22

Q^2_{L100} = 0.80

Q^2_{L200} = 0.81

Q^2_{L250} = 0.82

Q^2_{L500} = 0.69

(c) R^2_p = 0.79

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

2 MMetal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y2O3

ZnO

ZrO2

Shape:NA

Coating:NA

Size(nm): 15-90

Other properties:

Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

CuO cytotoxicity was predicted (2.42) and compared with other studies.

Comparable results were obtained.

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The nano-QTTR is not clearly well fitted with the meaning of QSAR, but it is interesting the idea of the model, as they explain, where will be possible to use in vitro data to develop a model of in vivo toxicity, hence reducing the number of experimental work with animals.

There is a widespread Mechanistic Interpretation.

More statistical results were provided related with Roy et al., 2012, as in a few previous classified papers where we decided to avoid that data since, there is not a good parameter to compare with the rest of classified publication in the table

nano-QTTR: nano quantitative toxicity-toxicity relationship

R^2 : Correlation coefficient

RMSE_CV: Cross-validation Root-mean-square-error

Q^2_{LOO} : Leave-one-out cross-validation coefficient

Q^2_{L100} : Leave-10%-out cross-validation coefficient

Q^2_{L20}

9.2. Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ...

Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, 6(3), 175–178.

and

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Human keratinocyte cell line (HaCaT), QSAR,

- μ : electronic chemical potential

- $(\Delta H_f)^c$: enthalpy of formation of metal oxide nano-cluster representing fragment of surface

- pEC₅₀E.Coli : E. Coli cytotoxicity, PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting metal oxide Nps toxicity to E. Coli cell line by optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting metal oxide Nps toxicity to E. Coli cell line by optimal descriptors and CORAL software
(Model 1 - without size descriptor case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yong Pan

yongpan@njtech.edu.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Pan, Y., Li, T., Cheng, J., Telesca, D., Zink, J. I., & Jiang, J.
(2016). Nano-QSAR modelling for predicting the cytotoxicity of
metal oxide nanoparticles using novel descriptors. RSC
Advances, 6(31), 25766–25775.

<http://doi.org/10.1039/c6ra01298a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/EC50)

3.3. Comment on endpoint:

Determined the cytotoxicity of the metal oxide nanoparticles in terms of EC50 (concentration which cytotoxicity reduces bacteria viability up to 50%) based on the curve fitting least squares procedure.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

SMILES notation plus physicochemical properties encoded as following:

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).

-----Descriptor ----- Code -----

- Molecular weight	A
- Cationic charge	B
- Mass percentage of metal elements	C
- Individual size	D
- Aggregation size	E
; 0	

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

10/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-oxide NPs within the range of parameters (descriptors) of the training set with the same applied organism (E.Coli) in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

10 Metal Oxide

List:

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

6.6.Pre-processing of data before modelling:

The splitting algorithm was as follows:

- (1). 13 metal oxides for which toxicity data had been either taken from the previous paper, or they had been tested in Batch I were sorted based on decreasing toxicity.
- (2). In a next step they were split into two sets: the training set (T) and the validation set (V1) in a way ensured that the points from V1 were evenly distributed within the range of the toxicity of the training set compounds (T). We utilized the following pattern of splitting: T-T-V1-T-T-T-V1-T-T-T-V1-T-T.
- (3). Finally, three additional compounds tested in Batch II and La₂O₃ were additionally included in the validation set (those compounds are indicated with V2).

We split the data in an above discussed way because of three reasons:

- (i) to ensure that the compounds V1 are evenly distributed within the range of toxicity log (1/EC₅₀),
- (ii) to have both experimental batches represented in the validation set, whereas only compounds from the Batch I were used for training,
- (iii) to include to the validation set some additional compounds (V2) having toxicity not necessarily within the range of the training set (this would be impossible, if we have merged compounds from Batch I and II together and then labeled every third compound as a member of the validation set). Indeed, observed toxicity of CoO was higher than toxicity of the most toxic compound in the training set (ZnO).

6.7.Statistics for goodness-of-fit:

$R^2 = 0.8891$

RMSE = 0.181

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LMO} = 0.8378$

10-round Y-scrambling:

Training Average $R^2 = 0.157$

Test Average $R^2 = 0.122$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

7 MMetal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size(nm): 15-90

Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the

source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.8181$

RMSE = 0.257

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Comparison with a previous work (source publication) was done, and a widespread Mechanistic Interpretation was performed (see section 3.5 Mechanistic Interpretation). To do that it was applied a sensitivity analysis of the physicochemical descriptors: The R^2 value for the new reduced model on the training set was computed when the i th feature is excluded from the original model.

NPs: nanoparticles

SMILES: Simplified Molecular Input Line Entry Specification

R^2 : correlation coefficient

RMSE: Root-mean-square error

Q^2_{LMO} : correlation coefficient for leave-many-out cross-validation

CORAL: CORrelation And Logic

9.2.Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ...

Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, 6(3), 175–178.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, SMILES notation plus physicochemical properties encoded as following:

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).


-----Descriptor ----- Code -----

- | | |
|-------------------------------------|---|
| - Molecular weight | A |
| - Cationic charge | B |
| - Mass percentage of metal elements | C |
| - Individual size | D |
| - Aggregation size | E |

,Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting metal oxide Nps toxicity to E. Coli cell line by optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting metal oxide Nps toxicity to E. Coli cell line by optimal descriptors and CORAL software
(Model 2 - with size descriptor case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yong Pan

yongpan@njtech.edu.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Pan, Y., Li, T., Cheng, J., Telesca, D., Zink, J. I., & Jiang, J.
(2016). Nano-QSAR modelling for predicting the cytotoxicity of
metal oxide nanoparticles using novel descriptors. RSC
Advances, 6(31), 25766–25775.

<http://doi.org/10.1039/c6ra01298a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/EC50)

3.3. Comment on endpoint:

Determined the cytotoxicity of the metal oxide nanoparticles in terms of EC50 (concentration which cytotoxicity reduces bacteria viability up to 50%) based on the curve fitting least squares procedure.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

SMILES notation plus physicochemical properties encoded as following:

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).

-----Descriptor ----- Code -----

- | | |
|-------------------------------------|------|
| - Molecular weight | A |
| - Cationic charge | B |
| - Mass percentage of metal elements | C |
| - Individual size | D |
| - Aggregation size | E; 0 |

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

9/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-oxide NPs within the range of parameters (descriptors) of the training set with the same applied organism (E.Coli) in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

9 Metal Oxide

List:

ZnO

CuO

Al₂O₃Fe₂O₃SnO₂TiO₂V₂O₃Y₂O₃Bi₂O₃In₂O₃Sb₂O₃SiO₂ZrO₂

CoO

NiO

Cr₂O₃La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

6.6.Pre-processing of data before modelling:

The splitting algorithm was as follows:

- (1). 13 metal oxides for which toxicity data had been either taken from the previous paper, or they had been tested in Batch I were sorted based on decreasing toxicity.
- (2). In a next step they were split into two sets: the training set (T) and the validation set (V1) in a way ensured that the points from V1 were evenly distributed within the range of the toxicity of the training set compounds (T). We utilized the following pattern of splitting: T-T-V1-T-T-T-V1-T-T-V1-T-T.
- (3). Finally, three additional compounds tested in Batch II and La₂O₃ were additionally included in the validation set (those compounds are indicated with V2).

We split the data in an above discussed way because of three reasons:

- (i) to ensure that the compounds V1 are evenly distributed within the range of toxicity log (1/EC₅₀),
- (ii) to have both experimental batches represented in the validation set, whereas only compounds from the Batch I were used for training,
- (iii) to include to the validation set some additional compounds (V2) having toxicity not necessarily within the range of the training set (this would be impossible, if we have merged compounds from Batch I and II together and then labeled every third compound as a member of the validation set). Indeed, observed toxicity of CoO was higher than toxicity of the most toxic compound in the training set (ZnO).

Neither individual nor aggregation size data for CuO were available in the literature, hence the data was reduced in 1, removing CuO in this case.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.9824$$

$$RMSE = 0.065$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$$Q^2_{LMO} = 0.9745$$

10-round Y-scrambling:

$$\text{Average } R^2 = 0.157$$

$$\text{Test Average } R^2 = 0.170$$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

7 MMetal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape:NA

Coating:NA

Size(nm): 15-90

Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the

source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.8670$

RMSE = 0.216

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Comparison with a previous work Sizochenko et al. 2014 (already reported in this table) was done, and a widespread Mechanistic Interpretation was performed (see section 3.5 Mechanistic Interpretation). To do that it was applied a sensitivity analysis of the physicochemical descriptors: The R^2 value for the new reduced model on the training set was computed when the i th feature is excluded from the original model.

NPs: nanoparticles

SMILES: Simplified Molecular Input Line Entry Specification

R^2 : correlation coefficient

RMSE: Root-mean-square error

Q^2_{LMO} : correlation coefficient for leave-many-out cross-validation

CORAL: CORrelation And Logic

9.2.Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ... Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, 6(3), 175–178.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli), QSAR, SMILES notation plus physicochemical properties encoded as following:

From all the normalized initial 5 descriptors:


Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).

-----Descriptor ----- Code -----

- | | |
|-------------------------------------|----------------------------|
| - Molecular weight | A |
| - Cationic charge | B |
| - Mass percentage of metal elements | C |
| - Individual size | D |
| - Aggregation size | E, Linear regression model |

based on SMILES-based optimal descriptors by the software CORAL.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting metal oxide Nps toxicity to HaCaT cell line by optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting metal oxide Nps toxicity to HaCaT cell line by optimal descriptors and CORAL software
(Model 3 - without size descriptor case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yong Pan

yongpan@njtech.edu.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Pan, Y., Li, T., Cheng, J., Telesca, D., Zink, J. I., & Jiang, J.
(2016). Nano-QSAR modelling for predicting the cytotoxicity of
metal oxide nanoparticles using novel descriptors. RSC
Advances, 6(31), 25766–25775.

<http://doi.org/10.1039/c6ra01298a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human keratinocyte cell line (HaCaT)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/LC50)

3.3. Comment on endpoint:

Cell viability was measured using the CytoTox-Glo Cytotoxicity Assay from Promega (Madison, WI). LC50 values for all MeOx were extrapolated using the third order polynomial equation of the log transformed data with the least squares fit in GraphPad (GraphPad Software, Inc., La Jolla, CA)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

SMILES notation plus physicochemical properties encoded as following:

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).

-----Descriptor ----- Code -----

- Molecular weight	A
- Cationic charge	B
- Mass percentage of metal elements	C
- Individual size	D
- Aggregation size	E
; 0	

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

13/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-oxide NPs within the range of parameters (descriptors) of the training set with the same applied organism (HaCaT) in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

13 Metal Oxide

List: Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size (nm): 15-150

Other info: To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

6.6.Pre-processing of data before modelling:

According to the following principles:

- (i) the split is random keeping the highest and lowest toxic NPs in the training set;
- (ii) the test set NPs should lie within the chemical space occupied by the training set NPs and cover all types of oxides (MeO, MeO₂, Me₂O₃)

6.7.Statistics for goodness-of-fit:

$R^2 = 0.9606$

RMSE = 0.075

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LMO} = 0.9393$

10-round Y-scrambling:

Average $R^2 = 0.090$

Test Average $R^2 = 0.164$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

5 M Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size(nm): 15-150

Other properties:

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.8281$

RMSE = 0.250

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

Comparison with a previous work (source publication) was done, and a widespread Mechanistic Interpretation was performed (see section 3.5 Mechanistic Interpretation). To do that it was applied a sensitivity analysis of the physicochemical descriptors: The R^2 value for the new reduced model on the training set was computed when the ith feature is excluded from the original model.

NPs: nanoparticles

MeOx: Metal Oxide

SMILES: Simplified Molecular Input Line Entry Specification

R^2 : correlation coefficient

RMSE: Root-mean-square error

Q^2_{LMO} : correlation coefficient for leave-many-out cross-validation

CORAL: CORrelation And Lo

9.2.Bibliography:

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Human keratinocyte cell line (HaCaT), QSAR, SMILES notation plus physicochemical properties encoded as following:

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).


-----Descriptor ----- Code -----

- | | |
|-------------------------------------|---|
| - Molecular weight | A |
| - Cationic charge | B |
| - Mass percentage of metal elements | C |
| - Individual size | D |
| - Aggregation size | E |

,Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting metal oxide Nps toxicity to HaCaT cell line by optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting metal oxide Nps toxicity to HaCaT cell line by optimal descriptors and CORAL software
(Model 4 - with size descriptor case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yong Pan

yongpan@njtech.edu.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Pan, Y., Li, T., Cheng, J., Telesca, D., Zink, J. I., & Jiang, J.
(2016). Nano-QSAR modelling for predicting the cytotoxicity of
metal oxide nanoparticles using novel descriptors. RSC
Advances, 6(31), 25766–25775.

<http://doi.org/10.1039/c6ra01298a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Human keratinocyte cell line (HaCaT)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as log(1/LC50)

3.3. Comment on endpoint:

Cell viability was measured using the CytoTox-Glo Cytotoxicity Assay from Promega (Madison, WI). LC50 values for all MeOx were extrapolated using the third order polynomial equation of the log transformed data with the least squares fit in GraphPad (GraphPad Software, Inc., La Jolla, CA)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

SMILES notation plus physicochemical properties encoded as following:

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).

-----Descriptor ----- Code -----

- Molecular weight	A
- Cationic charge	B
- Mass percentage of metal elements	C
- Individual size	D
- Aggregation size	E
; 0	

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

12/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of metal-oxide NPs within the range of parameters (descriptors) of the training set with the same applied organism (HaCaT) in the study.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

12 Metal Oxide

List: Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO
ZrO₂

Shape: NA

Coating: NA

Size (nm): 15-150

Other info: To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

6.6.Pre-processing of data before modelling:

According to the following principles:

- (i) the split is random keeping the highest and lowest toxic NPs in the training set;
- (ii) the test set NPs should lie within the chemical space occupied by the training set NPs and cover all types of oxides (MeO, MeO₂, Me₂O₃)

Neither individual nor aggregation size data for Mn₂O₃ were available in the literature, hence the data was reduced in 1, removing Mn₂O₃ in this case.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.9997$

RMSE = 0.007

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LMO} = 0.9996$

10-round Y-scrambling:

Average $R^2 = 0.048$

Test Average $R^2 = 0.123$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

5 M Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

Mn₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size(nm): 15-150

Other properties:

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al.,

(2008)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.9905$

RMSE = 0.206

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Comparison with a previous work Sizochenko et al. 2014 (already reported in this table) was done, and a widespread Mechanistic Interpretation was performed (see section 3.5 Mechanistic Interpretation). To do that it was applied a sensitivity analysis of the physicochemical descriptors: The R^2 value for the new reduced model on the training set was computed when the i th feature is excluded from the original model.

NPs: nanoparticles

SMILES: Simplified Molecular Input Line Entry Specification

R^2 : correlation coefficient

RMSE: Root-mean-square error

Q^2_{LMO} : correlation coefficient for leave-many-out cross-validation

CORAL: CORrelation And Logic

9.2.Bibliography:

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Human keratinocyte cell line (HaCaT), QSAR, SMILES notation plus physicochemical properties encoded as following:

From all the normalized initial 5 descriptors:

Discrimination of physicochemical features according to scale (Figure 1 in the publication). Classified into 9 categories (from 0, Norm(X)<0.1 to Norm(X)>0.9 by increase of 0.1).


-----Descriptor ----- Code -----

- | | |
|-------------------------------------|---|
| - Molecular weight | A |
| - Cationic charge | B |
| - Mass percentage of metal elements | C |
| - Individual size | D |
| - Aggregation size | E |

,Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting metal oxide Nps toxicity to E. Coli cell line by MLR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting metal oxide Nps toxicity to E. Coli cell line by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Fengchang Wu

wufengchang@vip.skleg.cn

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Mu, Y., Wu, F., Zhao, Q., Ji, R., Qie, Y., Zhou, Y., ... Xing, B.
(2016). Predicting toxic potencies of metal oxide nanoparticles by means of nano-QSARs. Nanotoxicology. State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Ac

<http://doi.org/10.1080/17435390.2016.1202352>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

and

Human keratinocyte cell line (HaCaT)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as $\log(1/EC_{50})$

3.3.Comment on endpoint:

Determined the cytotoxicity of the metal oxide nanoparticles in terms of EC_{50} (concentration which cytotoxicity reduces bacteria viability up to 50%) based on the curve fitting least squares procedure.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

Calculations were made with the QSAR toolbox in the SYBYL X1.1 program (Tripos, Inc. Co, Princeton, NJ, USA) and SPSS statistics 17.0 (IBM, Chicago, IL, USA)

4.3.Descriptors in the model:

- Z/r : Polarization force parameter

- ΔH_{Me+} : represents the enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure.; 2

4.4.Descriptor selection:

The same procedure as in the Puzyn et al. 2011 (already reported in this table) was followed in this study, with the use of PM6 method in MOPAC 2012 software package to prepare the structures and compute the initial set of descriptors.

Pearson and pair-wise correlation, clustering and principal component were applied in order to identify the most relevant descriptors. 4 descriptors from the initial 26 were screened and the different combinations of them were used to develop different models. Finally the best performed model was selected as the final model.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

16/2

Descriptor: Chemical ratio :2:16 ~ 1/8

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure S2 in Supplementary material)

$h^* = 0.5625$

No NPs was detected as an outlier of the AD.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

16 Metal Oxide

List: ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size (nm): 15-90

Other info: Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory.

The same procedure as in the Puzyn et al. 2011 (already reported in this table) was followed in this study, with use of PM6 method in MOPAC 2012 software package.

6.6.Pre-processing of data before modelling:

From 18 NPs two of them Mn₂O₃ and Co₃O₄ were used as an external validation. Also with the two previous mentioned NPs, 51 NPs toxicity were predicted (For specific details see (in the publication) Table S4 in supplementary material).

Two cross-validation procedures were applied to ensure the robustness of data and the predictivity of the model.

A bootstrapping was also applied to check the robustness of the model.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.8793$

RMSE = 0.442

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LOO} = 0.855$

$Q^2_{LMO} = 0.846$

Bootstrapping $R^2 = 0.910$

Progressive scrambling analysis:

$Q^2 = 0.579$

SEP_{CV} = 0.366

$dq^2/dr^2_{yy} = 1.109$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable
 INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA
 NP size: Yes
 NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

2 M Metal Oxide

List

ZnO

CuO

Al₂O₃

Fe₂O₃

SnO₂

TiO₂

V₂O₃

Y₂O₃

Bi₂O₃

In₂O₃

Sb₂O₃

SiO₂

ZrO₂

CoO

NiO

Cr₂O₃

La₂O₃

Shape: NA

Coating: NA

Size(nm): 15-90

Other properties:

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material in the source publication of reference data) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory.

The same procedure as in the Puzyn et al. 2011 (already reported in this table) was followed in this study, with use of PM6 method in MOPAC 2012 software package.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

RMSEP = 0.228

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

A set of 51 NPs without bibliographic data was predicted and compare the obtained results with other papers. The final descriptors were explained in a Mechanistic interpretation of the model.

MLR: Multiple linear regression

R²: correlation coefficient

RMSE: Root-mean-square error

RMSE: Root-mean-square error of prediction

Q²_LOO: correlation coefficient for leave-one-out cross-validation

Q²_LMO: correlation coefficient for leave-many-o

9.2.Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ...

Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nature Nanotechnology, 6(3), 175–178.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Bacteria Escherichia Coli (E. Coli)

and


Human keratinocyte cell line (HaCaT), QSAR,

- Z/r : Polarization force parameter

- ΔH_{Me+} : represents the enthalpy of formation of a gaseous cation having the same oxidation state as that in the metal oxide structure., MLR: Multiple Linear Regression

Calculations were made with the QSAR toolbox in the SYBYL X1.1 program (Tripos, Inc. Co, Princeton, NJ, USA) and SPSS statistics 17.0 (IBM, Chicago, IL, USA)

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting metal oxide Nps toxicity to E. Coli and HaCaT by SMILES-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting metal oxide Nps toxicity to E. Coli and HaCaT by SMILES-based optimal descriptor and Monte Carlo technique and CORAL software

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

A.P. Toropova

alla.toropova@marionegri.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software

package:

Toropova, A. P., Toropov, A. A., Manganelli, S., Leone, C., Baderna, D., Benfenati, E., & Fanelli, R. (2016). Quasi-SMILES as a tool to utilize eclectic data for predicting the behaviour of nanomaterials. *NanoImpact*, 1, 60–64.

<http://doi.org/10.1016/j.impact.2016.04.003>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Bacteria Escherichia Coli (E. Coli)

3.2. Endpoint:

In vitro - Cytotoxicity - measured as pEC50

3.3. Comment on endpoint:

Determined the cytotoxicity of the metal oxide nanoparticles in terms of EC50 (concentration which cytotoxicity reduces bacteria viability up to 50%) based on the curve fitting least squares procedure.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

- %11: HaCaT

- %12: E.Coli

- SMILES attributes (? , Al, Bi, Co, Cr, Fe, O, In, La, Ni, V, Sb, Si, Y, Sn, Ti, [, Zn, Zr); 21

4.4.Descriptor selection:

Optimal descriptors based on SMILES and Monte-Carlo optimization by software CORAL

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/21

Descriptor: Chemical ratio :21:22 ~ 1:1

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Probabilistic criteria to define domain of applicability according to distribution of available data into the training and calibration set (Toropov and Toropova, 2015a; Manganelli et al., 2016; Toropov and Toropova, 2015b; Toropov et al., 2015; Toropova and Toropov, 2015a; <http://www.insilico.eu/coral>). But it was not showed in the published work.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Metal Oxide

List: Al₂O₃Bi₂O₃

CoO

Cr₂O₃Fe₂O₃In₂O₃La₂O₃

NiO

Sb₂O₃SiO₂SnO₂TiO₂V₂O₃Y₂O₃

ZnO

ZrO₂Shape: NACoating: NASize (nm): 15-90Other info: Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

6.6.Pre-processing of data before modelling:

The total set of available data has been split (three times) into the training (n = 22), calibration (n = 5), and validation (n = 5) sets.

These splits are built up according to principles:

- (i) these splits are random;
- (ii) the ranges of endpoints are similar for each sub-set (i.e. for the training, calibration, and validation set);
- (iii) these splits are different. It is possible to notice that there is a good balance of cytotoxicity data between the two sets of values

6.7.Statistics for goodness-of-fit:

- Training -

Split1:

$$r^2 = 0.79$$

$$RMSE = 0.23$$

Split2:

$$r^2 = 0.74$$

$$RMSE = 0.227$$

Split3:

$$r^2 = 0.85$$

$$RMSE = 0.191$$

- Calibration -

Split1:

$$r^2 = 0.84$$

RMSE = 0.248

Split2:

$r^2 = 0.90$

RMSE = 0.237

Split3:

$r^2 = 0.90$

RMSE = 0.441

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

- Training -

Split1:

$q^2 = 0.76$

Split2:

$q^2 = 0.69$

Split3:

$q^2 = 0.83$

- Calibration -

Split1:

(c) $R^2_p = 0.76$

Split2:

(c) $R^2_p = 0.76$

Split3:

(c) $R^2_p = 0.70$

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

5 M Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Cr₂O₃

Fe₂O₃

In₂O₃

La₂O₃

NiO

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

Y₂O₃

ZnO

ZrO₂

Shape: NA

Coating: NA

Size(nm): 15-90

Other properties:

Experimental details in the previous works, from where data was obtained:

From Puzyn et al., 2011 (already reported in this table)

Initial unit cell coordinates for the different NPs were taken from publically available crystallographic data (Table 2 of supplementary material) to be used on the calculations of the descriptors performed at the semi-empirical level of the theory with use of PM6 method in MOPAC 2009 software package.

and

From Gajewicz et al., 2015 (already reported in this table)

To verify morphology and size, one drop of a 100mg/mL solution was spotted on a formvar/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Philips/FEI CM200 TEM (Hillsboro, OR) at 120kV.

Sphericity and circularity analysis data from TEM images were computed based on pixel count on a gray scale images

Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ZP) in cell culture media was done using on a Malvern Instruments Zetasizer Nano-ZS instrument as described by Murdock et al., (2008)

Calculated selected electronic properties based on small, stoichiometric clusters, reflecting all characteristics of fragments of crystal structures (surface) of particular oxides. Molecular geometries were optimized at the level of semi-empirical PM6 method (Stewart, 2007) implemented in the MOPAC 2009 package (Stewart, 2009)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Split1:

$r^2 = 0.96$

RMSE = 0.242

Split2:

$r^2 = 0.88$

RMSE = 0.257

Split3:

$r^2 = 0.87$

RMSE = 0.244

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The obtained attributes from the publication's Table 2 are not consistent with the example presented in the table 4. It should be checked.

In this paper, the group also calculate a group of statistics from " Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm2

metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205 " Since those statistics can not be compared with the majority of the classified models, we have decided only to mention that it was applied. It will be interesting to be careful about if the use of this statistics increase in the future classified models.

Developing different models with a different splitting of data into training and validation tests can be considered as a robustness evaluation methodology.

SMILES: Simplified Molecular Input Line Entry Specification

r^2 : correlation coefficient

RMSE: root-mean-square error

q^2 : cross-validation correlation coefficient

CORAL: CORrelation And Logic

(c) R^2_p = Parameter computed from correlations coefficient

9.2.Bibliography:

(already reported in this table)

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., ...

Leszczynski, J. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nature Nanotechnology, 6(3), 175–178.

and

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. Nanotoxicology, 9(3), 313–325

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


Cell, Bacteria Escherichia Coli (E. Coli), QSAR, - %11: HaCaT

- %12: E.Coli

- SMILES attributes (?, Al, Bi, Co, Cr, Fe, O, In, La, Ni, V, Sb, Si, Y, Sn, Ti, [, Zn, Zr),Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Bioactivity response (active/inactive) classification of coated iron oxide
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Bioactivity response (active/inactive) classification of coated iron oxide NPs by Decision tree (J48)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

G. Melagraki

A. Afantitis

melagraki@novamechanics.com

melagraki@insilicolab.eu

afantitis@novamechanics.com

afantitis@insilicolab.eu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Melagraki, G., & Afantitis, A. (2015). A risk assessment tool for the virtual screening of metal oxide nanoparticles through enalos insiliconano platform. Current Topics in Medicinal Chemistry, 15(18), 1827–1836.

<http://doi.org/10.2174/1568026615666150506144536>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Endothelial cells (human aorta)
 Vascular smooth muscle cells (human coronary artery)
 Hepatocytes (human HepG2 cells)
 Murine RAW 264.7 leukemic monocyte/macrophage cell

3.2.Endpoint:

In vitro - Cytotoxicity - measured as biological response by H4

3.3.Comment on endpoint:

The dataset provided measurements of biological response for four cell types (Endothelial cells (human aorta), Vascular smooth muscle cells (human coronary artery), Hepatocytes (human HepG2 cells), and Murine RAW 264.7 leukemic monocyte/macrophage cell), exposed to the NPs at four concentrations (0.01, 0.03, 0.1, and 0.3 mg/mL Fe), determined based on four different assays (Apo: apoptosis, Mito: mitochondrial potential, Red: reducing equivalents, and ATP: ATP content) With a certainty above 95%, NP induced response that is above that of the control. In the present work, SNR(Signal-to-Noise ratio) > 1.645 was identified as a hit for a given NP. A 5% chance of miss-identifying a non-hit as "hit" for a given NP, would be equivalent to a miss-identification of 3.2 out of the 64 measurements in its HTS profile. Therefore, even if 5% uncertainty would be acceptable it would be more practical to set the threshold to or above the next higher integer value, i.e., N hit \geq 4.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

J48 classification tree
 by KMINE software

4.3.Descriptors in the model:

- R1: spin-lattice Relaxivity
- R2: spin-spin Relaxivity
- Coating; 3

4.4.Descriptor selection:

Based on the data included in the training set, the InfoGain variable selection with Ranker evaluator selected the most significant descriptors among the available.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

31/3

Descriptor: Chemical ratio :3:31 ~ 1:10

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Domain of applicability was calculated using Euclidean distances to assess similarity between NPs included in the training and test set. For each NP included in the test set the distance is calculated to its nearest neighbour in the training set and is then compared to a calculated threshold (0.906)

All predictions (from test set) could be considered reliable under above conditions.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

31 Metal

Metal Oxide

List: Fe₂O₃

Fe₃O₄

Shape: NA

Coating: Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488
 Cross-linked dextran :: Alexa Fluor 750
 Cross-linked dextran :: FITC, R-COOH
 Cross-linked dextran :: biotin
 Cross-linked dextran :: FITC, COOH
 Cross-linked dextran :: Cy3.5
 Cross-linked dextran :: Cy5.5, protamine
 Cross-linked dextran :: Cy5.5, tat
 Cross-linked dextran :: Cy5.5
 Cross-linked dextran :: Cy5
 Cross-linked dextran :: Cy7
 Cross-linked dextran :: FITC
 Cross-linked dextran :: FITC, Glutamic acid
 Cross-linked dextran :: glycine
 Cross-linked dextran :: rhodamine, protamine
 Cross-linked dextran :: FITC, succinimidyl iodoacetate
 Cross-linked dextran :: Tat peptide
 Cross-linked dextran :: VT680
 Cross-linked dextran :: VT680, protamine
 Dextran :: NA
 Sucrose :: NA
 PVA :: COOH
 PVA :: Ethylene diamine
 PVA :: Ethylene diamine, VT680
 PVA :: protamine, rhodamine
 PVA :: L-arg8-COOH
 PVA :: COOH
 PVA :: AminoSPARK™680
 PVA :: PEG Ethylene diamine, AminoSPARK™680
 PVA :: Ethylene diamine, AminoSPARK™680
 PVA, PEG :: AngioSPARK™680- IVM
 PVA :: 15-mer peptide
 PVA :: L-arg7-COOH
 PVA :: Ethylene diamine, VT750
 PVA, PEG :: Ethylene diamine, VT750
 PVA :: D-arg7-COOH
 PVA, PEG :: NA
 Arabino-galactan :: NA
 Carboxymethyldextran :: NA
 Amphiphilic polymer – PEG :: NH₂
 Amphiphilic polymer :: COOH
Size (nm): 20-74

Other info: Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

6.6.Pre-processing of data before modelling:

Data were divided by applying the default random seed within the Partitioning KNIME node. Random seed provides reproducible results upon re-execution of the node. Finally, among the 44 initially available NPs, 31 were included in the training set and 13 in the test set (external test)

6.7.Statistics for goodness-of-fit:

Specificity = 0.733

Sensitivity = 1.000

Precision = 0.800

Accuracy = 0.871

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Accuracy_L100 = 0.710

Y-randomization was performed, and no statistically significant models were developed. (Data not reported)

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

13 MMetal

Metal Oxide

List

Fe₂O₃

Fe₃O₄

Shape:NA

Coating:Coating :: Surface modification

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: NA

Cross-linked dextran :: NH₂

Cross-linked dextran :: Alexa Fluor 488

Cross-linked dextran :: Alexa Fluor 750

Cross-linked dextran :: FITC, R-COOH

Cross-linked dextran :: biotin

Cross-linked dextran :: FITC, COOH

Cross-linked dextran :: Cy3.5

Cross-linked dextran :: Cy5.5, protamine

Cross-linked dextran :: Cy5.5, tat

Cross-linked dextran :: Cy5.5

Cross-linked dextran :: Cy5

Cross-linked dextran :: Cy7

Cross-linked dextran :: FITC

Cross-linked dextran :: FITC, Glutamic acid

Cross-linked dextran :: glycine

Cross-linked dextran :: rhodamine, protamine

Cross-linked dextran :: FITC, succinimidyl iodoacetate

Cross-linked dextran :: Tat peptide

Cross-linked dextran :: VT680

Cross-linked dextran :: VT680, protamine

Dextran :: NA

Sucrose :: NA

PVA :: COOH

PVA :: Ethylene diamine

PVA :: Ethylene diamine, VT680

PVA :: protamine, rhodamine

PVA :: L-arg8-COOH

PVA :: COOH

PVA :: AminoSPARK™680

PVA :: PEG Ethylene diamine, AminoSPARK™680

PVA :: Ethylene diamine, AminoSPARK™680

PVA, PEG :: AngioSPARK™680- IVM

PVA :: 15-mer peptide

PVA :: L-arg7-COOH

PVA :: Ethylene diamine, VT750

PVA, PEG :: Ethylene diamine, VT750

PVA :: D-arg7-COOH

PVA, PEG :: NA

Arabino-galactan :: NA

Carboxymethyldextran :: NA

Amphiphilic polymer – PEG :: NH₂

Amphiphilic polymer :: COOH

Size(nm): 20-74

Other properties:

Nanoparticle size and zeta potential were measured by using a Zetasizer 1000 (Malvern Instruments); relaxivities were determined by using a Bruker Minispec MQ20 NMR

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Specificity = 0.800

Sensitivity = 0.878

Precision = 0.875

Accuracy = 0.846

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The proposed model was made publicly available online through Enalos InSilicoNano platform. Enalos InSilicoNano platform is a webservice that can host several validated and predictive models that can be utilized in the NPs design process

NP: Nanoparticle

J48: open source Java implementation of the C4.5 (an algorithm used to generate a decision tree) in the Weka data mining tool

L10O: Leave-10-out cross-validation

9.2.Bibliography:

(already reported in this table)

Liu, R., Rallo, R., Weissleder, R., Tassa, C., Shaw, S., & Cohen, Y. (2013). Nano-SAR development for bioactivity of nanoparticles with considerations of decision boundaries. *Small*, 9(9-10), 1842–1852.

and

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21), 7387–7392. <http://doi.org/10.1073/pnas.0802878105>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Endothelial cells (human aorta)

Vascular smooth muscle cells (human coronary artery)

Hepatocytes (human HepG2 cells)


Murine RAW 264.7 leukemic monocyte/macrophage cell, QSAR, - R1: spin-lattice Relaxivity

- R2: spin-spin Relaxivity

- Coating,J48 classification tree

by KMINE software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells by SVR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells by SVR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yoram Cohen

yoram@ucla.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Liu, R., Rallo, R., Bilal, M., & Cohen, Y. (2015). Quantitative structure-activity relationships for cellular uptake of surface-modified nanoparticles. *Combinatorial Chemistry and High Throughput Screening*, 18(4), 365–375.

<http://doi.org/10.2174/1386207318666150306105525>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3. Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVR: Support Vector Regression

4.3.Descriptors in the model:

- GCUT_PEOE_1: The 1/3-ile of the eigenvalues of the modified graph distance adjacency matrix (the diagonal takes the value of the PEOE partial charges)
- vsa_hyd: Approximation to the sum of VDW surface areas of hydrophobic atoms (\AA^2)
- SlogP_VSA1: Sum of v_i (the accessible VDW surface area (\AA^2) for atom i) such that L_i (the contribution to $\log P(o/w)$ from atom i) is in $(-0.4, -0.2]$
- b_double: Number of double bonds (aromatic bonds are not considered to be double bonds)
- GCUT_SLOGP_0: The smallest eigenvalue of the modified graph distance adjacency matrix (the diagonal takes atomic contribution to $\log P$ instead of partial charge)
- BCUT_SLOGP_0: The smallest eigenvalue of the modified adjacency matrix (the diagonal takes atomic contribution to $\log P$ (using the Wildman and Crippen SlogP method) instead of partial charge); 6

4.4.Descriptor selection:

The initial set of descriptors were calculated by MOE (Molecular Operating Environment) system based on their SMILES representations.

Descriptor selection was accomplished by sequential forward floating selection (SFFS). At each selection step, SFFS first conducts a forward selection to identify the descriptor that leads to the greatest increase in model performance, then backward elimination to evaluate whether previously selected descriptors should be removed due to the addition of the newly selected one.

Based on the on model performance, with respect to the selected descriptors, the suitable descriptor number was then determined by locating the "turning point" being defined when the addition of a new descriptor led to insignificant improvement (e.g., $\lesssim 1\%$ increase in R^2) in model performance.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

109/6

Descriptor: Chemical ratio :6:109 ~ 1:18

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

AD was verified with average kernel similarity approach and Williams plot. (For specific details see the publication's Figure 5b)

$g^* = 0.0113$

Covering all but two of 109 NPs

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

109 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione

Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid

2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

The applied splitting was in order to perform a k-fold cross-validation test (k=5), ten times.

6.7.Statistics for goodness-of-fit:

$R^2_{E632} = 0.806$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-round Y-randomization:

$R^2_{E632} = -0.221 \pm 0.146$

10-round 5-fold cross-validation:

$R^2_{5cv} = 0.759$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable
 MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA Metal Oxide

List

(Fe₂O₃)_n(Fe₃O₄)_m

Shape:NA

Coating:Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4,3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7,3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic dianhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione

4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione

Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride

3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione

Dibenz(c,e)oxepin-5,7-dione

6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone

Lauric anhydride

1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine

Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Not external validation test was applied, hence we cannot say that the obtained results will be totally reliable.

NP: Nanoparticle

SVR: Support Vector Regression

R^2_{E632} : The 0.632 estimator. Suitable for performance validation of models based on small datasets.

$R^2_{E632} = 0.368 \cdot R^2_{resub} + 0.632 \cdot R^2_{boot}$, where

R^2_{resub} is the model prediction accuracy assessed

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - GCUT_PEOE_1: The 1/3-ile of the eigenvalues of the modified graph distance adjacency matrix (the diagonal takes the value of the

PEOE partial charges)

- vsa_hyd: Approximation to the sum of VDW surface areas of hydrophobic atoms (\AA^2)
- SlogP_VSA1: Sum of v_i (the accessible VDW surface area (\AA^2) for atom i) such that L_i (the contribution to $\log P(o/w)$ from atom i) is in $(-0.4, -0.2]$
- b_double: Number of double bonds (aromatic bonds are not considered to be double bonds)
- GCUT_SLOGP_0: The smallest eigenvalue of the modified graph distance adjacency matrix (the diagonal takes atomic contribution to $\log P$ instead of partial charge)
- BCUT_SLOGP_0: The smallest eigenvalue of the modified adjacency matrix (the diagonal takes atomic contribution to $\log P$ (using the Wildman and Crippen SlogP method) instead of partial charge), SVR: Support Vector Regression

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction model of nanoparticles uptake by PaCa2 cells by MLR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction model of nanoparticles uptake by PaCa2 cells by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yoram Cohen

yoram@ucla.edu

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Liu, R., Rallo, R., Bilal, M., & Cohen, Y. (2015). Quantitative structure-activity relationships for cellular uptake of surface-modified nanoparticles. *Combinatorial Chemistry and High Throughput Screening*, 18(4), 365–375.

<http://doi.org/10.2174/1386207318666150306105525>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

Pancreatic human cancer cells (PaCa2)

3.2. Endpoint:

In vitro - Cellular uptake - measured as log(pM) /cell

3.3. Comment on endpoint:

Cellular uptake is expressed as decadic logarithm of the concentration (pM) of NP per cell

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- GCUT_SLOGP_1: The 1/3-ile of the eigenvalues of the modified graph distance adjacency matrix (the diagonal takes atomic contribution to logP instead of partial charge)
- SlogP_VSA1: Sum of v_i (the accessible VDW surface area (\AA^2) for atom i) such that L_i (the contribution to logP(o/w) from atom i) is in $(-0.4, -0.2]$
- GCUT_SLOGP_0: The smallest eigenvalue of the modified graph distance adjacency matrix (the diagonal takes atomic contribution to logP instead of partial charge)
- SMR_VSA2: Sum of v_i such that R_i (the contribution to molar refractivity for atom i) is in $(0.26, 0.35]$
- vsa_acc: Approximation to the sum of van der Waals (VDW) surface areas (\AA^2) of pure hydrogen bond acceptors (without counting acidic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH)
- PEOE_VSA_POL: Total polar VDW surface area SlogP_VSA1
- radius: The smallest one among the largest entries of each row of the distance matrix
- opr_leadlike: Whether the number of violations of Oprea's lead-like test < 2
- BCUT_PEOE_3: The largest eigenvalue of the modified adjacency matrix (the diagonal takes the value of the partial charges calculated by partial equalization of orbital electronegativities method (PEOE))
- vsa_don: Approximation to the sum of VDW surface areas of pure hydrogen bond donors (without counting basic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH) (\AA^2)
- SlogP_VSA9: Sum of v_i such that $L_i > 0.40$; 11

4.4.Descriptor selection:

The initial set of descriptors were calculated by MOE (Molecular Operating Environment) system based on their SMILES representations.

Descriptor selection was accomplished by sequential forward floating selection (SFFS). At each selection step, SFFS first conducts a forward selection to identify the descriptor that leads to the greatest increase in model performance, then backward elimination to evaluate whether previously selected descriptors should be removed due to the addition of the newly selected one.

Based on the on model performance, with respect to the selected descriptors, the suitable descriptor number was then determined by locating the "turning point" being defined when the addition of a new descriptor led to insignificant improvement (e.g., $\leq 1\%$ increase in R^2) in model performance.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

109/11

Descriptor: Chemical ratio :11:109 ~ 1:10

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

 $h^* = 0.43$

Covering all but 6 of 109 NPs

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

109 Metal Oxide

List: (Fe₂O₃)_n(Fe₃O₄)_m

Shape: NA

Coating: Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione
 3-Methylfuran-2,5-dione
 7,3,4-Dimethylfuran-2,5-dione
 Hexanoic anhydride
 3-Methyldihydrofuran-2,5-dione
 5,5'-Carbonylbis(2-benzofuran-1,3-dione)
 5-Nitro-2-benzofuran-1,3-dione
 6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione
 1,4,5, 8-Naphthalenetetracarboxylic acidanhydride
 4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione
 5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 4-Hydroxy-2-benzofuran-1,3-dione
 4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione
 6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione
 3H-2,1-benzoxathiol-3-one 1,1-dioxide
 3,4-Dichlorofuran-2,5-dione
 S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate
 5,6-Dichloro-2-benzofuran-1,3-dione
 4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diylmorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride

8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine
 Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
 NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid

NC(C(=O)O)C1CCC(Cl)CC1

2-Aminopropanoic acid

2-Amino-5-carbamimidamidopentanoic acid

2-Aminobutanedioic acid

2,5-Diamino-5-oxopentanoic acid

2-Aminopentanedioic acid

2-Amino-3-(1Himidazol-4-yl)propanoic acid

2-Amino-4-(methylsulfanyl)butanoic acid

2-Amino-3-phenylpropanoic acid

Dihydrofuran-2,5-dione

Acetic anhydride

3-Methylidenedihydrofuran-2,5-dione

1,4-Dioxane-2,6-dione

2-Benzofuran-1,3-dione

(2,5-Dioxotetrahydrofuran-3-yl)acetic acid

4,7-Difluoro-2-benzofuran-1,3-dione

{Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size (nm): 38

Other info: The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

6.6.Pre-processing of data before modelling:

The applied splitting was in order to perform a k-fold cross-validation test (k=5), ten times.

6.7.Statistics for goodness-of-fit:

$R^2_{E632} = 0.751$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

10-round Y-randomization:

$R^2_{E632} = -0.170 \pm 0.067$

10-round 5-fold cross-validation:

$R^2_{5cv} = 0.737$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal Oxide

List(Fe₂O₃)_n(Fe₃O₄)_mShape:NACoating:Trifluoroacetic anhydride

Chlorodifluoroacetic anhydride

Pentafluoropropanoic anhydride

4 3,3-Dimethyldihydrofuran-2,5-dione

Furan-2,5-dione

3-Methylfuran-2,5-dione

7 3,4-Dimethylfuran-2,5-dione

Hexanoic anhydride

3-Methyldihydrofuran-2,5-dione

5,5'-Carbonylbis(2-benzofuran-1,3-dione)

5-Nitro-2-benzofuran-1,3-dione

6-Bromo-1H,3Hbenzo[de]isochromene-1,3-dione

1,4,5, 8-Naphthalenetetracarboxylic acidanhydride

4,5,6,7-Tetrafluoro-2-benzofuran-1,3-dione

5-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione

4-Hydroxy-2-benzofuran-1,3-dione

4-Oxatricyclo[5.2.2.0^{2,6}]undec-8-ene-3,5-dione

6-Chloro-2H-3,1-benzoxazine-2,4(1H)-dione

3H-2,1-benzoxathiol-3-one 1,1-dioxide

3,4-Dichlorofuran-2,5-dione

S-(2,5-dioxotetrahydrofuran-3-yl) ethanethioate

5,6-Dichloro-2-benzofuran-1,3-dione

4,10-Dioxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Bicyclo[2.2.2]-7-octene-2,3,5,6-tetracarboxylic Dianhydride
 3a,4,7,7a-Tetrahydro-2-benzofuran-1,3-dione
 Dibenz(c,e)oxepin-5,7-dione
 6-Nitro-1H,3Hbenzo[de]isochromene-1,3-dione
 Tetrahydrofuro[3',4':3,4]cyclobuta[1,2-c]furan-1,3,4,6-tetrone
 Lauric anhydride
 1,3-Dioxo-1,3-dihydro-2-benzofuran-5-carboxylic acid
 5-Methyl-2-benzofuran-1,3-dione
 4-Nitro-2-benzofuran-1,3-dione
 1H-isochromene-1,3(4H)-dione
 Dihydro-2H-pyran-2,6(3H)-dione
 4,4'-Ethane-1,2-diyl dimorpholine-2,6-dione
 2H-3,1-benzoxazine-2,4(1H)-dione
 1-Methyl-2H-3,1-benzoxazine-2,4(1H)-dione
 4-Methyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrahydro-2-benzofuran-1,3-dione
 2,5-Dioxotetrahydrofuran-3,4-diyl diacetate
 4,5,6,7-Tetrabromo-2-benzofuran-1,3-dione
 Hexahydro-2-benzofuran-1,3-dione
 5,6-Dihydro-1Hcyclopenta[c]furan-1,3(4H)-dione
 Iodoacetic anhydride
 Chloroacetic anhydride
 1,7,8,9,10,10-Hexachloro-4-oxatricyclo[5.2.1.0^{2,6}]dec-8-ene-3,5-dione
 Palmitic anhydride
 5-amino-1H,3Hbenzo[de]isochromene-1,3-dione
 Decanoic anhydride
 8-Oxaspiro[4.5]decane-7,9-dione
 4-Oxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione
 1H,3Hbenzo[de]isochromene-1,3-dione
 3-Phenyldihydro-2Hpyran-2,6(3H)-dione
 4,5,6,7-Tetrachloro-2-benzofuran-1,3-dione
 4,7-Dichloro-2-benzofuran-1,3-dione
 3,3-Dimethyldihydro-2H-pyran-2,6(3H)-dione
 Pentan-1-amine
 4-Methylpentan-2-amine
 3-Amino-6-(hydroxymethyl)cyclohexane-1,2,4-triol
 Hexan-1-amine
 2-Methylpropan-2-amine
 2-Methylpropan-1-amine
 2,2-Dimethylpropan-1-amine
 3-Methylbutan-1-amine
 Pentan-3-amine
 2-Methylbutan-2-amine
 Ethane-1,2-diamine

Pentadecan-1-amine
 Propane-1,3-diamine
 Butane-1,4-diamine
 Hexane-1,6-diamine
 2-Ethylhexan-1-amine
 1-Hexadecylamine
 Heptan-2-amine
 Tetradecan-1-amine
 N-(2-Aminoethyl)ethane-1,2-diamine
 Tricyclo[3.3.1.1^{3,7}]decane-1-methanamine
 4-(2-Aminoethyl)benzene-1,2-diol
 4-(2-Aminoethyl)phenol
 N-(2-Aminoethyl)-N'-(3-aminopropyl)butane-1,4-diamine
 N,N'-Bis(2-aminoethyl)propane-1,3-diamine
 3,6,9,12-Tetraazatetradecane-1,14-diamine
 Tricyclo[3.3.1.0^{3,7}]nonan-3-amine
 Tricyclo[3.3.1.1^{3,7}]decan-2-amine
 Aminoacetic acid
 Methyl 2-amino-3-phenylpropanoate
 2-Amino-3-hydroxypropanoic acid
 2-Amino-3-hydroxybutanoic acid
 2-Amino-3-(1H-indol-3-yl)propanoic acid
 2-ammonio-3-(4-hydroxyphenyl)propanoate
 2-Amino-3-methylbutanoic acid
 2,6-Diaminohexanoic acid
NCCCCC(N)C(O)=O
 Amino(4-chlorophenyl)acetic acid
NC(C(O)=O)c1ccc(Cl)cc1
 2-Aminopropanoic acid
 2-Amino-5-carbamimidamidopentanoic acid
 2-Aminobutanedioic acid
 2,5-Diamino-5-oxopentanoic acid
 2-Aminopentanedioic acid
 2-Amino-3-(1Himidazol-4-yl)propanoic acid
 2-Amino-4-(methylsulfanyl)butanoic acid
 2-Amino-3-phenylpropanoic acid
 Dihydrofuran-2,5-dione
 Acetic anhydride
 3-Methylidenedihydrofuran-2,5-dione
 1,4-Dioxane-2,6-dione
 2-Benzofuran-1,3-dione
 (2,5-Dioxotetrahydrofuran-3-yl)acetic acid
 4,7-Difluoro-2-benzofuran-1,3-dione
 {Bis[2-(2,6-dioxomorpholin-4-yl)ethyl]amino}acetic acid

Size(nm): 38

Other properties:

The metal oxide NP were covered with a layer of 10 kDa dextran, that was cross-linked with epichlorohydrin and aminated by reaction with ammonia, hence the NPs were called:

Cross-Linked Iron Oxide (CLIO-NH₂)

NPs were made magnetofluorescent with the addition of FITC (fluorescein isothiocyanate)

Overall size (volume weighted) in aqueous solution.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Not external validation test was applied, hence we cannot say that the obtained results will be totally reliable.

NP: Nanoparticle

MLR: Multiple Linear Regression

R²_{E632}: The 0.632 estimator. Suitable for performance validation of models based on small datasets.

R²_{E632} = 0.368 * R²_{resub} + 0.632 * R²_{boot}, where

R²_{resub} is the model prediction accuracy assessed

9.2.Bibliography:

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, 23(11), 1418–1423. <http://doi.org/10.1038/nbt1159>

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, Pancreatic human cancer cells (PaCa2), QSAR, - GCUT_SLOGP_1: The 1/3-ile of the eigenvalues of the modified graph distance adjacency matrix (the diagonal takes atomic contribution to logP instead of partial charge)

- SlogP_VSA1: Sum of v_i (the accessible VDW surface area (\AA^2) for atom i) such that L_i (the contribution to logP(o/w) from atom i) is in $(-0.4,-0.2]$

- GCUT_SLOGP_0: The smallest eigenvalue of the modified graph distance adjacency matrix (the diagonal takes atomic contribution to logP instead of partial charge)

- SMR_VSA2: Sum of v_i such that R_i (the contribution to molar refractivity for atom i) is in $(0.26,0.35]$

- vsa_acc: Approximation to the sum of van der Waals (VDW) surface areas (\AA^2) of pure hydrogen bond acceptors (without counting acidic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH)

- PEOE_VSA_POL: Total polar VDW surface area SlogP_VSA1

- radius: The smallest one among the largest entries of each row of the distance matrix


- opr_leadlike: Whether the number of violations of Oprea's lead-like test < 2

- BCUT_PEOE_3: The largest eigenvalue of the modified adjacency matrix (the diagonal takes the value of the partial charges calculated by partial equalization of orbital electronegativities method (PEOE))

- vsa_don: Approximation to the sum of VDW surface areas of pure hydrogen bond donors (without counting basic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH) (\AA^2)

- SlogP_VSA9: Sum of v_i such that $L_i > 0.40$, MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting the solubility of C-60 in various solvents by ANN
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting the solubility of C-60 in various solvents by ANN

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

István. Z. Kiss

ikiss@delfin.klte.hu

2.6. Date of model development and/or publication:

2000

2.7. Reference(s) to main scientific papers and/or software package:

Kiss, I. Z., Mandi, G., & Beck, M. T. (2000). Artificial Neural Network approach to predict the solubility of C-60 in various solvents. Journal Of Physical Chemistry A, 104, 8081–8088.

<http://doi.org/10.1021/jp000739v>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

Solubilities of C60 in different solvents (Table 1 in the publication , 134 different solvents) were taken

from a previous reference. Solubilities are given in terms of logarithmic values of molar fractions (log S) because the log S values correspond to the free energy changes in the solvation process. For some solvents, zero solubility values were reported, hence log S is undeterminable. Then the authors used "< -8" notation in the Tables to face this issue.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

ANN: Artificial Neural Network

Five input units, five hidden units and one output unit was applied. Jetnet 3.0 software was used for the training.

Previously, a Kohonen network approach was used to ensure that the chosen descriptors are suitable Follo

4.3.Descriptors in the model:

Descriptors available experimentally:

- Molar volume (V_m , cm^3/mol): calculated from density and molecular weight
- Polarizability parameter: determined using refracting index

Descriptors calculated numerically:

- LUMO energy (ELUMO, eV): obtained from MOPAC software
- Saturated surface area (S_{sat} , \AA^2): obtained from PCMMODEL software
- Average polarizability (P_a , $\times 10^{23}$ ESU): obtained from MOPAC software; 5

4.4.Descriptor selection:

Solvents were classified in 5 groups according to C60 solubility. Geremia Neural Network Development System 2.0 was used to classify solvents into correct groups. On the basis of results of one and two-parameter combinations, the 5 most succesful were chosen. With these parameters the network classify correctly 80% of the solvents (success rate can be estimated from the correlated parameters, R_m , α , α' , P_a , V_{pol})

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

134/5

Descriptor: Chemical ratio :5:134 ~ 1:27

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility (Table 1 in the publication).

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

134 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

All data was used to generated the model. A first model with standard deviation 0.58 was created. Then 8 solvents were removed and the model was generated again.

6.7.Statistics for goodness-of-fit:

$\sigma=0.45$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The solubility decreases with increasing molar volume, increases with polarizability parameter, saturated surface area and average polarizability, and does not have a definite tendency with LUMO energy.

The relations are chemically correct: the derivatives of the solubility with respect to the different molecular parameters are approximately constant for certain groups of solvents (not for alcohols, Figure 4 and 5 in the publication).

Prediction is similar to other published studies.

Outliers are due to chemical interactions between solvents and fullerenes.

In addition, authors create a model to predict the Hildebrand parameters based on solubility determined in the study and other descriptors (to accept the values both solubility of both C₆₀ and I₂ should be similar).

ANN: artificial neuronal network

σ : standard deviation

9.2. Bibliography:

Beck, M. T.; Mandi, G. Fullerene Sci. Technol. 1997, 5

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, Descriptors available experimentally:

- Molar volume (V_m , cm³/mol): calculated from density and molecular weight
- Polarizability parameter: determined using refracting index

Descriptors calculated numerically:


- LUMO energy (ELUMO, eV): obtained from MOPAC software
- Saturated surface area (S_{sat} , Å²): obtained from PCMODEL software

- Average polarizability (Pa, $\times 10^{23}$ ESU): obtained from MOPAC software, ANN: Artificial Neural Network

Five input units, five hidden units and one output unit was applied. Jetnet 3.0 software was used for the training.

Previously, a Kohonen network approach was used to ensure that the chosen descriptors are suitable Follo

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction of C60 solubilities from solvent molecular structures
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction of C60 solubilities from solvent molecular structures
MLR case

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Peter C. Jurs

NA

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Danauskas, S. M., & Jurs, P. C. (2001). Prediction of C60 solubilities from solvent molecular structures. *Journal of Chemical Information and Computer Sciences*, 41(2), 419–424.

<http://doi.org/10.1021/ci000140s>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

The set of organic solvents used for this study included 96 compounds compiled from several sources. The solubility was expressed in units of (1×10^4) (mole fraction of C60 in the saturated solution). To limit the range of the data, the base-ten logarithm was taken, and the new range extended from -3.00 (ethanol and propanone) to 2.12 log units (1-phenylnaphthalene). The identity of each compound and its experimental value for log solubility are presented in Table 1 in the publication. The data set consisted of a mixture of 45 alkanes, 36 benzene derivatives, 7 naphthalenes, 14 oxygen-containing compounds, 10 nitrogen-containing compounds, 21 chlorine-containing compounds, and 15 bromine-containing compounds. There is no groping, as in other papers.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

Generated using the ADAPT (Automated Data and Pattern Recognition Tool) software

4.3.Descriptors in the model:

Topological descriptors:

- S6CH: χ index of chains of length six
- NCl : number of chlorines
- NBr: number of bromines
- MDE 14 : distance edge for 1° to 4° carbons
- MDE 44 : distance edge for 4° to 4° carbons
- EMIN : minimum atomic e-state value
- MPOL : molecular polarizability

Electronic descriptors:

- LUMO : lowest unoccupied molecular orbital

Geometric/ electronic hybrid descriptors:

- MOMH 7 : radius of gyration; 9

4.4.Descriptor selection:

MOPAC software program was used to generate the initial set of descriptors.

Feature selection was performed using a routine that tests for identical values (85% of descriptor values are identical) and descriptors showing a pairwise correlation greater than 0.85. These descriptors are then eliminated from the original pool. If this reduced pool is not less than the theoretical cutoff limit of 60% of the data set, a vector space descriptor analysis routine is performed to further reduce the pool to reach the theoretical cutoff.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

86/9

Descriptor: Chemical ratio :9:86 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Range of solubility: base-ten logarithm, from -3.00 (ethanol and propanone) to 2.12 log units (1-phenylnaphthalene)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

86 Carbon-based

List: Fullerene C60Shape: SphericalCoating: NASize (nm): NA

Other info: The structure of each compound is sketched using Hyperchem and is placed in a preliminary conformation using a conjugate gradient descent algorithm. The compounds are then optimized again using the semiempirical MOPAC software program. The PM3 Hamiltonian is used for each optimization. To ensure the optimizations did not begin in a local

minimum, Peter's test for optimization had to be satisfied.

6.6.Pre-processing of data before modelling:

Random split between solvents

6.7.Statistics for goodness-of-fit:

rms error for the training set = 0.417 log solubility

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

10 MCarbon-based

List

Fullerene C60

Shape:Spherical

Coating:NA

Size(nm): NA

Other properties:

The structure of each compound is sketched using Hyperchem and is placed in a preliminary conformation using a conjugate gradient descent algorithm. The compounds are then optimized again using the semiempirical MOPAC software program. The PM3 Hamiltonian is used for each optimization. To ensure the optimizations did not begin in a local minimum, Peter's test for optimization had to be satisfied.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

rms error for the predicting set = 0.50 log solubility

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Prediction rms decreased to 0.40 log if 1-phenylnaphthalene was removed from the set.

Model was developed using standard tests of linear fitness and T-values greater than 4.

CNN: computational Neural Network

MLR: Multiple linear regression

rms: root mean square error

9.2.Bibliography:

Beck, M. T.; Mandi, G.; Keki, S. Solubility and Molecular Structure State of C60 Organic Solvents, 187th Electrochemical Society Meeting, Reno, NV, May 1995; paper 956.

Ruoff, R. S.; Tse, R.; Malhoutra, R.; Lorents, D. C. Solubility of C60 in a Variety of Solvents. J. Phys. Chem. 1993, 97, 3379-3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C60 and their Utility for the Rapid Acquisition of ¹³C NMR Data for Fullerenes. J. Chem. Soc., Chem. Commun. 1993, 15, 1207-1209.

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.;

Mathews, C. K. Solubility of C₆₀ in Organic Solvents. J. Org. Chem. 1992, 57, 6077-6079.

(12)

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, Topological descriptors:

- S6CH: χ index of chains of length siz
- NCl : number of chlorines
- NBr: number of bromines
- MDE 14 : distance edge for 1^o to 4^o carbons
- MDE 44 : distance edge for 4^o to 4^o carbons
- EMIN : minimum atomic e-state value
- MPOL : molecular polarizability

Electronic descriptors:


- LUMO : lowest unoccupied molecular orbital

Geometric/ electronic hybrid descriptors:

- MOMH 7 : radius of gyration,MLR: Multiple Linear Regression

Generated using the ADAPT (Automated Data and Pattern Recognition Tool) software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction of C60 solubilities from solvent molecular structures
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction of C60 solubilities from solvent molecular structures
CNN case

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Peter C. Jurs

NA

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Danauskas, S. M., & Jurs, P. C. (2001). Prediction of C60 solubilities from solvent molecular structures. *Journal of Chemical Information and Computer Sciences*, 41(2), 419–424.

<http://doi.org/10.1021/ci000140s>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

The set of organic solvents used for this study included 96 compounds compiled from several

sources. The solubility was expressed in units of (1×10^4) (mole fraction of C60 in the saturated solution). To limit the range of the data, the base-ten logarithm was taken, and the new range extended from -3.00 (ethanol and propanone) to 2.12 log units (1-phenylnaphthalene). The identity of each compound and its experimental value for log solubility are presented in Table 1 in the publication. The data set consisted of a mixture of 45 alkanes, 36 benzene derivatives, 7 naphthalenes, 14 oxygen-containing compounds, 10 nitrogen-containing compounds, 21 chlorine-containing compounds, and 15 bromine-containing compounds. There is no grouping, as in other papers.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

CNN: Computational Neural Network

Generated using the ADAPT (Automated Data and Pattern Recognition Tool) software

4.3.Descriptors in the model:

Topological descriptors:

- S6CH : χ index of chains of length six
- NCl : number of chlorines
- NBr : number of bromines
- MDE 14: distance edge for 1° to 4° carbons
- MDE 44 : distance edge for 4° to 4° carbons
- EMIN : minimum atomic e-state value
- MPOL : molecular polarizability

Electronic descriptors:

- LUMO : lowest unoccupied molecular orbital

Geometric/ electronic hybrid descriptors:

- MOMH 7 : radius of gyration; 9

4.4.Descriptor selection:

The multiple linear regression model (previous built model in the publication) descriptors are used as the input to a three-layer, feed-forward, fully connected Computational Neural Network. The descriptor values are transformed to restrict their range to the interval [0,1] and they are then sent to a hidden layer that employs a sigmoidal activation function.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

76/9

Descriptor: Chemical ratio :9:76 ~ 1:8

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Range of solubility: base-ten logarithm, from -3.00 (ethanol and propanone) to 2.12 log units (1-phenylnaphthalene)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

76 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6. Pre-processing of data before modelling:

Random split between solvents

6.7. Statistics for goodness-of-fit:

rms error for the training set = 0.30 log solubility

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

rms error for the cross-validation set = 0.45 log solubility

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

10 MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

rms error for the predicting set = 0.52 log solubility

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Prediction rms decreased to 0.34 log units if trichloromethane was removed from the set.

Model was developed using standard tests of linear fitness and T-values greater than 4.

Descriptors tend to agree with basic solvation principles.

Errors indicate a poor predictive model in that the external prediction set rms error was much higher than the training set.

CNN: computational Neural Network

MLR: Multiple linear regression

rms: root mean square error

9.2.Bibliography:

Beck, M. T.; Mandi, G.; Keki, S. Solubility and Molecular Structure State of C60 Organic Solvents, 187th Electrochemical Society Meeting, Reno, NV, May 1995; paper 956.

Ruoff, R. S.; Tse, R.; Malhoutra, R.; Lorents, D. C. Solubility of C60 in a Variety of Solvents. J. Phys. Chem. 1993, 97, 3379-3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C60 and their Utility for the Rapid Acquisition of ¹³C NMR Data for Fullerenes. J. Chem. Soc., Chem. Commun. 1993, 15, 1207-1209.

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Solubility of C60 in Organic Solvents. J. Org. Chem. 1992, 57, 6077-6079.

(12)

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, Topological descriptors:

- S6CH : χ index of chains of length siz
- NCI : number of chlorines
- NBr : number of bromines
- MDE 14: distance edge for 1° to 4° carbons
- MDE 44 : distance edge for 4° to 4° carbons
- EMIN : minimum atomic e-state value
- MPOL : molecular polarizability

Electronic descriptors:


- LUMO : lowest unoccupied molecular orbital

Geometric/ electronic hybrid descriptors:

- MOMH 7 : radius of gyration,CNN: Computational Neural Network

Generated using the ADAPT (Aumotated Data and Pattern Recognition Tool) software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Prediction of C60 solubilities from solvent molecular structures
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Prediction of C60 solubilities from solvent molecular structures
CNN case 2

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Peter C. Jurs

NA

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Danauskas, S. M., & Jurs, P. C. (2001). Prediction of C60 solubilities from solvent molecular structures. *Journal of Chemical Information and Computer Sciences*, 41(2), 419–424.

<http://doi.org/10.1021/ci000140s>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

The set of organic solvents used for this study included 96 compounds compiled from several

sources. The solubility was expressed in units of (1×10^4) (mole fraction of C60 in the saturated solution). To limit the range of the data, the base-ten logarithm was taken, and the new range extended from -3.00 (ethanol and propanone) to 2.12 log units (1-phenylnaphthalene). The identity of each compound and its experimental value for log solubility are presented in Table 1 in the publication. The data set consisted of a mixture of 45 alkanes, 36 benzene derivatives, 7 naphthalenes, 14 oxygen-containing compounds, 10 nitrogen-containing compounds, 21 chlorine-containing compounds, and 15 bromine-containing compounds. There is no grouping, as in other papers.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

CNN: Computational Neural Network

In this case, the model was developed using a cost function of the Neural Network itself as a fitness evaluator, as opposed to the linear regression model being used as the fitness evaluator.

Generated using the ADAPT

4.3.Descriptors in the model:

Topological descriptors:

- S6CH : χ index of chains of length six
- N5C : number of fifth-order paths
- MOLC 5: path three molecular connectivity
- NN : number of nitrogens
- MDE 14: molecular polarizability

Electronic descriptors:

- QNEG : charge on the most negative atom
- HARD): hardness coefficient

Geometric/ electronic hybrid descriptors:

- PNSA 3): partial negative surface area; 9

4.4.Descriptor selection:

The simulated annealing and linear regression routine was used to create a pool of linear regression models, containing five to nine descriptors each, with T values greater than 1.0 (same descriptors as selected in the Multiple Linear Regression model previously built in the publication).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

76/9

Descriptor: Chemical ratio :9:76 ~ 1:8

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Range of solubility: base-ten logarithm, from -3.00 (ethanol and propanone) to 2.12 log units (1-phenylnaphthalene)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

76 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6. Pre-processing of data before modelling:

Random split between solvents

6.7. Statistics for goodness-of-fit:

rms error for the training set = 0.255 log solubility

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

rms error for the cross-validation set = 0.253 log solubility

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

10 MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

rms error for the predicting set = 0.346 log solubility

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

With the exception of trichloromethane, the remaining prediction set compounds fall within the range of the training and crossvalidation sets, indicating the formation of a predictive model. The Type IIA model represents a 15.3% improvement in the training set rms error, a 43.5% improvement in the cross-validation error, and a 33.7% improvement in the prediction set rms error over the Type II model (see above).

Monte Carlos study was performed by randomizing the dependent variables and retraining the models. From the results obtained, it is highly unlikely that a predictive model could be developed for the scrambled dependent variables using the methods and cutoffs employed.

Errors indicate a poor predictive model in that the external prediction set rms error was much higher than the training set.

CNN: computational Neural Network

MLR: Multiple linear regression

rms: root mean square error

9.2.Bibliography:

Beck, M. T.; Mandi, G.; Keki, S. Solubility and Molecular Structure State of C60 Organic Solvents, 187th Electrochemical Society Meeting, Reno, NV, May 1995; paper 956.

Ruoff, R. S.; Tse, R.; Malhoutra, R.; Lorents, D. C. Solubility of C60 in a Variety of Solvents. J. Phys. Chem. 1993, 97, 3379-3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C60 and their Utility for the Rapid Acquisition of ¹³C NMR Data for Fullerenes. J. Chem. Soc., Chem. Commun. 1993, 15, 1207-1209.

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Solubility of C60 in Organic Solvents. J. Org. Chem. 1992, 57, 6077-6079.

(12)

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, Topological descriptors:

- S6CH : χ index of chains of length six
- N5C : number of fifth-order paths
- MOLC 5: path three molecular connectivity
- NN : number of nitrogens
- MDE 14: molecular polarizability

Electronic descriptors:

- QNEG : charge on the most negative atom
- HARD): hardness coefficient


Geometric/ electronic hybrid descriptors:

- PNSA 3): partial negative surface area, CNN: Computational Neural Network

In this case, the model was developed using a cost function of the Neural Network itself as a fitness evaluator, as opposed to the linear regression model being used as the fitness evaluator.

Generated using the ADAPT

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (alkanes) by MRA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents (alkanes) by MRA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).
In this case subset A: alkanes

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- h_{xc} : cluster connectivity index of order $h = 3$
- Φ : polarizability parameter (determined from the refractive index, eq (1)).; 2

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

6/2

Descriptor: Chemical ratio :2:6 ~ 1:3

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

6 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subset considered is A: alkanes (7). Model is generated for this subset.

6.7.Statistics for goodness-of-fit:

$r=0.997$

$r^2=0.993$

SE=0.0576

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape:SphericalCoating:NASize(nm): NAOther properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

From the original subset, the authors remove the solvent hexane, and the model generated is for $n=6$. Polarizability parameter explained 99% of data. Small number of data.

r : square root of correlation coefficient

r^2 : correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2. Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent Advances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C₆₀ in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C₆₀ and their Utility for the Rapid Acquisition of C₁₃ NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C₆₀. Full. Sci Tech. 1997, 5(2), 291-310.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, - h_{xc} : cluster connectivity index of order $h = 3$

- Φ : polarizability parameter (determined from the refractive index, eq (1))., MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (alkyl halides) by MRA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents (alkyl halides) by MRA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).

In this case subset B: alkyl halides

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- $h\chi$: connectivity index of order $h = 1$
- $h\chi$: connectivity index of order $h = 2$
- $h\chi_v$: valence connectivity index of order $h = 1$
- $h\chi_{vc}$: valence cluster connectivity index of order $h = 3$
- $h\chi_{vpc}$: valence path-cluster connectivity index of order $h = 4$
- IP: linear combination of:
 - F: number of fluorine atoms
 - Cl: number of chlorine atoms
 - Br: number of bromine atoms
 - I: number of iodine atoms
- Φ : polarizability parameter (determined from the refractive index, eq (1)); 9

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

32/9

Descriptor: Chemical ratio :9:32 ~ 1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

32 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subset considered is b: alkyl halides (32). Model is generated for this subset.

6.7.Statistics for goodness-of-fit:

$r=0.967$

$r^2=0.935$

SE=0.220

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: SphericalCoating: NASize(nm): NAOther properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

In halogenated solvents, the solute-solvent interaction cannot be explained fully by polarizability.

r: square root of correlation coefficient

r^2 : correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2. Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent Advances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C₆₀ in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C₆₀ and their Utility for the Rapid Acquisition of C¹³ NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C₆₀. Full. Sci Tech. 1997, 5(2), 291-310.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, - $h\chi$: connectivity index of order $h = 1$

- $h\chi$: connectivity index of order $h = 2$

- $h\chi_v$: valence connectivity index of order $h = 1$

- $h\chi_{vc}$: valence cluster connectivity index of order $h = 3$

- $h\chi_{vpc}$: valence path-cluster connectivity index of order $h = 4$

- IP: linear combination of:

F: number of fluorine atoms


Cl: number of chlorine atoms

Br: number of bromine atoms

I: number of iodine atoms

- Φ : polarizability parameter (determined from the refractive index, eq (1)).,MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (alcohols) by MRA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents (alcohols) by MRA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).
In this case subset C: alcohols

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- $h\chi$: connectivity index of order $h = 3$
- Φ : polarizability parameter (determined from the refractive index, eq (1)).; 2

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

6/2

Descriptor: Chemical ratio :2:6 ~ 1:3

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

6 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subset considered is C: alcohols (6). Model is generated for this subset.

6.7.Statistics for goodness-of-fit:

$r=0.993$

$r^2=0.986$

SE=0.0997

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape:SphericalCoating:NASize(nm): NAOther properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Polarizability parameter explained 97,3% of data. Small number of data.

r: square root of correlation coefficient

r²: correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2. Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent Advances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C₆₀ in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C₆₀ and their Utility for the Rapid Acquisition of C¹³ NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C₆₀. Full. Sci Tech. 1997, 5(2), 291-310.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, - h_x: connectivity index of order h =3

- Φ: polarizability parameter (determined from the refractive index, eq (1))., MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (cycloalkanes) by MRA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents (cycloalkanes) by MRA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).

In this case subset D: cycloalkanes

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- $h\chi$: connectivity index of order $h = 5$
- Φ : polarizability parameter (determined from the refractive index, eq (1)).; 2

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

5/2

Descriptor: Chemical ratio :2:5 ~ 1:3

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

5 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subset considered is D: cycloalkanes (5). Model is generated for this subset.

6.7.Statistics for goodness-of-fit:

$r=0.999$

$r^2=0.999$

SE=0.0428

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape:SphericalCoating:NASize(nm): NAOther properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Being cyclic in nature, the structural parameter was able to explain a higher percentage of data variability. Small number of data.

r: square root of correlation coefficient

r²: correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2. Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent Advances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C₆₀ in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C₆₀ and their Utility for the Rapid Acquisition of C¹³ NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C₆₀. Full. Sci Tech. 1997, 5(2), 291-310.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, - h_x: connectivity index of order h =5

- Φ: polarizability parameter (determined from the refractive index, eq (1))., MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (alkylbenzenes) by MRA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents (alkylbenzenes) by MRA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).

In this case subset E: alkylbenzenes

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- h_{xvpc} : valence path-cluster connectivity index of order $h = 6$
- IP: lienar combination of:
 - o: indicator parameter for ortho substituents
 - m: indicator parameter for meta substituents
 - p: indicator parameter for para substituents
- ; 4

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

16/4

Descriptor: Chemical ratio :4:16 ~1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

16 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subset considered is E: alkylbenzenes (16). Model is generated for this subset.

6.7.Statistics for goodness-of-fit:

$r=0.974$

$r^2=0.949$

$SE=0.112$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: SphericalCoating: NASize(nm): NAOther properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Position of substitution played an important role in deciding the solubility. solubility of C60 depends on the structure of the solvent molecule. This observation is analogous to the situation in cycloalkanes.Small number of data.

r: square root of correlation coefficient

r²:correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2.Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent AdVances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C60 in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C60 and their Utility for the Rapid Acquisition of C13 NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C60. Full. Sci Tech. 1997, 5(2), 291-310.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, - h_xvpc: valence path-cluster connectivity index of order h = 6

- IP: lienar combination of:

o: indicator parameter for ortho substituents
m: indicator parameter for meta substituents
p: indicator parameter for para substituents
,MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (aryl halides) by MRA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents (aryl halides) by MRA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).

In this case subset F: aryl halides

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- h_{xpc} : path-cluster connectivity index of order $h = 6$
- IP: linear combination of:
 - F: number of fluorine atoms
 - Cl: number of chlorine atoms
 - Br: number of bromine atoms
 - o : indicator parameter for ortho substituents
 - m : indicator parameter for meta substituents
 - p : indicator parameter for para substituents
- Φ : polarizability parameter (determined from the refractive index, eq (1)).; 8

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

9/8

Descriptor: Chemical ratio :8:9 ~1:1

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

9 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6. Pre-processing of data before modelling:

Subset considered is F: aryl halides (9). Model is generated for this subset.

6.7. Statistics for goodness-of-fit:

$r=0.994$

$r^2=0.988$

SE=0.157

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The solubility of C60 in a cyclic solvent depends very much on the shape of the solvent molecule. Small number of data.

r: square root of correlation coefficient

r²:correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2.Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent AdVances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C60 in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C60 and their Utility for the Rapid Acquisition of C13 NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C60. Full. Sci Tech. 1997, 5(2), 291-310.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - h_{xpc}: path-cluster connectivity index of order h = 6

- IP: linear combination of:

F: number of fluorine atoms

Cl: number of chlorine atoms

Br: number of bromine atoms


o: indicator parameter for ortho substituents

m: indicator parameter for meta substituents

p: indicator parameter for para substituents

- Φ : polarizability parameter (determined from the refractive index, eq (1)).,MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (Aliphatic) by MRA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents (Aliphatic) by MRA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).

In this case subsets A+B: aliphatic (n=39)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- $h\chi$: connectivity index of order $h = 1$
- $h\chi$: cluster connectivity index of order $h = 3$
- $h\chi_v$: valence connectivity index of order $h = 0$
- $h\chi_{vc}$: valence cluster connectivity index of order $h = 3$
- $h\chi_{vpc}$: valence path-cluster connectivity index of order $h = 4$
- Φ : polarizability parameter (determined from the refractive index, eq (1)).; 6

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

39/6

Descriptor: Chemical ratio :6:39 ~ 1:6

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental parameters (descriptors) of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

39 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subsets considered are A + B: Aliphatic (39). Model is generated for this subset.

6.7.Statistics for goodness-of-fit:

$r=0.952$

$r^2=0.906$

$SE=0.241$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: SphericalCoating: NASize(nm): NAOther properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

No regression model with a standard error less than 0.2 was obtained.

r: square root of correlation coefficient

r²: correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2. Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent Advances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C₆₀ in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C₆₀ and their Utility for the Rapid Acquisition of C¹³ NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C₆₀. Full. Sci Tech. 1997, 5(2), 291-310.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, - h_x: connectivity index of order h = 1


- h_x: cluster connectivity index of order h = 3

- h_{xv}: valence connectivity index of order h = 0

- h_{xvc}: valence cluster connectivity index of order h = 3

- $h\chi_{\text{vpc}}$: valence path-cluster connectivity index of order $h = 4$
- Φ : polarizability parameter (determined from the refractive index, eq (1)).,MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (Aliphatic_2) by MRA
	Printing Date: 30/03/2017

1.QSAR identifier

1.1.QSAR identifier (title):

Predicting C60 solubility in organic solvents (Aliphatic_2) by MRA

1.2.Other related models:

NA

1.3.Software coding the model:

NA

2.General information

2.1.Date of QMRF:

30/03/2017

2.2.QMRF author(s) and contact details:

LEITAT

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6.Date of model development and/or publication:

2001

2.7.Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8.Availability of information about the model:

No information available

2.9.Availability of another QMRF for exactly the same model:

No information available

3.Defining the endpoint - OECD Principle 1

3.1.Species:

NA

NA

3.2.Endpoint:

Solubility in organic solvents

3.3.Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).

In this case subsets A+B+C: aliphatic (n=45)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- Φ : polarizability parameter (determined from the refractive index, eq (1)).

- IP: linear combination of:

F: number of fluorine atoms

Cl: number of chlorine atoms

Br: number of bromine atoms

I: number of iodine atoms

OH: number of alcohol groups; 10

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

45/10

Descriptor: Chemical ratio :10:45 ~ 1:5

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

45 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subsets considered are A + B+C: Aliphatic (45). Model is generated for this subset.

6.7.Statistics for goodness-of-fit:

$r=0.954$

$r^2=0.910$

SE=0.256

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

No regression model with a standard error less than 0.2 was obtained. Polarizability parameter (Φ) and indicator parameter (IP) were sufficient to explain 91% of the data variability.

r: square root of correlation coefficient

r²:correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2.Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent AdVances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C₆₀ in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C₆₀ and their Utility for the Rapid Acquisition of C₁₃ NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C₆₀. Full. Sci Tech. 1997, 5(2), 291-310.

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - Φ : polarizability parameter (determined fromt the refractive index, eq (1)).

- IP: linear combination of:

F: number of fluorine atoms


Cl: number of chlorine atoms

Br: number of bromine atoms

I: number of iodine atoms

OH: number of alcohol groups, MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (Aromatics) by MRA
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents (Aromatics) by MRA

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).

In this case subsets E+F: aromatics (n=25)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- h_{xv} : connectivity index of order $h = 0$
 - h_{xy} : connectivity index of order $h = 1$
 - h_{xv} : valence connectivity index of order $h = 2$
 - h_{xvpc} : valence path-cluster connectivity index of order $h = 6$
 - IP: linear combination of:
 - F: number of fluorine atoms
 - Cl: number of chlorine atoms
 - Br: number of bromine atoms
 - I: number of iodine atoms
 - o: indicator parameter for ortho substituents
 - m: indicator parameter for meta substituents
 - p: indicator parameter for para substituents
- ; 11

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

25/11

Descriptor: Chemical ratio :11:25 ~ 1:2

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

25 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subsets considered are E+F: Aromatic (25). Model is generated for this subset.

6.7.Statistics for goodness-of-fit:

$r=0.971$

$r^2=0.943$

SE=0.207

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: SphericalCoating: NASize(nm): NAOther properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

No regression model with a standard error less than 0.2 was obtained. As in the case of the individual Sets E and F, in the combined set also higher order χ terms correlated better with the solubility of C60.

r: square root of correlation coefficient

r^2 : correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2. Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent Advances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C60 in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C60 and their Utility for the Rapid Acquisition of C13 NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C60. Full. Sci Tech. 1997, 5(2), 291-310.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - $h\chi_v$: connectivity index of order $h = 0$

- $h\chi_v$: connectivity index of order $h = 1$

- $h\chi_v$: valence connectivity index of order $h = 2$

- $h\chi_{vpc}$: valence path-cluster connectivity index of order $h = 6$

- IP: linear combination of:

F: number of fluorine atoms

Cl: number of chlorine atoms

Br: number of bromine atoms

I: number of iodine atoms


o: indicator parameter for ortho substituents

m: indicator parameter for meta substituents

p: indicator parameter for para substituents

,MRA: Multiple Regression Analysis

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents (Aliphatic_3) by MRA
	Printing Date: 30/03/2017

1.QSAR identifier

1.1.QSAR identifier (title):

Predicting C60 solubility in organic solvents (Aliphatic_3) by MRA

1.2.Other related models:

NA

1.3.Software coding the model:

NA

2.General information

2.1.Date of QMRF:

30/03/2017

2.2.QMRF author(s) and contact details:

LEITAT

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

P.R. Vasudeva Rao

vasu@igcar.emet.in

2.6.Date of model development and/or publication:

2001

2.7.Reference(s) to main scientific papers and/or software package:

Sivaraman, N., Srinivasan, T. G., Vasudeva Rao, P. R., & Natarajan, R. (2001). QSPR Modeling for Solubility of Fullerene (C60) in Organic Solvents. Journal of Chemical Information and Computer Sciences, 41(4), 1067–1074.

<http://doi.org/10.1021/ci010003a>

2.8.Availability of information about the model:

No information available

2.9.Availability of another QMRF for exactly the same model:

No information available

3.Defining the endpoint - OECD Principle 1

3.1.Species:

NA

NA

3.2.Endpoint:

Solubility in organic solvents

3.3.Comment on endpoint:

A dataset containing solubility of C60 in 75 diverse solvents is taken from publication's bibliography (ref

1-3, Table 1). Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process and also in mg/ml. Solubility data set was divided into subsets based on the chemical nature of the solvents. Some solvents are not taken into account either because of detection limits (e.g. acetone) or lack of replicates (e.g. acetonitrile).

In this case subsets A+B+C: aliphatics (n=45)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MRA: Multiple Regression Analysis

4.3.Descriptors in the model:

- h_{xvpc} : valence path-cluster connectivity index of order $h = 4$
- Φ : polarizability parameter (determined from the refractive index, eq (1)).
- Br: number of bromine atoms
- I: number of iodine atoms
- OH: number of alcohol groups; 5

4.4.Descriptor selection:

Within the building model (MRA) the best descriptors were selected.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

29/5

Descriptor: Chemical ratio :5:29 ~ 1:6

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

29 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Subsets considered are A+B+C: Aliphatic (55). Split into training and validation is random.

6.7.Statistics for goodness-of-fit:

$r=0.965$

$SE=0.254$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

16 MCarbon-based

List

Fullerene C60

Shape:Spherical

Coating:NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Solubility is determined for the external data set (16) and values are given in the publication's Table 8. Only qualitative assessment (good in general terms) is provided.

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

The regression equation could be considered a three-parameter model for aliphatic solvents. A generalized equation was also made for aromatic solvent. Poor correlation was obtained, which was attributed to the limited set of data (only 25 solvents).

r: square root of correlation coefficient

r^2 : correlation coefficient

SE: Standard error of the estimate

MRA: Multiple Regression Analysis

9.2. Bibliography:

Sivaraman, N.; Dhamodaran, R.; Kaliappan, I.; Srinivasan, T. G.; Vasudeva Rao, P. R.; Mathews, C. K. Recent Advances in The Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.: U.S.A., 1994.

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. Solubility of C₆₀ in a variety of Solvents. J. Phys. Chem. 1993, 97, 3379- 3383.

Scrivens, W. A.; Tour, J. M. Potent Solvents for C₆₀ and their Utility for the Rapid Acquisition of C¹³ NMR data for Fullerenes. J. Chem. Soc. Chem. Commun. 1993, 1207-1209.

Scrivens, W. A.; Cassell, A. M.; Kinsey, K. E.; Tour, J. M. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Kadish, K. M., Ruoff, R. S., Eds.; U.S.A., 1994.

(7).

Beck, M. T.; Mandi, G.; Keki, S. In Fullerenes Vol. 2; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1995.

Beck, M. T.; Mandi, G. In Fullerenes Vol. 3; Ruoff, R. S., Kadish, K. S., Eds.; The Electrochem. Soc.: Pennington, NJ, 1996.

Beck, M. T.; Mandi, G. Solubility of C₆₀. Full. Sci Tech. 1997, 5(2), 291-310.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC


10.3. Keywords:

NA, NA, QSPR, - $h\chi_{vpc}$: valence path-cluster connectivity index of order $h = 4$

- Φ : polarizability parameter (determined from the refractive index, eq (1)).
- Br: number of bromine atoms
- I: number of iodine atoms

- OH: number of alcohol groups, MRA: Multiple Regression Analysis

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting solubility of C60 fullerene by LSER approach
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting solubility of C60 fullerene by LSER approach
(Stepwise-MLR)
- Solvents at 298K

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yizhak Marcus

ymarcus@vms.huji.ac.il

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Marcus, Y., Smith, A. L., Korobov, M. V., Mirakyan, A. L., Avramenko, N. V., & Stukalin, E. B. (2001). Solubility of C 60 Fullerene. The Journal of Physical Chemistry B, 105(13), 2499–2506.

<http://doi.org/10.1021/jp0023720>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3.Comment on endpoint:

Solubility data is used from previous references (source publications). Properties could be either measurable physical and chemical properties or ones computed by means of molecular mechanics or semiempirical quantum chemistry. 113 solubility data items (for 298 K, Table 1 in the publication) and 32 data items for 303 K (Table 2 in the publication), which could be employed with as many independent variables (solvent properties) as were deemed pertinent.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSPR

4.2.Explicit algorithm:

Stepwise-MLR (Multiple Linear Regression)
by Crunc v. 4.0 software

4.3.Descriptors in the model:

- Dimroth-Reichardt "general polarity" parameter (ET(30)1)
- Molar refraction (R1)
- Molar volume (V1)
- Electron pair donicity (β_1); 4

4.4.Descriptor selection:

Stepwise linear regression method

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

95/4

Descriptor: Chemical ratio :4:95 ~ 1:24

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Ranges of solubility for the different solvents (Tables 1 and 2).

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

95 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: For some solvents (mostly aromatic), DSC was used to identify solid solvates and to determine the enthalpy of incongruent melting or decomposition of the solvate and the incongruent melting temperature. With this data the hypothetical solubility (together with measured solubility) can be determined. This is important for some solvents for which significant decrease of C60 solubility is observed because the formation of solvates with aromatic solvents. Then solubility of the unsolvated C60 is calculated.

6.6.Pre-processing of data before modelling:

Model was generated with 95 out of the 113 items, 18 outliers with deviation $> 2\sigma$.

6.7.Statistics for goodness-of-fit:

$R^2=0.9915$

$\sigma=0.411$ in $\log x_2$

$F_{4,91}=2781$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

For some solvents (mostly aromatic), DSC was used to identify solid solvates and to determine the enthalpy of incongruent melting or decomposition of the solvate and the incongruent melting temperature. With this data the hypothetical solubility (together with measured solubility) can be determined.

This is important for some solvents for which significant decrease of C60 solubility is observed because the formation of solvates with aromatic solvents. Then solubility of the unsolvated C60 is calculated.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

None of the iodoalkanes fit the correlation, methanol and others because of the extrem low solubility. Others because possible interactions C60-solvent.

Compared to previous reports, to point out is the absence of a constant in the regression.

It is possible to predict the solubility of the C60 fullerene in solvents in which it is so low that it could not so far be determined reliably.

The present approach permits predictions of solubilities within 1 order of magnitude (2σ of the correlations) with much less effort.

For aromatic solvents where the fullerene formed crystalline solvates, the enthalpy of incongruent melting and the temperature of maximum solubility were used to determine the "hypothetical solubility" of the unsolvated C60, which was then used in the statistical procedure instead of the solubility of the solvate.

Increasing molar volume and solvent polarity (as measured by the Dimroth-Reichardt "general polarity" parameter, ET(30)) diminished the solubility of C60, whereas electron pair donation ability and polarizability enhanced solubility.

LSER: linear solvation energy relationship

R^2 : correlation coefficient

σ : estandar error in log x2

F_{4,91} : Fisher statistic value

9.2. Bibliography:

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. J. Phys. Chem. 1993, 97, 3379-3383.

Heymann, D. Carbon 1996, 34, 627-631.

Kimata, K.; Hirose, T.; Moriuchi, K.; Hosoya, K.; Araki, T.; Tanaka, N. Anal. Chem. 1995, 67, 2556.

Beck, M. T.; Mandi, G. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Los Angeles, CA, 1996; p 32.

Beck, M. T.; Mandi, G.; Keki, S. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Reno, NV, 1995; p 1510.

Scrivens, W. A.; Tour, J. M. J. Chem. Soc., Chem. Commun. 1993, 15, 1207-1209.

Zhou, X.; Liu, J.; Jin, Z.; Gu, Z.; Wu, Y.; Sun, Y. Fullerene Sci. Technol. 1997, 5, 285-290.

Ref17. Mandi, G.; Beck, M. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Montreal, Canada, 1997; p 382.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - Dimroth-Reichardt "general polarity" parameter (ET(30)1)


- Molar refraction (R1)

- Molar volume (V1)

- Electron pair donicity (β_1); Stepwise-MLR (Multiple Linear Regression)

by Crunc v. 4.0 software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting solubility of C60 fullerene by LSER approach
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting solubility of C60 fullerene by LSER approach
(Stepwise-MLR)
- Solvents at 303K

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yizhak Marcus

ymarcus@vms.huji.ac.il

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Marcus, Y., Smith, A. L., Korobov, M. V., Mirakyan, A. L., Avramenko, N. V., & Stukalin, E. B. (2001). Solubility of C 60 Fullerene. The Journal of Physical Chemistry B, 105(13), 2499–2506.

<http://doi.org/10.1021/jp0023720>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3.Comment on endpoint:

Solubility data is used from previous references (source publications). Properties could be either measurable physical and chemical properties or ones computed by means of molecular mechanics or semiempirical quantum chemistry. 113 solubility data items (for 298 K, Table 1 in the publication) and 32 data items for 303 K (Table 2 in the publication), which could be employed with as many independent variables (solvent properties) as were deemed pertinent.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSPR

4.2.Explicit algorithm:

Stepwise-MLR (Multiple Linear Regression)
by Crunc v. 4.0 software

4.3.Descriptors in the model:

- Dimroth-Reichardt "general polarity" parameter (ET(30)1)
 - Molar volume (V1)
 - Polarity/polarizability parameter (π^*)
- ; 4

4.4.Descriptor selection:

Stepwise linear regression method

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

25/4

Descriptor: Chemical ratio :4:25 ~ 1:6

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Ranges of solubility for the different solvents (Tables 1 and 2).

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

25 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: For some solvents (mostly aromatic), DSC was used to identify solid solvates and to determine the enthalpy of incongruent melting or decomposition of the solvate and the incongruent melting temperature. With this data the hypothetical solubility (together with measured solubility) can be determined. This is important for some solvents for which significant decrease of C60 solubility is observed because the formation of solvates with aromatic solvents.

6.6.Pre-processing of data before modelling:

Model was generated with 25 out of the 32 items, 7 outliers with deviation $> 2\sigma$.

6.7.Statistics for goodness-of-fit:

$R^2=0.9907$

$\sigma=0.388$ in log x2

$F_{4,91}=886$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

For some solvents (mostly aromatic), DSC was used to identify solid solvates and to determine the enthalpy of incongruent melting or decomposition of the solvate and the incongruent melting temperature. With this data the hypothetical solubility (together with measured solubility) can be determined.

This is important for some solvents for which significant decrease of C60 solubility is observed because the formation of solvates with aromatic solvents.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Better performance when polarizability R and electron pair donicity are replaced by the polarity/polarizability parameter.

Compared to previous reports, to point out is the absence of a constant in the regression.

It is possible to predict the solubility of the C60 fullerene in solvents in which it is so low that it could not so far be determined reliably.

The present approach permits predictions of solubilities within 1 order of magnitude (2σ of the correlations) with much less effort.

For aromatic solvents where the fullerene formed crystalline solvates, the enthalpy of incongruent melting and the temperature of maximum solubility were used to determine the "hypothetical solubility" of the unsolvated C60, which was then used in the statistical procedure instead of the solubility of the solvate.

Increasing molar volume and solvent polarity (as measured by the Dimroth-Reichardt "general polarity" parameter, ET(30)) diminished the solubility of C60, whereas electron pair donation ability and polarizability enhanced solubility.

LSER: linear solvation energy relationship

R^2 : correlation coefficient

σ : estandar error in log x2

$F_{4,91}$: Fisher statistic value

9.2. Bibliography:

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. J. Phys. Chem. 1993, 97, 3379-3383.

Heymann, D. Carbon 1996, 34, 627-631.

Kimata, K.; Hirose, T.; Moriuchi, K.; Hosoya, K.; Araki, T.; Tanaka, N. Anal. Chem. 1995, 67, 2556.

Beck, M. T.; Mandi, G. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Los Angeles, CA, 1996; p 32.

Beck, M. T.; Mandi, G.; Keki, S. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Reno, NV, 1995; p 1510.

Scrivens, W. A.; Tour, J. M. J. Chem. Soc., Chem. Commun. 1993, 15, 1207-1209.

Zhou, X.; Liu, J.; Jin, Z.; Gu, Z.; Wu, Y.; Sun, Y. Fullerene Sci. Technol. 1997, 5, 285-290.

Ref17. Mandi, G.; Beck, M. In Recent AdVances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Montreal, Canada, 1997; p 382.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - Dimroth-Reichardt "general polarity" parameter (ET(30)1)


- Molar volume (V1)

- Polarity/polarizability parameter (π^*)

,Stepwise-MLR (Multiple Linear Regression)

by Crunc v. 4.0 software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting solubility of C60 fullerene by LSER approach (Stepwise-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting solubility of C60 fullerene by LSER approach (Stepwise-MLR)
All solvents

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Yizhak Marcus

ymarcus@vms.huji.ac.il

2.6. Date of model development and/or publication:

2001

2.7. Reference(s) to main scientific papers and/or software package:

Marcus, Y., Smith, A. L., Korobov, M. V., Mirakyan, A. L., Avramenko, N. V., & Stukalin, E. B. (2001). Solubility of C 60 Fullerene. The Journal of Physical Chemistry B, 105(13), 2499–2506.

<http://doi.org/10.1021/jp0023720>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

Solubility data is used from previous references (source publications). Properties could be either measurable physical and chemical properties or ones computed by means of molecular mechanics or semiempirical quantum chemistry. 113 solubility data items (for 298 K, Table 1 in the publication) and 32 data items for 303 K (Table 2 in the publication), which could be employed with as many independent variables (solvent properties) as were deemed pertinent.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Stepwise-MLR (Multiple Linear Regression)

by Crunc v. 4.0 software

4.3.Descriptors in the model:

- Dimroth-Reichardt "general polarity" parameter (ET(30)1)
- Molar refraction (R1)
- Molar volume (V1)
- Electron pair donicity (β_1); 4

4.4.Descriptor selection:

Stepwise linear regression method

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

120/4

Descriptor: Chemical ratio :4:120 ~ 1:30

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Ranges of solubility for the different solvents (Tables 1 and 2).

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

120 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: For some solvents (mostly aromatic), DSC was used to identify solid solvates and to determine the enthalpy of incongruent melting or decomposition of the solvate and the incongruent melting temperature. With this data the hypothetical solubility (together with measured solubility) can be determined. This is important for some solvents for which significant decrease of C60 solubility is observed because the formation of solvates with aromatic solvents.

6.6.Pre-processing of data before modelling:

Model was generated with 120 out of the 145 items, 7 outliers with deviation $> 2\sigma$.

6.7.Statistics for goodness-of-fit:

$R^2=0.9909$

$\sigma=0.418$ in log x2

$F_{4,91}=3268$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

For some solvents (mostly aromatic), DSC was used to identify solid solvates and to determine the enthalpy of incongruent melting or decomposition of the solvate and the incongruent melting temperature. With this data the hypothetical solubility (together with measured solubility) can be determined.

This is important for some solvents for which significant decrease of C60 solubility is observed because the formation of solvates with aromatic solvents.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

NA

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

Outliers essentially the same as those defined above.

Compared to previous reports, to point out is the absence of a constant in the regression.

It is possible to predict the solubility of the C60 fullerene in solvents in which it is so low that it could not so far be determined reliably.

The present approach permits predictions of solubilities within 1 order of magnitude (2σ of the correlations) with much less effort.

For aromatic solvents where the fullerene formed crystalline solvates, the enthalpy of incongruent melting and the temperature of maximum solubility were used to determine the "hypothetical solubility" of the unsolvated C60, which was then used in the statistical procedure instead of the solubility of the solvate.

Increasing molar volume and solvent polarity (as measured by the Dimroth-Reichardt "general polarity" parameter, ET(30)) diminished the solubility of C60, whereas electron pair donation ability and polarizability enhanced solubility.

LSER: linear solvation energy relationship

R^2 : correlation coefficient

σ : standard error in $\log x_2$

$F_{4,91}$: Fisher statistic value

9.2. Bibliography:

Ruoff, R. S.; Tse, D. S.; Malhotra, R.; Lorents, D. C. J. Phys. Chem. 1993, 97, 3379-3383.

Heymann, D. Carbon 1996, 34, 627-631.

Kimata, K.; Hirose, T.; Moriuchi, K.; Hosoya, K.; Araki, T.; Tanaka, N. Anal. Chem. 1995, 67, 2556.

Beck, M. T.; Mandi, G. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Los Angeles, CA, 1996; p 32.

Beck, M. T.; Mandi, G.; Keki, S. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Reno, NV, 1995; p 1510.

Scrivens, W. A.; Tour, J. M. J. Chem. Soc., Chem. Commun. 1993, 15, 1207-1209.

Zhou, X.; Liu, J.; Jin, Z.; Gu, Z.; Wu, Y.; Sun, Y. Fullerene Sci.

Technol. 1997, 5, 285-290.

Ref17. Mandi, G.; Beck, M. In Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials; Electrochemical Society: Montreal, Canada, 1997; p 382.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - Dimroth-Reichardt "general polarity" parameter (ET(30)1)


- Molar refraction (R1)

- Molar volume (V1)

- Electron pair donicity (β_1); Stepwise-MLR (Multiple Linear Regression)

by Crunc v. 4.0 software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: QSPR model to predict the solubility of C60 various solvents
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

QSPR model to predict the solubility of C60 various solvents

Linear model

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Huanxiang Liu

xiaojunyao@yahoo.com

2.6. Date of model development and/or publication:

2005

2.7. Reference(s) to main scientific papers and/or software package:

Liu, H., Yao, X., Zhang, R., Liu, M., Hu, Z., & Fan, B. (2005).
Accurate Quantitative Structure–Property Relationship Model To
Predict the Solubility of C60 in Various Solvents Based on a Novel
Approach Using a Least-Squares Support Vector Machine. The Jo
<http://doi.org/10.1021/jp052223n>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

Solubilities of 128 different solvents are compiled from the source publication which collected the data from previous studies. Solubilities are given in terms of logarithmic values of molar fractions (log S) because the log S values correspond to the free energy changes in the solvation process

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- Randic index (order3): topological descriptor, reflects molecular size and branching
 - Relative molecular weight: constitutional descriptor, accounts both for the atomic masses (volumes) and for their distribution within the molecular space and seems to quantify effectively the bulk cohesiveness of compounds arising from the dispersion and hydrophobic interactions
 - HOMO-1 energy: quantum mechanical descriptor, energy of the second highest occupied molecular orbital
 - RNCG: quantum mechanical descriptor, relative negative charge (QMNEG/QTMINUS) quantum mechanical PC
 - ABIC1: topological descriptor, average bonding information content, reflects the connectivity of atom-atom in the molecule at the first coordination sphere
 - avg one-electron react. index for a C atom: can estimate the relative reactivity of the compounds
- ; 6

4.4.Descriptor selection:

All molecules were drawn into Hyperchem and preoptimized using an MM+ molecular mechanics force field. A more precise optimization was done with a semi-empirical AM1 method in MOPAC. The MOPAC output files were used by the CODESSA program to calculate the different descriptors.

Heuristic Method in CODESSA program was used to accomplish the preselection of descriptors.

- All descriptors are checked to ensure (a) that values of each descriptor are available for each structure and (b) that there is a variation in these values. Descriptors for which values are not available for every structure in the data in question are discarded. Descriptors having a constant value for all structures in the data set are also discarded.
- Thereafter, all possible one- parameter regression models are tested and insignificant descriptors are removed. As a next step, the program calculates the paircorrelation matrix of descriptors and further reduces the descriptor pool by eliminating highly correlated descriptors. All two-parameter regression models with remaining descriptors are subsequently developed and ranked by the regression correlation coefficient, R². A stepwise addition of further descriptor scales is performed to find the best multi-parameter regression models with the optimum values of statistical criteria (highest values of R², the cross-validated R_{cv}², and the F value).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

122/6

Descriptor: Chemical ratio :6:122 ~ 1:20

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

122 Carbon-based

List: Fullerene C60Shape: SphericalCoating: NASize (nm): NAOther info: NA**6.6.Pre-processing of data before modelling:**

6 values were removed from the first model generated

6.7. Statistics for goodness-of-fit:

$R^2=0.892$

$s^2=0.134$

RMS=0.126

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

First model is generated, but six solvents are removed because outliers. Cyclopentane, diiodomethane, 1,1,1-trichloroethane, o-cresol, nitroethane and n-butylamine can react with fullerenes.

Good predictive capability of the model.

R^2 : correlation coefficient

F: Fisher statistic value

s^2 : variance

RMS: root mean square

9.2.Bibliography:

(already reported in this table)

Kiss, I. Z., Mandi, G., & Beck, M. T. (2000). Artificial Neural Network approach to predict the solubility of C-60 in various solvents. Journal Of Physical Chemistry A, 104, 8081–8088.

Kiss et al., catch their data from a previous work:

Beck, M. T.; Mandi, G. Fullerene Sci. Technol. 1997,5

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, - Randic index (order3): topological descriptor, reflects molecular size and branching

- Relative molecular weight: constitutional descriptor, accounts both for the atomic masses (volumes) and for their distribution within the molecular space and seems to quantify effectively the bulk cohesiveness of compounds arising from the dispersion and hydrophobic interactions
- HOMO-1 energy: quantum mechanical descriptor, energy of the second highest occupied molecular orbital
- RNCG: quantum mechanical descriptor, relative negative charge (QMNEG/QTMINUS) quantum mechanical PC
- ABIC1: topological descriptor, average bonding information content, reflects the connectivity of atom-atom in the molecule at the first coordination sphere
- avg one-electron react. index for a C atom: can estimate the relative reactivity of the compounds

,MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: QSPR model to predict the solubility of C60 various solvents
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

QSPR model to predict the solubility of C60 various solvents
Nonlinear model

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Huanxiang Liu

xiaojunyao@yahoo.com

2.6. Date of model development and/or publication:

2005

2.7. Reference(s) to main scientific papers and/or software package:

Liu, H., Yao, X., Zhang, R., Liu, M., Hu, Z., & Fan, B. (2005).
Accurate Quantitative Structure–Property Relationship Model To
Predict the Solubility of C60 in Various Solvents Based on a Novel
Approach Using a Least-Squares Support Vector Machine. The Jo
<http://doi.org/10.1021/jp052223n>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

Solubilities of 128 different solvents are compiled from a paper (originally from previous papers). Solubilities are given in terms of logarithmic values of molar fractions (log S) because the log S values correspond to the free energy changes in the solvation process

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

LSSVM: Least-Squares Support Vector Machine

Using Matlab/C toolbox

4.3.Descriptors in the model:

- Randic index (order3): topological descriptor, reflects molecular size and branching
- Relative molecular weight: constitutional descriptor, accounts both for the atomic masses (volumes) and for their distribution within the molecular space and seems to quantify effectively the bulk cohesiveness of compounds arising from the dispersion and hydrophobic interactions
- HOMO-1 energy: quantum mechanical descriptor, energy of the second highest occupied molecular orbital
- RNCG: quantum mechanical descriptor, relative negative charge (QMNEG/QTMINUS) quantum mechanical PC
- ABIC1: topological descriptor, average bonding information content, reflects the connectivity of atom-atom in the molecule at the first coordination sphere
- avg one-electron react. index for a C atom: can estimate the relative reactivity of the compounds

; 6

4.4.Descriptor selection:

All molecules were drawn into Hyperchem and preoptimized using an MM+ molecular mechanics force field. A more precise optimization was done with a semi-empirical AM1 method in MOPAC. The MOPAC output files were used by the CODESSA program to calculate the different descriptors.

Heuristic Method in CODESSA program was used to accomplish the preselection of descriptors.

- All descriptors are checked to ensure (a) that values of each descriptor are available for each structure and (b) that there is a variation in these values. Descriptors for which values are not available for every structure in the data in question are discarded. Descriptors having a constant value for all structures in the data set are also discarded.
- Thereafter, all possible one- parameter regression models are tested and insignificant descriptors are removed. As a next step, the program calculates the paircorrelation matrix of descriptors and further reduces the descriptor pool by eliminating highly correlated descriptors. All two-parameter regression models with remaining descriptors are subsequently developed and ranked by the regression correlation coefficient, R². A stepwise addition of further descriptor scales is performed to find the best multi-parameter regression models with the optimum values of statistical criteria (highest values of R², the cross-validated R_{cv}², and the F value).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

92/6

Descriptor: Chemical ratio :6:92 ~ 1:15

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

92 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Random split

6.7.Statistics for goodness-of-fit:

$R^2=0.910$

RMS=0.104

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Leave-one-out cross validation of the whole training set was performed.

RMS=0.116

$R^2=0.903$

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

30 MCarbon-based

List

Fullerene C60

Shape:Spherical

Coating:NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

RMS=0.153

$R^2=0.908$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Performance of the LSSVM model is better than that of the linear model described above. Results are most satisfactory than results obtained in the source publication.

LSSVM: Least-squares support vector machine

R^2 : correlation coefficient

RMS: root mean square

9.2.Bibliography:

(already reported in this table)

Kiss, I. Z., Mandi, G., & Beck, M. T. (2000). Artificial Neural Network approach to predict the solubility of C-60 in various solvents. Journal Of Physical Chemistry A, 104, 8081–8088.

Kiss et al., catch their data from a previous work:

Beck, M. T.; Mandi, G. Fullerene Sci. Technol. 1997,5

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


NA, NA, QSPR, - Randic index (order3): topological descriptor, reflects molecular size and branching

- Relative molecular weight: constitutional descriptor, accounts both for the atomic masses (volumes) and for their distribution within the molecular space and seems to quantify effectively the bulk cohesiveness of compounds arising from the dispersion and hydrophobic interactions
- HOMO-1 energy: quantum mechanical descriptor, energy of the second highest occupied molecular orbital
- RNCG: quantum mechanical descriptor, relative negative charge (QMNEG/QTMINUS) quantum mechanical PC
- ABIC1: topological descriptor, average bonding information content, reflects the connectivity of atom-atom in the molecule at the first coordination sphere
- avg one-electron react. index for a C atom: can estimate the relative reactivity of the compounds

,LSSVM: Least-Squares Support Vector Machine

Using Matlab/C toolbox

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting Young's modulus by correlation weighting by SMILES-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting Young's modulus by correlation weighting by SMILES-based optimal descriptor and Monte Carlo technique

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Andrey A. Toropov

aatoropov@yahoo.com

2.6. Date of model development and/or publication:

2006

2.7. Reference(s) to main scientific papers and/or software

package:

Toropov, A. A., & Leszczynski, J. (2006). A new approach to the characterization of nanomaterials: Predicting Young's modulus by correlation weighting of nanomaterials codes. Chemical Physics Letters, 433(1–3), 125–129.

<http://doi.org/http://dx.doi.org/10.1016/j.cplett.2006.11.010>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Young's Modulus

3.3. Comment on endpoint:

Young's modulus is the measure of the stiffness of an elastic material and is used to characterise materials

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors.

4.3.Descriptors in the model:

Three SMILES attributes :

- atomic composition
- type of substance (bulk or not)
- temperature of synthesis

; 3

4.4.Descriptor selection:

SMILES-based optimal descriptors and Monte-Carlo optimization

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

21/3

Descriptor: Chemical ratio :3:21 ~ 1:7

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental Young's modulus..

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

21 Metal Oxide

List: AlN

Al₂O₃

TiC

ZrO₂

SiC

3Al₂O₃ 2SiO₂

MoSi₂

Shape: Film

Coating: NA

Size (nm): NA

Other info: Data on Young's modulus (from source publication) does not refer to nanomaterials specifically, but they are coatings in ceramic form or simple bulk ceramic materials.

6.6.Pre-processing of data before modelling:

Random split, following rules:

- all components of the considered species are included in the training set
- diapasons of Young's modulus values for the training and test sets are approximately the same

6.7.Statistics for goodness-of-fit:

From the three Monte Carlo runs (all similar) the best is presented:

$r^2=0.9757$

S=18.25 Gpa

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

8 Metal Oxide

List

AlN

Al₂O₃

TiC

ZrO₂

SiC

3Al₂O₃ 2SiO₂MoSi₂Shape: FilmCoating: NASize(nm): NAOther properties:

Data on Young's modulus (from source publication) does not refer to nanomaterials specifically, but they are coatings in ceramic form or simple bulk ceramic materials.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

From the three Monte Carlo runs (all similar) the best is presented:

$r^2=0.8952$

$r^2_{\text{pred}}=0.8880$

$S=34.69$ Gpa

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

Robustness is not assessed. Description of statistical characteristics is not provided. Monte Carlo method is used to calculate the values of the $CW(I_k)$ that yield correlations coefficients that are as large as possible between Young's modulus and the DWC for the training set. Three runs are done.

Apart from the training and the validation set (named test set in the paper), the authors divide randomly in four groups training and test (validation) sets and 4 other models are generated (Table 7 in the publication).

r^2 : correlation coefficient

r^2_{pred} : correlation prediction coefficient

S : root mean square error

F : Fisher statistic value

I_k : component information on the nanostructure

$CW(I_k)$: correlation weight of the component I_k

DCW: descriptor corel

9.2.Bibliography:

J.F. Shackelford, W. Alexander, The CRC Materials Science and Engineering Handbook, third ed., CRC press, Boca Raton FL 33431, 2000, p. 1980.

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


NA, NA, QSPR, Three SMILES attributes :

- atomic composition
- type of substance (bulk or not)
- temperature of synthesis

,Linear regression model

based on SMILES-based optimal descriptors.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents by SMILES-based optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents by SMILES-based optimal descriptor and Monte Carlo technique

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Andrey A. Toropov

aatoropov@yahoo.com

2.6. Date of model development and/or publication:

2007

2.7. Reference(s) to main scientific papers and/or software

package:

Toropov, A. A., Leszczynska, D., & Leszczynski, J. (2007). QSPR study on solubility of fullerene C60 in organic solvents using optimal descriptors calculated with SMILES. Chemical Physics Letters, 441(1-3), 119–122.

<http://doi.org/10.1016/j.cplett.2007.04.094>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A series of benzene derivatives have been selected. Modeling of solubility (10^{-4} molar fraction of C60 at T=298K) has been taken. These experimental values are taken from other literature sources.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors.

4.3.Descriptors in the model:

17 SMILES attributes:

- #
- (
- /
- 1, 2
- =
- C, Br, Cl, F, I, N, O, S
- c
- [N+], [O-]; 17

4.4.Descriptor selection:

SMILES-based optimal descriptors and Monte-Carlo optimization

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

25/17

Descriptor: Chemical ratio :17:25

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

25 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Data have been randomly divided into training and test sets, but interval of solubility values should be similar for the training and test sets

6.7.Statistics for goodness-of-fit:

From the three Monte Carlo runs (all similar) the best is presented:

$R^2=0.8161$

$S=3.60$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

11 MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

From the three Monte Carlo runs (all similar) the best is presented:

$R^2=0.7903$

$r^2_{pred}=0.7235$

$S=4.65$

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Not all descriptors are in both the training and test (validation) sets.

Monte Carlo method is used to calculate the values of the CW(I_k) that yield correlations coefficients that are as large as possible between solubility and the DWC for the training set. Three runs are performed.

Basic split into training and test (validation) sets is done by means of exchange some solvents from training into test set and vice-versa.

Model generation by Monte Carlo optimization is again performed for each one of these 4 splits. Statistical characteristics are reproduced well.

R²: correlation coefficient

r²_{pred}: correlation prediction coefficient. Details of the calculations are given in Table 1.

S: root mean square error

F: Fisher statistic value

I_k: component information on the nanostructure

CW(I_k): correlation w

9.2.Bibliography:

M.V. Korobov, A.L. Smith, Solubility of the fullerenes, in: K.M. Kadish, R.S. Ruoff (Eds.), Fullerenes: Chemistry, Physics, and Technology, Wiley Inter Science, 2000, p. 55 (Chapter 2).

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, 17 SMILES attributes:

- #
- (
- /
- 1, 2
- =


- C, Br, Cl, F, I, N, O, S

- c

-[N+], [O-], Linear regression model

based on SMILES-based optimal descriptors.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents by SMILES-based optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents by SMILES-based optimal descriptor and Monte Carlo technique

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Andrey A. Toropov

aatoropov@yahoo.com

2.6. Date of model development and/or publication:

2007

2.7. Reference(s) to main scientific papers and/or software package:

Toropov, A. A., Rasulev, B. F., Leszczynska, D., & Leszczynski, J. (2007). Additive SMILES based optimal descriptors: QSPR modeling of fullerene C60 solubility in organic solvents. Chemical Physics Letters, 444(1-3), 209–214.

<http://doi.org/10.1016/j.cplett.2007.07.024>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

Experimental values of fullerene solubilities (log S) are taken from a previous reference (n=122)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors.

4.3.Descriptors in the model:

21 SMILES attributes:

- #, =
- (
- /
- 1,2,3
- C, Br, Cl, F, I, N, O, S
- \
- c,n,s
- [N+], [O-]; 21

4.4.Descriptor selection:

SMILES-based optimal descriptors and Monte-Carlo optimization

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

92/21

Descriptor: Chemical ratio :21:92 ~ 1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

92 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Random split like in Liu, H., Yao, X., Zhang, R., Liu, M., Hu, Z., & Fan, B. (2005). Accurate Quantitative Structure–Property Relationship Model To Predict the Solubility of C60 in Various Solvents Based on a Novel Approach Using a Least-Squares Support Vector Machine. The Journal of Physical Chemistry B, 109(43), 20565–20571.

6.7.Statistics for goodness-of-fit:

From the three Monte Carlo runs (all similar) the best is presented:

$R^2=0.8612$

$s=0.401$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

$Q^2 = 0.8537$

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

30 MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

From the three Monte Carlo runs (all similar) the best is presented:

$R^2 = 0.8908$

$R^2_{\text{pred}} = 0.8748$

$s = 0.435$

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

This paper describes exactly the same type of model than Toropov, A. A., Leszczynska, D., & Leszczynski, J. (2007). QSPR study on solubility of fullerene C60 in organic solvents using optimal descriptors calculated with SMILES. Chemical Physics Letters, 441(1-3), 119–122. but with using another data set.

This is reliable model but with less accuracy than the nonlinear model developed in the source publication. (also the data source, which is based on quantum chemical descriptors.

Larger numbers of SFk in the training set are related to higher statistical significance of the SFk for a given model. SFk is a fragment of SMILES.

R²: correlation coefficient

Q²: cross validation coefficient. Details of the calculations are given in Table 1.

R²_{pred}: correlation prediction coefficient. Details of the calculations are given in Table 1.

s: root mean square error

F: Fisher sta

9.2. Bibliography:

(already reported in this table)

H. Liu, X. Yao, R. Zhang, M. Liu, Z. Hu, B. Fan, J. Phys. Chem. B 109 (2005) 20565.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC


10.3. Keywords:

NA, NA, QSPR, 21 SMILES attributes:

- #, =

- (
- /
- 1,2,3
- C, Br, Cl, F, I, N, O, S
- \
- c,n,s
- [N+], [O-], Linear regression model
based on SMILES-based optimal descriptors.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Model for predicting water solubility for CNT based on the chiral
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Model for predicting water solubility for CNT based on the chiral vector by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Andrey A. Toropov

aatoropov@yahoo.com

2.6. Date of model development and/or publication:

2007

2.7. Reference(s) to main scientific papers and/or software package:

Toropov, A. a, Leszczynska, D., & Leszczynski, J. (2007).
Predicting water solubility and octanol water partition coefficient
for carbon nanotubes based on the chiral vector. Computational
Biology and Chemistry, 31(2), 127–128.

<http://doi.org/10.1016/j.compbiolchem.2007.02.002>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in water

3.3. Comment on endpoint:

Data is taken for 16 CNT (with different components, i.e, n and m, of chiral vector). It is well known

that toxicity depends on chemical solubility

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- Chiral vector (n) which contains information about rolling up graphite layer in formation of CNT.
Taken from Torrens, 2005).

- Chiral vector (m) which contains information about rolling up graphite layer in formation of CNT.
Taken from Torrens, 2005); 2

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

8/2

Descriptor: Chemical ratio :2:8 ~ 1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No
 Chemical Name: not applicable
 SMILES: not applicable
 Formula: not applicable
 INChI: not applicable
 MOL file: not applicable
 Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes
 NP size: Yes
 NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

8 Carbon-based
List: CNT: Carbon nanotubes
Shape: Fiber
Coating: NA
Size (nm): NA
Other info: NA

6.6.Pre-processing of data before modelling:

Data have been randomly divided into training and test sets

6.7.Statistics for goodness-of-fit:

$r^2=0.99998$

$s=0.0534$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

8 MCarbon-based

List

CNT: Carbon nanotubes

Shape:Fiber

Coating:NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2=0.99990$

$s=0.0933$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Small number of data values. Robustness is not assessed.

r^2 : correlation coefficient

s: root mean square error

F: Fisher F-ratio

MRA: Multiple Regression Analysis

9.2.Bibliography:

Torrens, F., 2005. Partition of solvents and co-solvents of nanotubes: proteins and cyclopyranoses. In: Caldwell, G.W., Atta-ur-Rahman, B.A. (Eds.), *Frontiers in Drug Design and Discovery I*. Springer, Bentham, Hilversum (Holland), pp. 231–268.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - Chiral vector (n) which contains information about rolling up graphite layer in formation of CNT. Taken from Torrens, 2005).

- Chiral vector (m) which contains information about rolling up graphite layer in formation of CNT. Taken from Torrens, 2005),MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Model for predicting octanol water partition coefficient for CNT based
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Model for predicting octanol water partition coefficient for CNT based on the chiral vector by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Andrey A. Toropov

aatoropov@yahoo.com

2.6. Date of model development and/or publication:

2007

2.7. Reference(s) to main scientific papers and/or software package:

Toropov, A. a, Leszczynska, D., & Leszczynski, J. (2007).
Predicting water solubility and octanol water partition coefficient
for carbon nanotubes based on the chiral vector. Computational
Biology and Chemistry, 31(2), 127–128.

<http://doi.org/10.1016/j.compbiolchem.2007.02.002>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Octanol water partition coefficient

3.3. Comment on endpoint:

Data is taken for 16 CNT (with different components, i.e, n and m, of chiral vector). It is well known

that toxicity depends on chemical solubility

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- Chiral vector (n) which contains information about rolling up graphite layer in formation of CNT.
Taken from Torrens, 2005).

- Chiral vector (m) which contains information about rolling up graphite layer in formation of CNT.
Taken from Torrens, 2005); 2

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

8/2

Descriptor: Chemical ratio :2:8 ~ 1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No
 Chemical Name: not applicable
 SMILES: not applicable
 Formula: not applicable
 INChI: not applicable
 MOL file: not applicable
 Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes
 NP size: Yes
 NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

8 Carbon-based
List: CNT: Carbon nanotubes
Shape: Fiber
Coating: NA
Size (nm): NA
Other info: NA

6.6.Pre-processing of data before modelling:

Data have been randomly divided into training and test sets

6.7.Statistics for goodness-of-fit:

$r^2=0.99910$

$s=0.364$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

8 MCarbon-based

List

CNT: Carbon nanotubes

Shape:Fiber

Coating:NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$r^2=0.99960$

$s=0.287$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Small number of data values. Robustness is not assessed.

r^2 : correlation coefficient

s: root mean square error

F: Fisher F-ratio

MRA: Multiple Regression Analysis

9.2.Bibliography:

Torrens, F., 2005. Partition of solvents and co-solvents of nanotubes: proteins and cyclopyranoses. In: Caldwell, G.W., Atta-ur-Rahman, B.A. (Eds.), *Frontiers in Drug Design and Discovery I*. Springer, Bentham, Hilversum (Holland), pp. 231–268.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - Chiral vector (n) which contains information about rolling up graphite layer in formation of CNT. Taken from Torrens, 2005).

- Chiral vector (m) which contains information about rolling up graphite layer in formation of CNT. Taken from Torrens, 2005), MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents by SMILES-based optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents by SMILES-based optimal descriptor and Monte Carlo technique

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Andrey A. Toropov

aatoropov@yahoo.com

2.6. Date of model development and/or publication:

2008

2.7. Reference(s) to main scientific papers and/or software

package:

Toropov, A. A., Rasulev, B. F., Leszczynska, D., & Leszczynski, J. (2008). Multiplicative SMILES-based optimal descriptors: QSPR modeling of fullerene C60 solubility in organic solvents. Chemical Physics Letters, 457(4-6), 332–336.

<http://doi.org/10.1016/j.cplett.2008.04.013>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

Experimental values of the C60 solubility as log S, where S is expressed in molar fraction.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors.

4.3.Descriptors in the model:

25 SMILES attributes.

Nb, Number of braquets (branching);

- (000, (001, (002, (003

Ndb, Number of double bonds (=)

- =000, =001, =002, =003

SSk represents two SMILES consequent elements in the SMILES strings

- /, \, [, #, C, Br, Cl, F, I, N, O, S, c,n,s, [N+], [O-]; 25

4.4.Descriptor selection:

SMILES-based optimal descriptors and Monte-Carlo optimization

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/25

Descriptor: Chemical ratio :25: 92 ~ 1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Three splits into the training and test sets are examined in the present study. These splits obey the following principles:

(i) they are random;

(ii) the ranges of solubility for the training and test sets are similar.

For every split they perform 3 Monte Carlos runs.

6.7.Statistics for goodness-of-fit:

From the three Monte Carlo runs (all similar) the best is presented:

Split1:

$R^2 = 0.9381$

$s = 0.278$

Split2:

$R^2 = 0.9393$

$s = 0.257$

Split3:

$R^2 = 0.9349$

$s = 0.281$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Corresponding to previous column.

Split1:

$Q^2 = 0.9348$

Split2:

$q^2 = 0.9362$

Split3:

$q^2 = 0.9316$

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

Split1 : 30

Split2 : 30

Split3 : 30 MCarbon-based

List

Fullerene C60

Shape:Spherical

Coating:NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

From the three Monte Carlo runs (all similar) the best is presented:

Split1:

$R^2 = 0.9157$

$R^2_{pred} = 0.9031$

$s = 0.333$

Split2:

$R^2 = 0.9008$

$R^2_{pred} = 0.8823$

$s = 0.401$

Split3:

$R^2 = 0.9257$

$R^2_{pred} = 0.9148$

$s = 0.301$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Two SMILES attributes (Sak) are absent in the training set (C1 and Br2). Therefore correlation weight values have been fixed as 1.

Two outliers are detected and removed from the analysis.

Larger numbers of SAK in the training set are related to higher statistical significance of the SAK for a given model.

Despite Q^2 is not defined as cross-validation correlation coefficient in the paper, the meaning has been extracted from other papers

published by the same authors.

Optimal descriptors calculated with two components SAK together with the global SMILES attributes (Nb and Ndb) are based on more information related to the molecular structure than the previous version of the descriptors on previous works. However, further increase of the detailing may leads to the overtraining (i.e. a situation when an excellent model for the training set is accompanied by a poor model for the test).

Thus the proposed method has clear interpretations (each

SMILES attribute is promoter of increase or decrease of the fullerene C60 solubility and this is defined by the correlation weight

R^2 : correlation coefficient

s: root mean square error

F: Fisher statistic value

Q^2 : cross-validation correlation coefficient .

Definition is given in publication's Table 2

R^2_{pred} : predictive correlation coefficient. Defined in Table 2

9.2. Bibliography:

(already reported in this table)

Yu.Yu Prylutsky, et al. , Mater. Sci. Eng. Sect. C 23 (2003) 109.

(already reported in this table)

A.A. Toropov, B.F. Rasulev, D. Leszczynska, J. Leszczynski, Chem. Phys. Lett. 444 (2007) 209.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, 25 SMILES attributes.

Nb, Number of brackets (branching);

- (000, (001, (002, (003


Ndb, Number of double bonds (=)

- =000, =001, =002, =003

SSk represents two SMILES consequent elements in the SMILES strings

- /, \, [, #, C, Br, Cl, F, I, N, O, S, c,n,s, [N+], [O-], Linear regression model based on SMILES-based optimal descriptors.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents by means of a molecular-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents by means of a molecular-based model
GA-MLR model

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Farhard Gharagheizi

fghara@ut.ac.ir

2.6. Date of model development and/or publication:

2008

2.7. Reference(s) to main scientific papers and/or software package:

Gharagheizi, F., & Alamdari, R. F. (2008). A molecular-based model for prediction of solubility of C60 fullerene in various solvents. Fullerenes Nanotubes and Carbon Nanostructures, 16(1), 40–57.

<http://doi.org/10.1080/15363830701779315>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3.Comment on endpoint:

A dataset containing solubility of C60 in 128 diverse solvents is taken from bibliography (ref 14-17). The same dataset has been used by other two authors earlier (ref 13 and ref11). This paper compares two new models with those generated in these references. Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSPR

4.2.Explicit algorithm:

GA-MLR: Genetic Algorithm-based Multivariate Linear Regression

Using MATLAB software

4.3.Descriptors in the model:

Molecular descriptors are defined for solvents according to chemical structure using the Dragon Software.

- piPC03: Molecular multiple path count of order 03 (walk and path counts)
- ATS1m 2D: Broto-Mreanu autocorrelation of a topological structure-lag 1/weighted by atomic masses (2D autocorrelations)
- Seigp: Eigenvalue sum from polarizability weighted distance matrix (Eigenvalue 0 based indices)
- More23e: 3D-MORSE-signal 23/weighted by atomic sanderson electronegativities (More23e 3D-MORSE descriptors)
- H1m: H autocorrelation of lag 1/weighted by atomic masses (GETAWAY descriptors); 5

4.4.Descriptor selection:

The chemical structures of these 128 solvents were drawn in Hyperchem software and optimized using the MM+ molecular mechanics force field. Then, Dragon software was used to generate the initial set of descriptors.

After calculating molecular descriptors, the pool of molecular descriptors was reduced by removing descriptors that could not be calculated for every structure in the dataset, and those descriptors with an essentially constant value for all the structures.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

102/5

Descriptor: Chemical ratio :5:102 ~ 1:20

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

102 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Random split

6.7.Statistics for goodness-of-fit:

$R^2 = 0.9076$

MSE= 0.344

RMS= 0.118

SDE= 0.235

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

26 MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

$Q^2_{\text{ext}} = 0.8967$

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

From the first model generated using GA-MLR the authors remove 4 solvents of the dataset because of probable reaction between solvent-nanomaterial. The new model increased the accuracy and is the one included in goodness-of-fit.

FFNN results are better in comparison with MLR equation. MLR equation has one variable less than the best previous reported linear models and is also more accurate.

GA-MLR: Genetic algorithm-based multivariate linear regression

R^2 : squared correlation coefficient

MSE: mean square error

RMS: root mean square error

SDE: standard deviation error

F: Fisher factor

Q^2_{ext} : external explained variance

9.2. Bibliography:

Sivaraman, N., Dhamodaran, R., Kaliappan, I., Srinivasan, T.G., Rao, P.R.V., and Mathews, C.K. (1992) Solubility of C60 in organic solvents. *J. Org. Chem.*, 57: 6077–6079.

Ruoff, R.S., Tse, D.S., Malhotra, R., and Lorents, D.C. (1993) Solubility of fullerene (C60) in a variety of solvents. *J. Phys. Chem.*, 97: 3379–3383.

Scrivens, W.A. and Tour, J.M. (1993) Potent solvents for C60 and their utility for the rapid acquisition of ^{13}C NMR data for fullerenes. *J. Chem. Soc. Chem. Commun.*, 15: 1207–1209.

Beck, M.T. and Mandi, G. (1997) Solubility of C60. *Fuller. Nanotub. Car. N.*, 5: 291–310.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, Molecular descriptors are defined for solvents according to chemical structure using the Dragon Software.

- piPC03: Molecular multiple path count of order 03 (walk and path counts)

- ATS1m 2D: Broto-Mreanu autocorrelation of a topological structure-lag 1/weighted by atomic masses (2D autocorrelations)


- Seigp: Eigenvalue sum from polarizability weighted distance matrix (Eigenvalue 0 based indices)

- More23e: 3D-MORSE-signal 23/weighted by atomic sanderson electronegativities (More23e 3D-MORSE descriptors)

- H1m: H autocorrelation of lag 1/weighted by atomic masses (GETAWAY descriptors),GA-MLR: Genetic Algorithm-based Multivariate Linear Regression

Using MATLAB software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents by means of a molecular-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents by means of a molecular-based model
FFNN model

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Farhard Gharagheizi

fghara@ut.ac.ir

2.6. Date of model development and/or publication:

2008

2.7. Reference(s) to main scientific papers and/or software package:

Gharagheizi, F., & Alamdari, R. F. (2008). A molecular-based model for prediction of solubility of C60 fullerene in various solvents. Fullerenes Nanotubes and Carbon Nanostructures, 16(1), 40–57.

<http://doi.org/10.1080/15363830701779315>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

A dataset containing solubility of C60 in 128 diverse solvents is taken from bibliography (ref 14-17). The same dataset has been used by other two authors earlier (ref 13 and ref11). This paper compares two new models with those generated in these references. Solubility is given in terms of logarithmic values for molar fractions $\log(S)$ because the $\log(S)$ values correspond to the Gibbs free energy changes in the solvation process.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

FFNN: Three-layer Feed Forward Neural Network

Training was done by Levenberg-Marquart algorithm.

4.3.Descriptors in the model:

Molecular descriptors are defined for solvents according to chemical structure using the Dragon Software.

- piPC03: Molecular multiple path count of order 03 (walk and path counts)
- ATS1m 2D: Broto-Mreanu autocorrelation of a topological structure-lag 1/weighted by atomic masses (2D autocorrelations)
- Seigp: Eigenvalue sum from polarizability weighted distance matrix (Eigenvalue 0 based indices)
- More23e: 3D-MORSE-signal 23/weighted by atomic sanderson electronegativities (More23e 3D-MORSE descriptors)
- H1m: H autocorrelation of lag 1/weighted by atomic masses (GETAWAY descriptors); 5

4.4.Descriptor selection:

The chemical structures of these 128 solvents were drawn in Hyperchem software and optimized using the MM+ molecular mechanics force field. Then, Dragon software was used to generate the initial set of descriptors.

After calculating molecular descriptors, the pool of molecular descriptors was reduced by removing descriptors that could not be calculated for every structure in the dataset, and those descriptors with an essentially constant value for all the structures.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

102/5

Descriptor: Chemical ratio :5:102 ~ 1:20

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

102 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Random split

6.7.Statistics for goodness-of-fit:

$R^2 = 0.9427$

MSE= 0.267

RMS= 0.071

SDE= 0,265

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

26 MCarbon-based

List

Fullerene C60

Shape: SphericalCoating: NASize(nm): NAOther properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation: $Q^2_{\text{ext}} = 0.9413$ **7.8. Predictivity - Assessment of the external validation set:**

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

From the first model generated using GA-MLR the authors remove 4 solvents of the dataset because of probable reaction between solvent-nanomaterial. The new model increased the accuracy and is the one included in goodness-of-fit.

FFNN results are better in comparison with MLR equation. MLR equation has one variable less than the best previous reported linear models and is also more accurate.

FFNN: Three-layer feed forward Neural Network

R²: squared correlation coefficient

MSE: mean square error

RMS: root mean square error

SDE: standard deviation error

F: Fisher F-ratio

Q²_{ext}: external explained variance

9.2. Bibliography:

Sivaraman, N., Dhamodaran, R., Kaliappan, I., Srinivasan, T.G., Rao, P.R.V., and Mathews, C.K. (1992) Solubility of C60 in organic solvents. *J. Org. Chem.*, 57: 6077–6079.

Ruoff, R.S., Tse, D.S., Malhotra, R., and Lorents, D.C. (1993) Solubility of fullerene (C60) in a variety of solvents. *J. Phys. Chem.*, 97: 3379–3383.

Scrivens, W.A. and Tour, J.M. (1993) Potent solvents for C60 and their utility for the rapid acquisition of ¹³C NMR data for fullerenes. *J. Chem. Soc. Chem. Commun.*, 15: 1207–1209.

Beck, M.T. and Mandi, G. (1997) Solubility of C60. *Fuller. Nanotub. Car. N.*, 5: 291–310.

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, Molecular descriptors are defined for solvents according to chemical structure using the Dragon Software.

- piPC03: Molecular multiple path count of order 03 (walk and path counts)
- ATS1m 2D: Broto-Mreanu autocorrelation of a topological structure-lag 1/weighted by atomic masses (2D autocorrelations)
- Seigp: Eigenvalue sum from polarizability weighted distance matrix (Eigenvalue 0 based indices)
- More23e: 3D-MORSE-signal 23/weighted by atomic sanderson electronegativities (More23e 3D-MORSE descriptors)
- H1m: H autocorrelation of lag 1/weighted by atomic masses (GETAWAY descriptors),FFNN: Three-layer Feed Forward Neural Network

Training was done by Levenberg-Marquart algorithm.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Binding interactions between fullerene inhibitors and HIV-1 PR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Binding interactions between fullerene inhibitors and HIV-1 PR residues using molecular docking and molecular dynamics simulations, by 3D QSAR applying PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Serdar Durdagi

durdagis@eie.gr

2.6. Date of model development and/or publication:

2008

2.7. Reference(s) to main scientific papers and/or software

package:

Durdagi, S., Mavromoustakos, T., Chronakis, N., & Papadopoulos, M. G. (2008). Computational design of novel fullerene analogues as potential HIV-1 PR inhibitors: Analysis of the binding interactions between fullerene inhibitors and HIV-1 PR residues using

<http://doi.org/10.1016/j.bmc.2008.10.039>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Binding interaction between fullerenes and HIV-1 PR inhibitor (Human immunodeficiency virus type I

aspartic protease)

3.3.Comment on endpoint:

Experimentally reported (Table 1 in publication) and computationally designed fullerene analogues (Table 2 in publication) and their activities (measured and calculated binding affinities) have been used. Since the experimental binding activities of most of the derivatives, used in this study are only reported as median effective concentration (EC50), these values are assumed to be equal with K_i in the calculations of the free binding energies. The logarithmic values of $1/EC_{50}$ (pEC_{50}) were used in the 3D QSAR correlations, as they are related to changes in the free energy of binding.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

PLS: Partial Least Squares regression

4.3.Descriptors in the model:

In CoMSIA analysis, descriptors were treated as independent variables:

- Steric: STR
- Electrostatic: ES
- H-bond acceptor: ACC
- H-bond donor: DON
- Hydrophobic: HYD

; 4

4.4.Descriptor selection:

A series of fullerene derivatives have been designed and their binding energies with HIV-1 PR have been computed with molecular docking analysis. Therefore the structure of the compounds together with their activities (determined) are available for QSAR studies. Among the synthesized analogues, compound 23 was selected as a template, because it has the highest binding affinity at the HIV-1 PR in the training set. Several variations in the alignment schemes are considered by superimposing the similar pharmacophoric features. Highlighted carbon atoms (32 central carbon atoms of fullerene) for the template ligand 23 are selected for the structural superimposition processes. The alignment of the molecules was based on atom-by-atom superimposition of selected atoms, which are common in all compounds. The criteria applied for the selection were: (i) the overlap of the putative biologically relevant pharmacophore groups (with minimum RMSD); (ii) the use of the conformations of subgroups of fullerene derivatives obtained from the docking simulations and (iii) the statistical significance of the 3D-QSAR/CoMSIA models. Then different combinations of stereoelectronic fields of 3D QSAR/CoMSIA models are obtained.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

43/4

Descriptor: Chemical ratio :5:43 ~ 1:9

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of functionalized fullerenes in range of binding energies for the whole data set (Tables 1 and 2 in the publication)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

43 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6. Pre-processing of data before modelling:

External (test) set includes compounds representing all categories of activity of the training set

(inactive and active compounds)

6.7.Statistics for goodness-of-fit:

Best model (CoMSIA4):

$r^2=0.933$

Std error= 0.130

Relative contribution Steric=0.426

Relative contribution Electrostatic =0.127

Relative contribution H-bond donor =0.167

Relative contribution H-bond acceptor =0.280

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$r^2_{cv}=0.739$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

6 MCarbon-based

List

Fullerene C60

Shape:Spherical

Coating:NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Only qualitative assessment of 6 compounds used as external (test in the paper) validation. Range of difference in predicted pEC50: from -0.94 to 0.51.

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Novel fullerene derivatives with high inhibition effect can be designed by using this model. In last step Leapfrog de novo program and the 3D QSAR / CoMSIA contour maps have been used in order to generate a series of potent fullerenes.

Since experimental and computed binding energies showed good correlation, both experimental binding affinities of structures (Table 1 in the publication) and estimated binding affinities (Table 2 in the publication) have been used to form the 3D QSAR models.

Same model is also applied in

Durdagi, S., Mavromoustakos, T., & Papadopoulos, M. G. (2008). 3D QSAR CoMFA/CoMSIA, molecular docking and molecular dynamics studies of fullerene-based HIV-1 PR inhibitors. *Bioorganic and Medicinal Chemistry Letters*, 18(23), 6283–6289.

where their state it was first reported.

r^2 : correlation coefficient

F: Fisher factor

Std error: standard error of prediction

r^2_{cv} : cross-validation correlation coefficient

CoMSIA: Comparative molecular similarity indices analysis

9.2.Bibliography:

Friedman, H. S.; Ganapathi, P. S.; Rubin, P. S.; Kenyon, G. L. J. Med. Chem. 1998, 41, 2424.

Durdagi, S., Mavromoustakos, T., & Papadopoulos, M. G. (2008). 3D QSAR CoMFA/CoMSIA, molecular docking and molecular dynamics studies of fullerene-based HIV-1 PR inhibitors. Bioorganic and Medicinal Chemistry Letters, 18(23), 6283–6289.

Bingel, C. Chem. Ber. 1993, 126, 1957.

Ganapathi, P. S.; Friedman, S. H.; Kenyon, G. L.; Rubin, Y. J. Org. Chem. 1995, 60, 2954.

Ref 12: Schuster, D. I.; Wilson, S. R.; Schinazi, R. F. Bioorg. Med. Chem. Lett. 1996, 6, 1253.

Available from: <http://www.chemdb.niaid.nih.gov>.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, In CoMSIA analysis, descriptors were treated as independent variables:

- Steric: STR
- Electrostatic: ES
- H-bond acceptor: ACC
- H-bond donor: DON
- Hydrophobic: HYD

,PLS: Partial Least Squares regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents by InChI-based optimal
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents by InChI-based optimal descriptors and Monte-Carlo optimization

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Andrey A. Toropov

aatoropov@yahoo.com

2.6. Date of model development and/or publication:

2009

2.7. Reference(s) to main scientific papers and/or software

package:

Toropov, A. A., Toropova, A. P., Benfenati, E., Leszczynska, D., & Leszczynski, J. (2009). Additive InChI-based optimal descriptors: QSPR modeling of fullerene C60 solubility in organic solvents. Journal of Mathematical Chemistry, 46(4), 1232–1251.

<http://doi.org/10.1007/s10910-008-9514-0>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

The endpoint is not described in the paper. It can be deduced from the text that solubility data is taken from a paper published by the same group A.A Toropov, B.F. Rasulev, D. Leszczynska, J. Leszczynski, Chem. Phys. Lett. 444, 209–214 (2007).. They just compare predicting ability of both models.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Linear regression model based on additive InChI-based optimal descriptors

4.3.Descriptors in the model:

116 InChI-based descriptors / attributes (lk).

Formula layer

- Br2, Br3, Br (bromine atoms); C10; C11; C12; C14; C16; C2; C3; C4; C5; C6;C7; C8; C9; C; (carbon atom); Cl2; Cl3; Cl4; Cl (chlorine atom); F (fluorine atom); I (iodine atom); N (nitrogen atom); O2; O3; O (oxygen atom); S (sulphur atom);

Connectivity layer

(10; (11; (14; (2; (3; (4; (5; (6; (7; (8; (9; (; ,10; ,11; ,1; ,2; ,3; ,4; ,5; ,6; ,7; ,8; ,9; -10; -11; -12;-13;-14;-15;- 1;-2;-3;-4;-5;-6;-7;-8;- 9; 0; 1; 2; 3; 4; 5; 6; 7; 8; 9; c11; c1; c2; c3; c4; c6; c7; c8; c9;

Hydrogen atoms:

h1; h2; h3; h4; h5; h6; h7; h8; h9; H10; H11; H12; H14; H16; H18; H22; H26; H30; H2; H3; H4; H5; H6; H7; H8; H9; H;

Electronic charge and double bonds

+;-;b2;

Symbol \ is also included; 116

4.4.Descriptor selection:

InChI-based optimal descriptors and Monte-Carlo optimization

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/116

Descriptor: Chemical ratio :116:92

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected applicability domain of nanomaterials within the range of experimental solubility (Table 6 in the publication). Probabilistic analysis that is represented in publication's Table 3 can be used as a tool for definition of the applicability domain for this model: substances for which should be done prediction must have InChI without of rare (in the training set) attributes.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Not specified in the paper. But from other papers of the same authors:

Three splits into the training and test sets are examined in the present study. These splits obey the following principles:

- (i) they are random;
 - (ii) the ranges of solubility for the training and test sets are similar.
- For every split they perform 3 Monte Carlos runs.

6.7.Statistics for goodness-of-fit:

From the three Monte Carlo runs (all similar) the average is presented
(publication's Table 2):

Split1:

$$R^2 = 0.9463$$

$$s = 0.249$$

Split2:

$$R^2 = 0.9489$$

$$s = 0.251$$

Split3:

$$R^2 = 0.9501$$

$$s = 0.239$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

Corresponding to Split 1, first Monte Carlo Run (the only specified in the paper).

Split1, first run:

$$Q^2 = 9418$$

$$SDEP=0.258$$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

Split1 : 30

Split2 : 30

Split3 : 30 MCarbon-based

List

Fullerene C60

Shape:Spherical

Coating:NA

Size(nm): NA

Other properties:

NA

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

From the three Monte Carlo runs (all similar) the average is presented (publication's Table 2). For R^2_{pred} and SDEP the values refer to the first Monte Carlo run.

Split1:

$R^2 = 0.9405$

$s = 0.348$

$R^2_{pred} = 0.9305$

Split2:

$R^2 = 0.8853$

$s = 0.334$

Split3:

R^2

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

International Chemical Identifier can be used as elucidation of the molecular structure for the QSPR analysis of the solubility of fullerene C60 in organic solvents.

The statistical quality of the InChI-based model for fullerene C60 solubility is better than the quality of SMILES-based model (source publication) of this parameter.

Majority of the correlation weights ($W(I_k)$) have similar values for the 3 splits (either promote or decrease solubility). Some others are unstable (e.g. h_5 , h_7), probably because of a weak influence of relevant molecular phenomenon. The conclusion is that some optimal descriptors can therefore hint that some mechanistic interpretations for the C60 solubility exists (tendencies can be modified).

It should be also pointed out that for the first split of the three splits some descriptors are present (e.g. $once$), while for the other two not. This leads to significant differences in between the three values. In the paper is stated that the model should describe substances must have InChI without rare (in the training set) attributes.

R^2 : correlation coefficient

s : standard error of estimation

F : Fisher statistic value

Q^2 : cross-validated coefficient

R^2_{pred} : predictive correlation coefficient

SDEP: standard deviation of error prediction

InChI: International Chemical Identifier

9.2. Bibliography:

A.A Toropov, B.F. Rasulev, D. Leszczynska, J. Leszczynski, Chem. Phys. Lett. 444, 209–214 (2007).

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, 116 InChI-based descriptors / attributes (I_k).

Formula layer

- Br2, Br3, Br (bromine atoms); C10; C11; C12; C14; C16; C2; C3; C4; C5; C6; C7; C8; C9; C; (carbon

atom); Cl2; Cl3; Cl4; Cl (chlorine atom); F (fluorine atom); I (iodine atom); N (nitrogen atom); O2; O3;
O (oxygen atom); S (sulphur atom);

Connectivity layer

(10; (11; (14; (2; (3; (4; (5; (6; (7; (8; (9; (; ,10; ,11; ,1; ,2; ,3; ,4; ,5; ,6; ,7; ,8; ,9; -10; -11;
-12;-13;-14;-15;- 1;-2;-3;-4;-5;-6;-7;-8;- 9; 0; 1; 2; 3; 4; 5; 6; 7; 8; 9; c11; c1; c2; c3; c4; c6; c7;
c8; c9;

Hydrogen atoms:


h1; h2; h3; h4; h5; h6; h7; h8; h9; H10; H11; H12; H14; H16; H18; H22; H26; H30; H2; H3; H4; H5;
H6; H7; H8; H9; H;

Electronic charge and double bonds

+;-;b2;

Symbol \ is also included,Linear regression model based on additive InChI-based optimal descriptors

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 - C70 solubility in chlorobenzene by SMILES-based
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 - C70 solubility in chlorobenzene by SMILES-based optimal descriptor and Monte Carlo technique

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Alla A. Toropova

aatoropov@yahoo.com

2.6. Date of model development and/or publication:

2011

2.7. Reference(s) to main scientific papers and/or software package:

Toropova, A. P., Toropov, A. A., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2011). CORAL: QSPR models for solubility of [C60] and [C70] fullerene derivatives. *Molecular Diversity*, 15(1), 249–256.

<http://doi.org/10.1007/s11030-010-9245-6>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in chlorobenzene

3.3. Comment on endpoint:

NA

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

4.3.Descriptors in the model:

SMILE attributes:

- %10, %11, %12, %13, %14, %15, %16, %17, %18, %19, %20, %21, %22, %23, %24, %25, %26, %27, %28, %29, %30, %31, %32, %33, %34, %35, %36, %37, %38

- (, O, -, [, c, C, o, s

; 47

4.4.Descriptor selection:

SMILES-based optimal descriptors and Monte-Carlo optimization by CORAL software

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/47

Descriptor: Chemical ratio :47:18

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper. Expected applicability domain of nanomaterials within the range of experimental solubility (supporting information of publication, Table S2). The list of SMILES attributes and their correlation weights can be used to define the applicability domain of examined models: firstly, the models can be used for fullerene derivatives, and secondly, SMILES of these substances must contain attributes which take place in SMILES of the training set.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Carbon-based

List: Fullerene C60 and C70

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: Molecular structure of the molecules are given in Table 1 of the publication.

6.6.Pre-processing of data before modelling:

Not specified in the paper. But from other papers of the same authors:

Three splits into the training and test sets are examined in the present study. These splits obey the following principles:

(i) they are random;

(ii) the ranges of solubility for the training and test sets are similar.

For every split they perform 3 Monte Carlos runs.

6.7.Statistics for goodness-of-fit:

From the three Monte Carlo runs (all similar) the average is presented

(publication's Table 3):

Split1:

$R^2 = 0.8988$

S = 11.275 mg/mL

Split2:

$R^2 = 0.9336$

s = 8.907 mg/mL

Split3:

$R^2 = 0.8963$

s = 10.708 mg/mL

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

Split1 : 9

Split2 : 9

Split3 : 9 MCarbon-based

List

Fullerene C60 and C70

Shape:Spherical

Coating:NA

Size(nm): NA

Other properties:

Molecular structure of the molecules are given in Table 1 of the publication.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

From the three Monte Carlo runs (all similar) the average is presented (publication's Table 3).

Split1:

$R^2 = 0.9064$

$S = 14.407 \text{ mg/mL}$

$R^2_{\text{pred}} = 0.7751$

Split2:

$R^2 = 0.7946$

$S = 18.650 \text{ mg/mL}$

$R^2_{\text{pred}} = 0.7786$

Split3:

$R^2 = 0.9400$

$S = 13.887 \text{ mg/mL}$

$R^2_{\text{pred}} = 0$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Substance 5 is an outlier (model represented here is without this substance). Chemical features of the ligand in this molecule are briefly discussed. Removing 5 improves model for the training set, but not the predictivity ability of the model (validation set).

The threshold is the parameter for separation of SMILES elements into two classes: rare and not rare. We have used Threshold=1. This value indicates that Sk that takes place in the training less than 1 time should be blocked, i.e., its correlation weight should be equal to zero.

SMILES attributes, which are absent in the training set, have no influence on the model. SMILES attributes which are absent in the training set (i.e., attributes which take place only in the validation set) are not involved in the modeling process.

If the correlation weight for the SMILES attribute SA in sequence of the runs of the optimization has values which all are larger than zero, then the attribute can be estimated as stable promoter of increase of the endpoint. Vice versa, if the correlation weight for the SA has in sequence of the runs of optimization values which all are smaller than zero, the attribute can be estimated as a stable promoter of decrease of the given endpoint. Finally, if a SMILES attribute in three runs of the optimization has both correlation weights: smaller and larger than zero values, one can estimate the attribute as an attribute of undefined role.

R^2 : correlation coefficient

S: standard error of estimation

F: Fisher F-ratio

9.2. Bibliography:

Troshin PA, Hoppe H, Renz J et al (2009) Material solubility- photovoltaic performance relationship in the design of novel fullerene derivatives for bulk heterojunction solar cells. *Adv Funct Mater* 19:779–788. doi:10.1002/adfm.200801189

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, SMILE attributes:


- %10, %11, %12, %13, %14, %15, %16, %17, %18, %19, %20, %21, %22, %23, %24, %25, %26, %27, %28, %29, %30, %31, %32, %33, %34, %35, %36, %37, %38

- (, O, -, [, c, C, o, s

, Linear regression model

based on SMILES-based optimal descriptors by the software CORAL.

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 solubility in organic solvents based on quantum-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 solubility in organic solvents based on quantum-chemical and topological descriptors by GA-MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Bakhtiyor F. Rasulev

rasulev@icnanotox.org

2.6. Date of model development and/or publication:

2011

2.7. Reference(s) to main scientific papers and/or software

package:

Petrova, T., Rasulev, B. F., Toropov, A. A., Leszczynska, D., & Leszczynski, J. (2011). Improved model for fullerene C60 solubility in organic solvents based on quantum-chemical and topological descriptors. Journal of Nanoparticle Research, 13(8), 3235–32

<http://doi.org/10.1007/s11051-011-0238-x>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

Solubilities of 122 different solvents are compiled from Liu et al 2005 (published originally in Beck and Mandi 1977). Solubilities are given in terms of logarithmic values of molar fractions (log S) because the log S values correspond to the free energy changes in the solvation process

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

GA-MLR: Genetic Algorithm and Multiple Linear Regression
using BuildQSAR program

4.3.Descriptors in the model:

- T12: topological descriptor, second Mohar index T12.
- X1sol: topological descriptor representing solvation connectivity index. It is a bidimensional descriptor. Encodes several important for solubility characteristics.
- FDI: geometrical descriptor representing a folding degree index. Converges to one for linear molecules (of infinite length) and decreases in accord with the folding degree of the molecule. It can be used as indicator of the degree of departure of a molecule from a strict linearity.
- H052: descriptor is among atom-centered fragments, describing H (hydrogen) attached to C(sp³) with 1X (heteroatom) attached to the next C. It represents the number of hydrogens at the considered molecular fragment.; 4

4.4.Descriptor selection:

Structural based additive descriptors and quantum-chemical descriptors were computed by DRAGON software and Gaussian 03 software by Density Functional Theory methodology, respectively.

- Correlation coefficients for all pairs of descriptors were evaluated to identify highly correlated descriptors and to avoid redundancy. Hence some highly correlated and constant descriptors (cross-correlation $r^2[0.9]$) were removed from the further consideration.
- The descriptors with cross-correlation coefficient larger than 0.6 were avoided

Also within the performance of the model Genetic Algorithm (GA) was applied with the Multiple Linear Regression in order to select the most relevant descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

92/4

Descriptor: Chemical ratio :4:92 ~ 1:23

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper. Experimental solubility range (Table 2 in the

publication).

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

92 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6.Pre-processing of data before modelling:

Random splits are done according to the source publication and Toropov AA, Rasulev BF, Leszczynska D, Leszczynski J (2007c) Additive SMILES based optimal descriptors: QSPR modeling of fullerene C60 solubility in organic solvents. Chem Phys Lett 444:209–214.

6.7.Statistics for goodness-of-fit:

The best model is presented

(4 variables, publication's Table 1):

$r^2 = 0.861$

$s = 0.411$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Leave-one-out technique was applied. The best model is presented (4 variables, publication's Table 1):

$$q^2 = 0.841$$

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

30 MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

The best model is presented (4 variables, publication's Table 1):

$$r^2 = 0.903$$

$$s = 0.355$$

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information

9.1. Comments:

Simple and transparent descriptors and mechanistically interpretable.

Models with 1 to 5 descriptors are generated. The best is described with 4 descriptors.

Although the X1sol descriptor alone displayed the statistically not so significant correlation, it still provides a main contribution to solubility (positive value).

The presence of quantum- chemical descriptor—HOMO energy confirmed the importance of nucleophilic properties of solvents for solubility of C60. Higher HOMO value for solvent results in the higher solubility of C60. (positive influence). Also HOMO-LUMO gap parameter might be used to build individual models for families of solvents (Figure 3 in the publication), despite the aim of the paper was to build general models.

Surprisingly, the folding descriptor FDI showed certain correlation with solubility, asserting that the closer structure of solvent to linearity the higher solubility of C60 in this solvent is.

The presence of nHAcc descriptor in the model confirms importance of H bonds.

Original data was taken by the source publication from Beck MT, Mandi G (1997) Solubility of C60. Fuller Sci Technol 5:291–310

r^2 : correlation coefficient

F: Fisher F-ratio

s: standard error of estimation

q^2 : cross-validation correlation coefficient

GA: Genetic algorithm

MLRA: multiple linear regression analysis

9.2.Bibliography:

Liu H, Yao X, Zhang R, Liu M, Hu Z, Fan B (2005) Accurate quantitative structure-property relationship model to predict the solubility of C60 in various solvents based on a novel approach using a least-squares support vector machine. J Phys Chem B 109:20565–20571

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - T12: topological descriptor, second Mohar index TI2.


- X1sol: topological descriptor representing solvation connectivity index. It is a bidimensional descriptor. Encodes several important for solubility characteristics.

- FDI: geometrical descriptor representing a folding degree index. Converges to one for linear molecules (of infinite length) and decreases in accord with the folding degree of the molecule. It can be used as indicator of the degree of departure of a molecule from a strict linearity.

- H052: descriptor is among atom-centered fragments, describing H (hydrogen) attached to C(sp3) with 1X (heteroatom) attached to the next C. It represents the number of hydrogens at the considered molecular fragment.,GA-MLR: Genetic Algorithm and Multiple Linear Regression

using BuildQSAR program

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Young's modulus prediction in CNTs (taking into account vacancies) by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Young's modulus prediction in CNTs (taking into account vacancies) by PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tammie L. Borders

tammie.l.borders@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Borders, T. L., Fonseca, A. F., Zhang, H., Cho, K., & Rusinko, A. (2013). Developing descriptors to predict mechanical properties of nanotubes. *Journal of Chemical Information and Modeling*, 53(4), 773–782.

<http://doi.org/10.1021/ci300482n>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Young's Modulus

3.3. Comment on endpoint:

Young's modulus (Y) is a measure of stiffness defined as the ratio of axial stress (σ) to axial strain (ϵ)

over ranges of stress in which Hooke's law holds true.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

R_T: Theoretical radius

CA: Chiral angle (radians)

C_N2/C_T: ratio of nonsp² hybridized carbons to total number of carbons; 3

4.4.Descriptor selection:

Descriptors selection was inherent to the building model (PLS), also it was computed a PCA and a Star Plots(see remarks cell) to evaluate the correlation, the relevance and the stability of the different descriptors.

Additionally, model with the whole set of descriptors was computed in order to compare the effect in the accuracy of the model (Table 2 in the publication).

In order to evaluate sensivity of the model to the nanotube's radius, data with range of radius from 0.2 to 0.7 nm, from 0.7 to 2.1 nm and the full set, under the whole set of descriptors were compared in the Table 2.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/3

Descriptor: Chemical ratio :3:53 ~ 1:18

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of CNTs within the range of parameters (descriptors) of the training set, for instance the radius range from 0.2 to 2.1 nm.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Carbon-based

List: CNT: Carbon nanotubes

Shape: Fiber

Coating: NA

Size (nm): Length ~ 10

Radius: range of 0.2 to 2.1

Other info: CNTs were created with JCrystalSoft Nanotube Modeler version 1.6.1 and individually processed with Python scripts to create four types of surface defects: single vacancy, double vacancy, mixed single and double vacancy.

The adaptive intermolecular reactive empirical bond order (AIREBO) potential was used for molecular dynamics (MD) simulations.

The endpoint was calculated, with the set parameters and the obtained ones in the simulations, through a theoretical equation.

6.6. Pre-processing of data before modelling:

From the initial 78 NPs, those ones without functionalization were used to this case. The 66 NPs were splitted in 80% to train the model and 20% to test it.

6.7. Statistics for goodness-of-fit:

$R^2_{\text{Train}} = 0.94$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Y-scrambling to the whole set of descriptors:

$$R^2 = 0.16$$

(Since the final descriptors are the ones which explain almost the total of the variance we consider the result also consistent for the final model)

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20% of data (~13) MCarbon-based

List

CNT: Carbon nanotubes

Shape: Fiber

Coating: NA

Size(nm): Length ~ 10

Radius: range of 0.2 to 2.1

Other properties:

CNTs were created with JCrystalSoft Nanotube Modeler version 1.6.1 and individually processed with Python scripts to create four types of surface defects: single vacancy, double vacancy, mixed single and double vacancy.

The adaptive intermolecular reactive empirical bond order (AIREBO) potential was used for molecular dynamics (MD) simulations.

The endpoint was calculated, with the set parameters and the obtained ones

in the simulations, through a theoretical equation.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2_{\text{Test}} = 0.85$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

A traditional star plot represents a unique data point and is composed of a sequence of equiangular spokes. Each spoke represents a variable with the spoke length proportional to the magnitude of that variable in that data point. Modifying the definition of the traditional star plot, each star represents a variable, and the spokes are observations of that variable over a set of conditions.⁴⁰ Spoke length is proportional to the normalized descriptor strength and provides a measure of descriptor importance. The spoke color indicates the correlation of the descriptor in the model with blue being positive (correlated) and red being negative (anticorrelated).

One of the aims of the work was to obtain the critical descriptors to obtain the desired endpoint, which was also applied in a third analysis against experimental data with the obtained descriptors, but the obtained model was not applied in that analysis.

An exhaustive analysis of the descriptors was done, which could be interpreted as an Mechanistic Interpretation.

NPs: Nanoparticles

PCA: Principal Component Analysis

PLS: Partial Least Squares

R^2 : Correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, R_T: Theoretical radius

CA: Chiral angle (radians)

C_N2/C_T: ratio of nonsp² hybridized carbons to total number of carbons,PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Poisson's ratio prediction in CNTs (taking into account vacancies) by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Poisson's ratio prediction in CNTs (taking into account vacancies) by PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tammie L. Borders

tammie.l.borders@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Borders, T. L., Fonseca, A. F., Zhang, H., Cho, K., & Rusinko, A. (2013). Developing descriptors to predict mechanical properties of nanotubes. *Journal of Chemical Information and Modeling*, 53(4), 773–782.

<http://doi.org/10.1021/ci300482n>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Poisson's ratio

3.3. Comment on endpoint:

Poisson's ratio (PR) is the negative ratio of the lateral to longitudinal strain as an axial load is applied.

For this study, values were calculated at an initial and final strain of 0% and 5%, respectively.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

R_T: Theoretical radius

CA: Chiral angle (radians)

C_M/C_T: ratio of missing carbons to total number of carbons; 3

4.4.Descriptor selection:

Descriptors selection was inherent to the building model (PLS), also it was computed a PCA and a Star Plots (see remarks cell) to evaluate the correlation, the relevance and the stability of the different descriptors.

Additionally, model with the whole set of descriptors was computed in order to compare the effect in the accuracy of the model (Table 2 in the publication).

In order to evaluate sensitivity of the model to the nanotube's radius, data with range of radius from 0.2 to 0.7 nm, from 0.7 to 2.1 nm and the full set, under the whole set of descriptors were compared in the Table 2.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/3

Descriptor: Chemical ratio :3:53 ~ 1:18

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of CNTs within the range of parameters (descriptors) of the training set, for instance the radius range from 0.2 to 2.1 nm.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

0 Carbon-based

List: CNT: Carbon nanotubes

Shape: Fiber

Coating: NA

Size (nm): Length ~ 10

Radius: range of 0.2 to 2.1

Other info: CNTs were created with JCrystalSoft Nanotube Modeler version 1.6.1 and individually processed with Python scripts to create four types of surface defects: single vacancy, double vacancy, mixed single and double vacancy.

The adaptive intermolecular reactive empirical bond order (AIREBO) potential was used for molecular dynamics (MD) simulations.

The endpoint was calculated, with the set parameters and the obtained ones in the simulations, through a theoretical equation.

6.6. Pre-processing of data before modelling:

From the initial 78 NPs, those ones without functionalization were used to this case. The 66 NPs were splitted in 80% to train the model and 20% to test it.

6.7. Statistics for goodness-of-fit:

$R^2_{\text{Train}} = 0.85$

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

Y-scrambling to the whole set of descriptors:

$$R^2 = 0.16$$

(Since the final descriptors are the ones which explain almost the total of the variance we consider the result also consistent for the final model)

7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20% of data (~13) MCarbon-based

List

CNT: Carbon nanotubes

Shape: Fiber

Coating: NA

Size(nm): Length ~ 10

Radius: range of 0.2 to 2.1

Other properties:

CNTs were created with JCrystalSoft Nanotube Modeler version 1.6.1 and individually processed with Python scripts to create four types of surface defects: single vacancy, double vacancy, mixed single and double vacancy.

The adaptive intermolecular reactive empirical bond order (AIREBO) potential was used for molecular dynamics (MD) simulations.

The endpoint was calculated, with the set parameters and the obtained ones

in the simulations, through a theoretical equation.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2_{\text{Test}} = 0.94$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

A traditional star plot represents a unique data point and is composed of a sequence of equiangular spokes. Each spoke represents a variable with the spoke length proportional to the magnitude of that variable in that data point. Modifying the definition of the traditional star plot, each star represents a variable, and the spokes are observations of that variable over a set of conditions.⁴⁰ Spoke length is proportional to the normalized descriptor strength and provides a measure of descriptor importance. The spoke color indicates the correlation of the descriptor in the model with blue being positive (correlated) and red being negative (anticorrelated).

One of the aims of the work was to obtain the critical descriptors to obtain the desired endpoint, which was also applied in a third analysis against experimental data with the obtained descriptors, but the obtained model was not applied in that analysis.

An exhaustive analysis of the descriptors was done, which could be interpreted as a Mechanistic Interpretation.

NPs: Nanoparticles

PCA: Principal Component Analysis

PLS: Partial Least Squares

R^2 : Correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, R_T: Theoretical radius

CA: Chiral angle (radians)

C_M/C_T: ratio of missingcarbons to total number of carbons,PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Young's modulus prediction in CNTs (taking into account vacancies
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Young's modulus prediction in CNTs (taking into account vacancies plus methyl functionalization) by PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tammie L. Borders

tammie.l.borders@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software

package:

Borders, T. L., Fonseca, A. F., Zhang, H., Cho, K., & Rusinko, A. (2013). Developing descriptors to predict mechanical properties of nanotubes. *Journal of Chemical Information and Modeling*, 53(4), 773–782.

<http://doi.org/10.1021/ci300482n>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Young's Modulus

3.3. Comment on endpoint:

Young's modulus (Y) is a measure of stiffness defined as the ratio of axial stress (σ) to axial strain (ϵ) over ranges of stress in which Hooke's law holds true.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

CA: Chiral angle (radians)

C_N2/C_T: ratio of nonsp² hybridized carbons to total number of carbons

M_N/C_T: ratio of methyl groups to total number of carbons; 3

4.4.Descriptor selection:

Descriptors selection was inherent to the building model (PLS), also it was computed a PCA and a Star Plots(see remarks cell) to evaluate the correlation, the relevance and the stability of the different descriptors.

Additionally, model with the whole set of descriptors was computed in order to compare the effect in the accuracy of the model (Table 5 in the publication).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/3

Descriptor: Chemical ratio :3:43 ~ 1:14

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of CNTs within the range of parameters (descriptors) of the training set, for instance the radius range from 0.35 to 0.7 nm.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Carbon-based

List: CNT: Carbon nanotubes

Shape: Fiber

Coating: Methyl functionalization

Size (nm): Length ~ 10

Radius: range of 0.2 to 2.1

Other info: CNTs were created with JCrystalSoft Nanotube Modeler version 1.6.1 and individually processed with Python scripts to create four types of surface defects: single vacancy, double vacancy, mixed single and double vacancy, and methyl functionalization.

The adaptive intermolecular reactive empirical bond order (AIREBO) potential was used for molecular dynamics (MD) simulations.

The endpoint was calculated, with the set parameters and the obtained ones in the simulations, through a theoretical equation.

6.6.Pre-processing of data before modelling:

From the initial 78 NPs, those ones within the radius range 0.35 to 0.7 nm were used to this case. The 43 NPs were splitted in 80% to train the model and 20% to test it.

6.7.Statistics for goodness-of-fit:

R^2_{Train} = 0.92

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20% of data (~11) MCarbon-based

List

CNT: Carbon nanotubes

Shape: Fiber

Coating: Methyl functionalization

Size(nm): Length ~ 10

Radius: range of 0.2 to 2.1

Other properties:

CNTs were created with JCrystalSoft Nanotube Modeler version 1.6.1 and individually processed with Python scripts to create four types of surface defects: single vacancy, double vacancy, mixed single and double vacancy, and methyl functionalization.

The adaptive intermolecular reactive empirical bond order (AIREBO) potential was used for molecular dynamics (MD) simulations.

The endpoint was calculated, with the set parameters and the obtained ones in the simulations, through a theoretical equation.

7.6. Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2_{\text{Test}} = 0.93$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

A traditional star plot represents a unique data point and is composed of a sequence of equiangular spokes. Each spoke represents a variable with the spoke length proportional to the magnitude of that variable in that data point. Modifying the definition of the traditional star plot, each star represents a variable, and the spokes are observations of that variable over a set of conditions.⁴⁰ Spoke length is proportional to the normalized descriptor strength and provides a measure of descriptor importance. The spoke color indicates the correlation of the descriptor in the model with blue being positive (correlated) and red being negative (anticorrelated).

One of the aims of the work was to obtain the critical descriptors to obtain the desired endpoint, which was also applied in a third analysis against experimental data with the obtained descriptors, but the obtained model was not applied in that analysis.

An exhaustive analysis of the descriptors was done, which could be interpreted as a Mechanistic Interpretation.

NPs: Nanoparticles

PCA: Principal Component Analysis

PLS: Partial Least Squares

R^2 : Correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, CA: Chiral angle (radians)

C_N2/C_T: ratio of nonsp² hybridized carbons to total number of carbons

M_N/C_T: ratio of methyl groups to total number of carbons,PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Poisson's ratio prediction in CNTs (taking into account vacancies plus
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Poisson's ratio prediction in CNTs (taking into account vacancies plus methyl functionalization) by PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tammie L. Borders

tammie.l.borders@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software

package:

Borders, T. L., Fonseca, A. F., Zhang, H., Cho, K., & Rusinko, A. (2013). Developing descriptors to predict mechanical properties of nanotubes. *Journal of Chemical Information and Modeling*, 53(4), 773–782.

<http://doi.org/10.1021/ci300482n>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Poisson's ratio

3.3. Comment on endpoint:

Poisson's ratio (PR) is the negative ratio of the lateral to longitudinal strain as an axial load is applied. For this study, values were calculated at an initial and final strain of 0% and 5%, respectively.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

CA: Chiral angle (radians)

C_N2/C_T: ratio of nonsp² hybridized carbons to total number of carbons

M_N/C_T: ratio of methyl groups to total number of carbons; 3

4.4.Descriptor selection:

Descriptors selection was inherent to the building model (PLS), also it was computed a PCA and a Star Plots(see remarks cell) to evaluate the correlation, the relevance and the stability of the different descriptors.

Additionally, model with the whole set of descriptors was computed in order to compare the effect in the accuracy of the model.

In order to evaluate sensivity of the model to the nanotube's radius, data with range of radius from 0.2 to 0.7 nm, from 0.7 to 2.1 nm and the full set, under the whole set of descriptors were compared in the publication's Table 2.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

0/3

Descriptor: Chemical ratio :3:43 ~ 1:14

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of CNTs within the range of parameters (descriptors) of the training set, for instance the radius range from 0.35 to 0.7 nm.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

0 Carbon-based

List: CNT: Carbon nanotubes

Shape: Fiber

Coating: Methyl functionalization

Size (nm): Length ~ 10

Radius: range of 0.2 to 2.1

Other info: CNTs were created with JCrystalSoft Nanotube Modeler version 1.6.1 and individually processed with Python scripts to create four types of surface defects: single vacancy, double vacancy, mixed single and double vacancy, and methyl functionalization.

The adaptive intermolecular reactive empirical bond order (AIREBO) potential was used for molecular dynamics (MD) simulations.

The endpoint was calculated, with the set parameters and the obtained ones in the simulations, through a theoretical equation.

6.6.Pre-processing of data before modelling:

From the initial 78 NPs, those ones within the radius range 0.35 to 0.7 nm were used to this case. The 43 NPs were splitted in 80% to train the model and 20% to test it.

6.7.Statistics for goodness-of-fit:

$R^2_{\text{Train}} = 0.91$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

20% of data (~11) MCarbon-based

List

CNT: Carbon nanotubes

Shape: Fiber

Coating: Methyl functionalization

Size(nm): Length ~ 10

Radius: range of 0.2 to 2.1

Other properties:

CNTs were created with JCrystalSoft Nanotube Modeler version 1.6.1 and individually processed with Python scripts to create four types of surface defects: single vacancy, double vacancy, mixed single and double vacancy, and methyl functionalization.

The adaptive intermolecular reactive empirical bond order (AIREBO) potential was used for molecular dynamics (MD) simulations.

The endpoint was calculated, with the set parameters and the obtained ones in the simulations, through a theoretical equation.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2_{\text{Test}} = 0.81$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

A traditional star plot represents a unique data point and is composed of a sequence of equiangular spokes. Each spoke represents a variable with the spoke length proportional to the magnitude of that variable in that data point. Modifying the definition of the traditional star plot, each star represents a variable, and the spokes are observations of that variable over a set of conditions.⁴⁰ Spoke length is proportional to the normalized descriptor strength and provides a measure of descriptor importance. The spoke color indicates the correlation of the descriptor in the model with blue being positive (correlated) and red being negative (anticorrelated).

One of the aims of the work was to obtain the critical descriptors to obtain the desired endpoint, which was also applied in a third analysis against experimental data with the obtained descriptors, but the obtained model was not applied in that analysis.

An exhaustive analysis of the descriptors was done, which could be interpreted as a Mechanistic Interpretation.

NPs: Nanoparticles

PCA: Principal Component Analysis

PLS: Partial Least Squares

R^2 : Correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, CA: Chiral angle (radians)

C_N2/C_T: ratio of nonsp² hybridized carbons to total number of carbons

M_N/C_T: ratio of methyl groups to total number of carbons,PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: LSER model to predict the Solubility of fullerene C60 in various
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

LSER model to predict the Solubility of fullerene C60 in various solvents by stepwise-MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

S. Yousefinejad

yousefinejad.s@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Yousefinejad, S., Honarasa, F., Abbasitabar, F., & Arianezhad, Z. (2013). New LSER model based on solvent empirical parameters for the prediction and description of the solubility of buckminsterfullerene in various solvents. Journal of Solution Chemistry,

<http://doi.org/10.1007/s10953-013-0062-2>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in organic solvents

3.3. Comment on endpoint:

Solubilities of C60 in different solvents (Table 1 in the publication) were screened from a previous work (source reference) already reported in the table. From the whole set of solvents, only those ones which fit with the desired initial descriptors were selected for the study. Solubilities are given in terms of logarithmic values of molar fractions (log S) because the log S values correspond to the free energy changes in the solvation process.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Stepwise-MLR (Multiple Linear Regression)

by SPSS v15.0 and MATLAB v7 softwares

4.3.Descriptors in the model:

- XX: Dispersion and dipolar interactions. Solvent induced frequency shifts so SO₂
- log₁₀(L¹⁶): Dispersion interactions. Based on the logarithmic gas–liquid partition coefficient in n-hexadecane
- e(SVB): Branching of the interactions. Average equilibrium and chromatographic distribution constants on amberlite XAD-2, SM-2 and XAD-4
- α: Hydrogen bond donation ability. Solvatochromic parameter of solvent HBD (hydrogen bond donor) acidity
- B_KT: Basicity. Calculated from the difference of the longest wavelength band in the UV–Vis spectra measured for p-nitroaniline and N,N-diethyl-p-nitroaniline; 5

4.4.Descriptor selection:

Initial 127 solvents scales (descriptors, empirical parameters) selected from literature:

Katritzky, A.R., Fara, D.C., Kuanar, M., Hur, E., Karelson, M.: The classification of solvents by combining classical QSPR methodology with principal component analysis. J. Phys. Chem. A 109, 10323–10341 (2005).

The initial descriptors were screened in three steps:

- Before the model procedure, correlations among descriptors and the endpoint (log S) were examined, and the ones with the highest correlations were kept.
- Within the model building, stepwise-MLR, using the statistical values of R² and Q² the best number of descriptors were screened.
- Finally, the selected descriptors were evaluated by VIF (Variance Inflation Factor) to examine the multicollinearity and the relative importance of each one of the descriptors by the value MF (Mean Effect)

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

61/5

Descriptor: Chemical ratio :5:79 ~ 1:16

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 4)

$$h^* = 0.29$$

Two solvents (tetradecane and n-butylamine) were outside the AD of the model. The

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

61 Carbon-based

List: Fullerene C60

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: NA

6.6. Pre-processing of data before modelling:

The data set was divided into training (80 % of total solvents) and external test set (20 % of total solvents) in order to test the final model performance. These solvents were selected based on descriptor spaces. To do so, the data matrix containing the total descriptors was subjected to principal component analysis and the first two principal components (PCs) were plotted against each other. Among the points in the resulting plot, a homogenous set of solvents were selected as test samples.

A Leave-one-out Cross-validation was also applied to the training set.

Two of solvents were outside the AD, which were removed from the training set (initial number of training points: 63)

6.7.Statistics for goodness-of-fit:

$R^2_{cal} = 0.85$

RMSEC = 0.44

$F = 58.28$ ($F_{crit} = 2.38$)

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LOO} = 0.79$

RMSE_CV = 0.49

50 Y-Randomization:

$Q^2_{MP} = 0.23$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

16 MCarbon-based

List

Fullerene C60

Shape: Spherical

Coating: NA

Size(nm): NA

Other properties:

NA

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

$Q^2_{\text{ext}} = 0.87$

RMSEP = 0.37

7.8. Predictivity - Assessment of the external validation set:

No information available

7.9. Comments on the external validation of the model:

No information available

8. Providing a mechanistic interpretation - OECD Principle 5**8.1. Mechanistic basis of the model:**

No information available

8.2. A priori or a posteriori mechanistic interpretation:

No information available

8.3. Other information about the mechanistic interpretation:

No additional information available

9. Miscellaneous information**9.1. Comments:**

A previous model was presented in the publication before the Applicability Domain, and the statistical result are compared. A little improvement was obtained after the outliers were excluded.

The model was constructed based on some empirical parameters that were obtained using some experimental probes and these parameters can define some aspects of solvent interactions with the solute (here fullerene C60). Thus it could be said that the proposed model is a kind of LSER

A good Mechanistic Interpretation was provided.

R^2_{cal} : Calibration correlation coefficient

RMSEC: Calibration root-mean-square errors

Q^2_{LOO} : Leave-one-out cross-validation correlation coefficient

RMSE_CV: Leave-one-out cross-validation root-mean-square errors

Q²_{ext}: Correlation coefficient

9.2. Bibliography:

(already reported in this table)

Kiss, I. Z., Mandi, G., & Beck, M. T. (2000). Artificial Neural Network approach to predict the solubility of C-60 in various solvents. *Journal Of Physical Chemistry A*, 104, 8081–8088.

Kiss et al., catch their data from a previous work:

Beck, M. T.; Mandi, G. *Fullerene Sci. Technol.* 1997,5

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC


10.2. Publication date:

To be entered by JRC

10.3. Keywords:

NA, NA, QSPR, - XX: Dispersion and dipolar interactions. Solvent induced frequency shifts so SO₂
 - log₁₀(L¹⁶): Dispersion interactions. Based on the logarithmic gas–liquid partition coefficient in n-hexadecane
 - e(SVB): Branching of the interactions. Average equilibrium and chromatographic distribution constants on amberlite XAD-2, SM-2 and XAD-4
 - α: Hydrogen bond donation ability. Solvatochromic parameter of solvent HBD (hydrogen bond donor) acidity
 - B_{KT}: Basicity. Calculated from the difference of the longest wavelength band in the UV–Vis spectra measured for p-nitroaniline and N,N-diethyl-p-nitroaniline, Stepwise-MLR (Multiple Linear Regression)
 by SPSS v15.0 and MATLAB v7 softwares

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting phase behavior of amphiphilic nanostructured nanoparticle
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting phase behavior of amphiphilic nanostructured nanoparticle drug delivery vehicles by BRANN

(phytantriol batch 1 case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

dave.winkler@csiro.au

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Mulet, X., Burden, F. R., & Winkler, D. A. (2013).

Predicting the complex phase behavior of self-assembling drug delivery nanoparticles. *Molecular Pharmaceutics*, 10(4), 1368–1377.

(phytantriol batch 1 case)

<http://doi.org/10.1021/mp3006402>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

LC structure

3.3.Comment on endpoint:

Predicted the effect of different drugs on the complex phase behavior of lyotropic lipid-based LC (Liquied Crystal) nanoparticles

Two experiments used phytantriol from different batches to assess effects of batch-to-batch variability; another used monoolein, and the last used Myverol, the commercial product containing monoolein. These amphiphilic materials were used to prepare inverse-bicontinuous cubic and inverse-hexagonal liquid-crystalline nanoparticles loaded with 10 commonly used drugs. These drugs had a wide range of structures and lipophilicities and were loaded at six concentrations (0, 1, 2, 5, 10, and 15 mol %) and at two temperatures (25 and 37 °C).

The applied drugs for the training and testing were the following: levofloxacin, prednisolone, hydrocortisone, atropine, dexamethasone, diazepam, progesterone, indometacin, haloperidol, transretinol; and chlorambucil, cimetidine, β -estradiol, androsterone, flumequine, nifedipine, ibuprofen, curcumin, histamine, dopamine, calcein (Sigma-Aldrich).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

BRANN: Bayesian Regularization Artificial Neural Network

4.3.Descriptors in the model:

NA; 0

4.4.Descriptor selection:

Theoretical-based molecular descriptors, computed molecular descriptors by DRAGON software and experimental parameters were used as inputs for the BRANN model. No reduction of descriptors was applied for BRANN. MLREM model was computed but the performance was not as much accurate as the nonlienar model BRANN. However, the most relevant descriptors of both models were represented in the publication's Figure 5 where the overlaped descriptors from both models are detected, thus it was defined that those descriptors will be the most relevant for identify each of the different phases.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

96/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of drugs within the range of

parameters (descriptors) of the training set and for the specific lipid of the drug-loaded nanoparticle, phytantriol.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

96 Lipid-based

List: Phytantriol

Shape: NA

Coating: NA

Size (nm): NA

Other info: The internal liquid crystalline structure of the dispersed particles was determined by using SAXS (small-angle X-ray scattering) which provides information about the symmetry and the lattice parameter (unit cell size) of the internal structure.

The different structural phases used as endpoint (symmetry inside brackets) were:

- Inverse-bicontinuous diamond Q_{II}^D (Pn3m) cubic
- Hexagonal H_{II}
- Gyroid Q_{II}^G (Ia3d) cubic
- Crystalline L_c

6.6.Pre-processing of data before modelling:

The data points come from a combination of 10 drugs at 6 different concentrations, 2 different temperatures and as was reported by the authors, multiplied by the 4 possible phases, since a model for each phase was computed (4 endpoints) and all together were used as the final model.

From the 480 data points (120 for each phase model) the data was splitted in 80% for training set and 20% for external validation using K-means clustering.

6.7.Statistics for goodness-of-fit:

Accuracy = 99.4 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MLipid-based

List

Phytantriol

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

The internal liquid crystalline structure of the dispersed particles was determined by using SAXS (small-angle X-ray scattering) which provides information about the symmetry and the lattice parameter (unit cell size) of the internal structure.

The different structural phases used as endpoint (symmetry inside brackets) were:

- Inverse-bicontinuous diamond Q_{II}^D (Pn3m) cubic
- Hexagonal H_{II}
- Gyroid Q_{II}^G (Ia3d) cubic
- Crystalline L_c

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Accuracy_test = 98.96 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

The statistical parameters were not clearly reported for our classification, then the reported values were not obtained directly in the publication.

Due to the lack of robustness validation and the lack of information about the final applied descriptors, those were not classified and the reliability of the model can not be ensured..

SAXS: small-angle X-ray scattering

BRANN: Bayesian regularization Artificial Neural Network

MLREM: Multiple Linear Regression with Expectation Maximization

Accuracy_test : accuracy classification for test set (from splitted data)

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, NA,BRANN: Bayesian Regularization Artificial Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting phase behavior of amphiphilic nanostructured nanoparticle
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting phase behavior of amphiphilic nanostructured nanoparticle drug delivery vehicles by BRANN

(phytantriol batch 2 case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

dave.winkler@csiro.au

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Mulet, X., Burden, F. R., & Winkler, D. A. (2013).

Predicting the complex phase behavior of self-assembling drug delivery nanoparticles. *Molecular Pharmaceutics*, 10(4), 1368–1377.

(phytantriol batch 2 case)

<http://doi.org/10.1021/mp3006402>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

LC structure

3.3.Comment on endpoint:

Predicted the effect of different drugs on the complex phase behavior of lyotropic lipid-based LC (Liquied Crystal) nanoparticles

Two experiments used phytantriol from different batches to assess effects of batch-to-batch variability; another used monoolein, and the last used Myverol, the commercial product containing monoolein. These amphiphilic materials were used to prepare inverse-bicontinuous cubic and inverse-hexagonal liquid-crystalline nanoparticles loaded with 10 commonly used drugs. These drugs had a wide range of structures and lipophilicities and were loaded at six concentrations (0, 1, 2, 5, 10, and 15 mol %) and at two temperatures (25 and 37 °C).

The applied drugs for the training and testing were the following: levofloxacin, prednisolone, hydrocortisone, atropine, dexamethasone, diazepam, progesterone, indometacin, haloperidol, transretinol; and chlorambucil, cimetidine, β -estradiol, androsterone, flumequine, nifedipine, ibuprofen, curcumin, histamine, dopamine, calcein (Sigma-Aldrich).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

BRANN: Bayesian Regularization Artificial Neural Network

4.3.Descriptors in the model:

NA; 0

4.4.Descriptor selection:

Theoretical-based molecular descriptors, computed molecular descriptors by DRAGON software and experimental parameters were used as inputs for the BRANN model. No reduction of descriptors was applied for BRANN. MLREM model was computed but the performance was not as much accurate as the nonlienar model BRANN. However, the most relevant descriptors of both models were represented in the publication's Figure 5 where the overlaped descriptors from both models are detected, thus it was defined that those descriptors will be the most relevant for identify each of the different phases.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

96/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of drugs within the range of

parameters (descriptors) of the training set and for the specific lipid of the drug-loaded nanoparticle, phytantriol.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

96 Lipid-based

List: Phytantriol

Shape: NA

Coating: NA

Size (nm): NA

Other info: The internal liquid crystalline structure of the dispersed particles was determined by using SAXS (small-angle X-ray scattering) which provides information about the symmetry and the lattice parameter (unit cell size) of the internal structure.

The different structural phases used as endpoint (symmetry inside brackets) were:

- Inverse-bicontinuous diamond Q_{II}^D (Pn3m) cubic
- Hexagonal H_{II}
- Gyroid Q_{II}^G (Ia3d) cubic
- Fluid isotropic FI

6.6.Pre-processing of data before modelling:

The data points come from a combination of 10 drugs at 6 different concentrations, 2 different temperatures and as was reported by the authors, multiplied by the 4 possible phases, since a model for each phase was computed (4 endpoints) and all together were used as the final model.

From the 480 data points (120 for each phase model) the data was splitted in 80% for training set and 20% for external validation using K-means clustering.

Then also 11 extra drugs (at 6 different concentrations) were tested with the already trained model.

6.7.Statistics for goodness-of-fit:

Accuracy = 98.96 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

132 MLipid-based

List

Phytantriol

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

The internal liquid crystalline structure of the dispersed particles was determined by using SAXS (small-angle X-ray scattering) which provides information about the symmetry and the lattice parameter (unit cell size) of the internal structure.

The different structural phases used as endpoint (symmetry inside brackets) were:

- Inverse-bicontinuous diamond Q_{II}^D (Pn3m) cubic
- Hexagonal H_{II}
- Gyroid Q_{II}^G (Ia3d) cubic
- Fluid isotropic FI

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Accuracy_{test} = 98.96 %

Accuracy_{extra} = 91 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The statistical parameters were not clearly reported for our classification, then the reported values were not obtained directly in the publication.

Due to the lack of robustness validation and the lack of information about the final applied descriptors, those were not classified and the reliability of the model can not be ensured..

SAXS: small-angle X-ray scattering

BRANN: Bayesian regularization Artificial Neural Network

MLREM: Multiple Linear Regression with Expectation Maximization

Accuracy_test : accuracy classification for test set (from splitted data)

Accuracy_extra : accu

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, NA,BRANN: Bayesian Regularization Artificial Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting phase behavior of amphiphilic nanostructured nanoparticle
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting phase behavior of amphiphilic nanostructured nanoparticle drug delivery vehicles by BRANN

(monoolein case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

dave.winkler@csiro.au

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Mulet, X., Burden, F. R., & Winkler, D. A. (2013).

Predicting the complex phase behavior of self-assembling drug delivery nanoparticles. *Molecular Pharmaceutics*, 10(4), 1368–1377.

(monoolein case)

<http://doi.org/10.1021/mp3006402>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

LC structure

3.3.Comment on endpoint:

Predicted the effect of different drugs on the complex phase behavior of lyotropic lipid-based LC (Liquied Crystal) nanoparticles

Two experiments used phytantriol from different batches to assess effects of batch-to-batch variability; another used monoolein, and the last used Myverol, the commercial product containing monoolein. These amphiphilic materials were used to prepare inverse-bicontinuous cubic and inverse-hexagonal liquid-crystalline nanoparticles loaded with 10 commonly used drugs. These drugs had a wide range of structures and lipophilicities and were loaded at six concentrations (0, 1, 2, 5, 10, and 15 mol %) and at two temperatures (25 and 37 °C).

The applied drugs for the training and testing were the following: levofloxacin, prednisolone, hydrocortisone, atropine, dexamethasone, diazepam, progesterone, indometacin, haloperidol, transretinol; and chlorambucil, cimetidine, β -estradiol, androsterone, flumequine, nifedipine, ibuprofen, curcumin, histamine, dopamine, calcein (Sigma-Aldrich).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

BRANN: Bayesian Regularization Artificial Neural Network

4.3.Descriptors in the model:

NA; 0

4.4.Descriptor selection:

Theoretical-based molecular descriptors, computed molecular descriptors by DRAGON software and experimental parameters were used as inputs for the BRANN model. No reduction of descriptors was applied for BRANN. MLREM model was computed but the performance was not as much accurate as the nonlienar model BRANN. However, the most relevant descriptors of both models were represented in the publication's Figure 5 where the overlaped descriptors from both models are detected, thus it was defined that those descriptors will be the most relevant for identify each of the different phases.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

96/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of drugs within the range of

parameters (descriptors) of the training set and for the specific lipid of the drug-loaded nanoparticle, monoolein.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

96 Lipid-based

List: Monoolein

Shape: NA

Coating: NA

Size (nm): NA

Other info: The internal liquid crystalline structure of the dispersed particles was determined by using SAXS (small-angle X-ray scattering) which provides information about the symmetry and the lattice parameter (unit cell size) of the internal structure.

The different structural phases used as endpoint (symmetry inside brackets) were:

- Inverse-bicontinuous diamond Q_{II}^D (Pn3m) cubic
- Hexagonal H_{II}
- Primitive Q_{II}^P (Im3m) cubic

6.6.Pre-processing of data before modelling:

The data points come from a combination of 10 drugs at 6 different concentrations, 2 different temperatures and as was reported by the authors, multiplied by the 3 possible phases, since a model for each phase was computed (3 endpoints) and all together were used as the final model.

From the 360 data points (120 for each phase model) the data was splitted in 80% for training set and 20% for external validation using K-means clustering.

Then also 11 extra drugs (at 6 different concentrations) were tested with the already trained model.

6.7.Statistics for goodness-of-fit:

Accuracy = 100.00 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

132 MLipid-based

List

Monoolein

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

The internal liquid crystalline structure of the dispersed particles was determined by using SAXS (small-angle X-ray scattering) which provides information about the symmetry and the lattice parameter (unit cell size) of the internal structure.

The different structural phases used as endpoint (symmetry inside brackets) were:

- Inverse-bicontinuous diamond Q_{II}^D (Pn3m) cubic
- Hexagonal H_{II}
- Primitive Q_{II}^P (Im3m) cubic

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Accuracy_{test} = 100.00 %

Accuracy_{extra} = 85 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

The statistical parameters were not clearly reported for our classification, then the reported values were not obtained directly in the publication.

Due to the lack of robustness validation and the lack of information about the final applied descriptors, those were not classified and the reliability of the model can not be ensured..

SAXS: small-angle X-ray scattering

BRANN: Bayesian regularization Artificial Neural Network

MLREM: Multiple Linear Regression with Expectation Maximization

Accuracy_{test} : accuracy classification for test set (from splitted data)

Accuracy_extra : accu

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, NA,BRANN: Bayesian Regularization Artificial Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting phase behavior of amphiphilic nanostructured nanoparticle
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting phase behavior of amphiphilic nanostructured nanoparticle drug delivery vehicles by BRANN

(Myverol [monoolein with impurities] case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

D.A. Winkler

dave.winkler@csiro.au

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Le, T. C., Mulet, X., Burden, F. R., & Winkler, D. A. (2013).

Predicting the complex phase behavior of self-assembling drug delivery nanoparticles. *Molecular Pharmaceutics*, 10(4), 1368–1377.

(Myverol [monoolein with impurities] case)

<http://doi.org/10.1021/mp3006402>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

LC structure

3.3.Comment on endpoint:

Predicted the effect of different drugs on the complex phase behavior of lyotropic lipid-based LC (Liquid Crystal) nanoparticles

Two experiments used phytantriol from different batches to assess effects of batch-to-batch variability; another used monoolein, and the last used Myverol, the commercial product containing monoolein. These amphiphilic materials were used to prepare inverse-bicontinuous cubic and inverse-hexagonal liquid-crystalline nanoparticles loaded with 10 commonly used drugs. These drugs had a wide range of structures and lipophilicities and were loaded at six concentrations (0, 1, 2, 5, 10, and 15 mol %) and at two temperatures (25 and 37 °C).

The applied drugs for the training and testing were the following: levofloxacin, prednisolone, hydrocortisone, atropine, dexamethasone, diazepam, progesterone, indometacin, haloperidol, transretinol; and chlorambucil, cimetidine, β -estradiol, androsterone, flumequine, nifedipine, ibuprofen, curcumin, histamine, dopamine, calcein (Sigma-Aldrich).

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

BRANN: Bayesian Regularization Artificial Neural Network

4.3.Descriptors in the model:

NA; 0

4.4.Descriptor selection:

Theoretical-based molecular descriptors, computed molecular descriptors by DRAGON software and experimental parameters were used as inputs for the BRANN model. No reduction of descriptors was applied for BRANN. MLREM model was computed but the performance was not as much accurate as the nonlinear model BRANN. However, the most relevant descriptors of both models were represented in the publication's Figure 5 where the overlapped descriptors from both models are detected, thus it was defined that those descriptors will be the most relevant for identify each of the different phases.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

96/0

Descriptor: Chemical ratio :NA

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper.

Expected an applicability domain of drugs within the range of

parameters (descriptors) of the training set and for the specific lipid of the drug-loaded nanoparticle, Myverol [monoolein with impurities] .

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

96 Lipid-based

List: Monoolein

Shape: NA

Coating: NA

Size (nm): NA

Other info: The internal liquid crystalline structure of the dispersed particles was determined by using SAXS (small-angle X-ray scattering) which provides information about the symmetry and the lattice parameter (unit cell size) of the internal structure.

The different structural phases used as endpoint (symmetry inside brackets) were:

- Inverse-bicontinuous diamond Q_{II}^D (Pn3m) cubic
- Hexagonal H_{II}
- Primitive Q_{II}^P (Im3m) cubic

6.6.Pre-processing of data before modelling:

The data points come from a combination of 10 drugs at 6 different concentrations, 2 different temperatures and as was reported by the authors, multiplied by the 3 possible phases, since a model for each phase was computed (3 endpoints) and all together were used as the final model using K-means clustering.

From the 360 data points (120 for each phase model) the data was splitted in 80% for training set and 20% for external validation.

6.7.Statistics for goodness-of-fit:

Accuracy = 95.83 %

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MLipid-based

List

Monoolein

Shape:NA

Coating:NA

Size(nm): NA

Other properties:

The internal liquid crystalline structure of the dispersed particles was determined by using SAXS (small-angle X-ray scattering) which provides information about the symmetry and the lattice parameter (unit cell size) of the internal structure.

The different structural phases used as endpoint (symmetry inside brackets) were:

- Inverse-bicontinuous diamond Q_{II}^D (Pn3m) cubic
- Hexagonal H_{II}
- Primitive Q_{II}^P (Im3m) cubic

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

Accuracy_{test} = 83.33 %

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

The statistical parameters were not clearly reported for our classification, then the reported values were not obtained directly in the publication.

Due to the lack of robustness validation and the lack of information about the final applied descriptors, those were not classified and the reliability of the model can not be ensured..

SAXS: small-angle X-ray scattering

BRANN: Bayesian regularization Artificial Neural Network

MLREM: Multiple Linear Regression with Expectation Maximization

Accuracy_{test} : accuracy classification for test set (from splitted data)

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, NA,BRANN: Bayesian Regularization Artificial Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting dispersion of SWNTs in different organic solvents by MLR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting dispersion of SWNTs in different organic solvents by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

M. Salahinejad

salahinejad@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Salahinejad, M., & Zolfonoun, E. (2013). QSAR studies of the dispersion of SWNTs in different organic solvents. Journal of Nanoparticle Research, 15(11).

<http://doi.org/10.1007/s11051-013-2028-0>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Dispersion in organic solvents

3.3. Comment on endpoint:

Dispersion and solubilization are different phenomena but they are used interchangeably in the literature. Based on diameter and length of CNTs (Carbon Nanotubes) and their ability to form colloidal solutions, Geckeler and Premkumar recommended to use the dispersion term.

The dispersion of CNTs in different organic solvents was determined measuring the dispersion absorbance as concentration (C_max) of SWNTs (Single-Walled Nanotubes) after sonication and mild centrifugation. Detailed description of the method reported as reference in the publication (Giordani et al. 2006; Bergin et al. 2008)

The Log C_max was used as endpoint value in this model.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- Diameter: Largest value in the distance matrix. [Petitjean, M.; Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds; J. Chem. Inf. Comput. Sci. 32 (1992) 331-337.]
- GCUT_SLOGP_3: The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.
- b_rotR: Fraction of rotatable bonds: Number of rotatable bonds (A bond is rotatable if it has order 1, is not in a ring, and has at least two heavy neighbors) divided by number of bonds between heavy atoms
- AM_1_Eele: The electronic energy (kcal/mol) calculated using the AM1 Hamiltonian. [Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.]; 4

4.4.Descriptor selection:

Molecular operating environment (MOE) software was used to compute descriptors of each of the solvent molecules with MMFF94 force field and 0.01 kcal/Å gradient norm criterion.

After removing the descriptors containing only zero or constant values for all solvents, enhanced replacement method (ERM) on Multiple Linear Regression (MLR) from 1 descriptor up to 4 descriptors (Referring the rule of thumb of " five or six data points per descriptor) different models were computed.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

23/4

Descriptor: Chemical ratio :4:23 ~ 1:6

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 2)

$h^* = 0.52$

No outliers were detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

23 Carbon-based

List: SWNTs: Single-walled nanotubes

Shape: Fiber

Coating: NA

Size (nm): NA

Other info: The dispersibility of SWNTs in a series of 29 different organic solvents was extracted from Bergin et al. 2009.

6.6.Pre-processing of data before modelling:

A modified Kennard-Stone algorithm, where the response vector was replicated k (number of descriptors) times to enhance the influence of the response on the splitting results (Puzyn, T., Mostrag-Szlichtyng, A., Gajewicz, A., Skrzyński, M., & Worth, A. P. (2011). Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. Structural Chemistry, 22(4), 795–804. <http://doi.org/10.1007/s11224-011-9757-4>) was employed to split the data set into a training and a prediction set.

6.7.Statistics for goodness-of-fit:

$R^2_{cal} = 0.876$

SEC = 0.071

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

6 MCarbon-based

List

SWNTs: Single-walled nanotubes

Shape: Fiber

Coating: NA

Size(nm): NA

Other properties:

The dispersibility of SWNTs in a series of 29 different organic solvents was extracted from Bergin et al. 2009.

7.6. Experimental design of test set:

No information available

7.7. Predictivity - Statistics obtained by external validation:

$R^2_p = 0.924$

SEP = 0.078

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Compared the quality of the predictions of the log(Cmax) with three amidine solvents as an external test data, using model obtained.

1,1,3,3-tetramethylguan- idine (TMG), 1,5-diazabicyclo(4.3.0)non-5-ene (DBN) and 1,8-diazabicycloudec-7-ene (DBU) are consider as solvents which are incapable to disperse SWNTs, and correlated result were obtained after apply the built model.

Statistical value were provided in the table 2 in the publication which could be understood as cross-validation value but it was not specified in the paper, thus it was considered as not be included in the table.

CNTs: Carbon Nanotubes

SWNTs: Single walled Nanotubes

MLR: Multiple Linear Regression

ERM: Enhanced Replacement Method

R^2_{cal} : correlation coefficient for calibration set (training)

SEC: Standard Error of Calibration

R^2_p : correlationc oefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


NA, NA, QSPR, - Diameter: Largest value in the distance matrix. [Petitjean, M.; Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds; J. Chem. Inf. Comput. Sci. 32 (1992) 331-337.]

- GCUT_SLOGP_3: The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

- b_rotR: Fraction of rotatable bonds: Number of rotatable bonds (A bond is rotatable if it has order 1, is not in a ring, and has at least two heavy neighbors) divided by number of bonds between heavy atoms

- AM_1_Eele: The electronic energy (kcal/mol) calculated using the AM1 Hamiltonian. [Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.],MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting dispersion of SWNTs in different organic solvents by ANN
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting dispersion of SWNTs in different organic solvents by ANN

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

M. Salahinejad

salahinejad@gmail.com

2.6. Date of model development and/or publication:

2013

2.7. Reference(s) to main scientific papers and/or software package:

Salahinejad, M., & Zolfonoun, E. (2013). QSAR studies of the dispersion of SWNTs in different organic solvents. Journal of Nanoparticle Research, 15(11).

<http://doi.org/10.1007/s11051-013-2028-0>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Dispersion in organic solvents

3.3. Comment on endpoint:

Dispersion and solubilization are different phenomena but they are used interchangeably in the literature. Based on diameter and length of CNTs (Carbon Nanotubes) and their ability to form colloidal solutions, Geckeler and Premkumar recommended to use the dispersion term.

The dispersion of CNTs in different organic solvents was determined measuring the dispersion absorbance as concentration (C_max) of SWNTs (Single-Walled Nanotubes) after sonication and mild centrifugation. Detailed description of the method reported as reference in the publication (Giordani et al. 2006; Bergin et al. 2008)

The Log C_max was used as endpoint value in this model.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

ANN: Artificial Neural Network

4.3.Descriptors in the model:

- Diameter: Largest value in the distance matrix. [Petitjean, M.; Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds; J. Chem. Inf. Comput. Sci. 32 (1992) 331-337.]
- GCUT_SLOGP_3: The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.
- b_rotR: Fraction of rotatable bonds: Number of rotatable bonds (A bond is rotatable if it has order 1, is not in a ring, and has at least two heavy neighbors) divided by number of bonds between heavy atoms
- AM_1_Eele: The electronic energy (kcal/mol) calculated using the AM1 Hamiltonian. [Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.]; 4

4.4.Descriptor selection:

Molecular operating environment (MOE) software was used to compute descriptors of each of the solvent molecules with MMFF94 force field and 0.01 kcal/Å gradient norm criterion.

After removing the descriptors containing only zero or constant values for all solvents, enhanced replacement method (ERM) on Multiple Linear Regression (MLR) from 1 descriptor up to 4 descriptors (Referring the rule of thumb of " five or six data points per descriptor) different models were computed.

For each one of the generated models (different number of descriptors) the same descriptors selected in ERM-MLR were applied in the ANN models

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

23/4

Descriptor: Chemical ratio :4:23 ~ 1:6

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 2)

$h^* = 0.52$

No outliers were detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

23 Carbon-based

List: SWNTs: Single-walled nanotubes

Shape: Fiber

Coating: NA

Size (nm): NA

Other info: The dispersibility of SWNTs in a series of 29 different organic solvents was extracted from Bergin et al. 2009.

6.6.Pre-processing of data before modelling:

A modified Kennard-Stone algorithm, where the response vector was replicated k (number of descriptors) times to enhance the influence of the response on the splitting results (Puzyn, T., Mostrag-Szlichtyng, A., Gajewicz, A., Skrzyński, M., & Worth, A. P. (2011). Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. Structural Chemistry, 22(4), 795–804. <http://doi.org/10.1007/s11224-011-9757-4>) was employed to split the data set into a training and

a prediction set.

6.7. Statistics for goodness-of-fit:

$R^2_{cal} = 0.858$

SEC = 0.077

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

NA

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

6 MCarbon-based

List

SWNTs: Single-walled nanotubes

Shape: Fiber

Coating: NA

Size(nm): NA

Other properties:

The dispersibility of SWNTs in a series of 29 different organic solvents was extracted from Bergin et al. 2009.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2_p = 0.963$

SEP = 0.048

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

Compared the quality of the predictions of the log(C_{max}) with three amidine solvents as an external test data, using model obtained.

1,1,3,3-tetramethylguan- idine (TMG), 1,5-diazabicyclo(4.3.0)non-5-ene (DBN) and 1,8-diazabicycloudec-7-ene (DBU) are consider as solvents which are incapable to disperse SWNTs, and correlated result were obtained after apply the built model.

Statistical value were provided in the table 2 in the publication which could be understood as cross-validation value but it was not specified in the paper, thus it was considered as not be included in the table.

CNTs: Carbon Nanotubes

SWNTs: Single walled Nanotubes

ANN: Artificial Neural Network

ERM: Enhanced Replacement Method

R^2_{cal} : correlation coefficient for calibration set (training)

SEC: Standard Error of Calibration

R^2_p : correlation coefficient f

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


NA, NA, QSPR, - Diameter: Largest value in the distance matrix. [Petitjean, M.; Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds; J. Chem. Inf. Comput. Sci. 32 (1992) 331-337.]

- GCUT_SLOGP_3: The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

- b_rotR: Fraction of rotatable bonds: Number of rotatable bonds (A bond is rotatable if it has order 1, is not in a ring, and has at least two heavy neighbors) divided by number of bonds between heavy atoms

- AM_1_Eele: The electronic energy (kcal/mol) calculated using the AM1 Hamiltonian. [Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.], ANN: Artificial Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting dispersion of SWNTs in different organic solvents by PLS
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting dispersion of SWNTs in different organic solvents by PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Jahan B. Ghasemi

jahan.ghasemi@gmail.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Rofouei, M. K., Salahinejad, M., & Ghasemi, J. B. (2014). An alignment independent 3D-QSAR modeling of dispersibility of single-walled carbon nanotubes in different organic solvents. Fullerenes Nanotubes and Carbon Nanostructures, 22(7), 605–617.

<http://doi.org/10.1080/1536383X.2012.702157>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Dispersion in organic solvents

3.3. Comment on endpoint:

Dispersion and solubilization are different phenomena but they are used interchangeably in the literature. Based on diameter and length of CNTs (Carbon Nanotubes) and their ability to form colloidal solutions, Geckeler and Premkumar recommended to use the dispersion term. The dispersion of CNTs in different organic solvents was determined measuring the dispersion absorbance as concentration (C_{max}) of SWNTs (Single-Walled Nanotubes) after sonication and mild centrifugation. Detailed description of the method reported as reference in the publication (Giordani et al. 2006; Bergin et al. 2008). The Log C_{max} was used as endpoint value in this model.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

Generated molecular interaction fields (MIFs) for DRY, N1, O, and TIP probes defined as follows: the DRY probe represents hydrophobic interactions, N1 (amide) and O (carbonyl) probes represent hydrogen bond donor and acceptor groups, respectively, and the TIP probe represents a shape-field.

The five most relevant GRIND variables were:

- Variable No. 311, correlogram DRY-N1
- Variable No. 558, correlogram N1-TIP
- Variable No. 183, correlogram TIP-TIP
- Variable No. 373, correlogram DRY-TIP
- Variable No. 540, correlogram N1-TIP; 12

4.4.Descriptor selection:

GRIND-INdependent Descriptors (GRIND) was calculated automatically using the software Pentacle, version1.05 (Molecular Discovery Ltd., Oxford, UK). The Pentacle software uses alignment independent descriptors derived from GRID molecular interaction fields (MIF). In this study it was generated MIFs for DRY, N1, O, and TIP probes defined as follows: the DRY probe represents hydrophobic interactions, N1 (amide) and O (carbonyl) probes represent hydrogen bond donor and acceptor groups, respectively, and the TIP probe represents a shape-field. All molecular interaction fields were computed with the grid resolution of 0.5 Å with the smoothing window 0.8 Å. AMANDA algorithm were used for the extraction of nodes from the obtained MIF, the distance and relative position of nodes were described by MACC2

After removing those GRIND descriptors contained only zero values for all solvents, five variable selection techniques including: fractional factorial design (FFD), stepwise multiple linear regression, successive projection algorithm (SPA), genetic algorithm (GA), and enhanced replacement method (ERM). ERM was selected due it had the significantly better statistical values.

Also, the PLS model revealed the most relevant descriptor after model the system correctly.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

22/12

Descriptor: Chemical ratio :12:00

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

 $h^* = 0.54$

No outliers were detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

22 Carbon-based

List: SWNTs: Single-walled nanotubes

Shape: Fiber

Coating: NA

Size (nm): NA

Other info: The dispersibility of SWNTs in a series of 29 different organic solvents was extracted from Bergin et al. 2009.

6.6.Pre-processing of data before modelling:

A modified Kennard-Stone algorithm, where the response vector was replicated k (number of descriptors) times to enhance the influence of the response on the splitting results (Puzyn, T., Mostrag-Szlichtyng, A., Gajewicz, A., Skrzyński, M., & Worth, A. P. (2011). Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. Structural Chemistry, 22(4), 795–804. <http://doi.org/10.1007/s11224-011-9757-4>) was employed to split the data set into a training and a prediction set.

6.7.Statistics for goodness-of-fit:

$R^2_{cal} = 0.994$

RMSEC = 0.072

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2 = 0.936$

RMSECV = 0.238

Y-randomization (50 times):

- $R^2 = 0.255 \pm 0.112$

- $Q^2 = 0.042 \pm 0.048$

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

7 MCarbon-based

List

SWNTs: Single-walled nanotubes

Shape:Fiber

Coating:NA

Size(nm): NA

Other properties:

The dispersibility of SWNTs in a series of 29 different organic solvents was extracted from Bergin et al. 2009.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2_{Pred} = 0.980$

RMSEP = 0.072

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

NA

CNTs: Carbon Nanotubes

SWNTs: Single walled Nanotubes

PLS: Partial Least Squares

ERM: Enhanced Replacement Method

R^2_{cal} : correlation coefficient for calibration set (training)

RMSEC: Root Mean Square Error of Calibration

R^2_{Pred} : correlation coefficient

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, Generated molecular interaction fields (MIFs) for DRY, N1, O, and TIP probes defined as follows: the DRY probe represents hydrophobic interactions, N1 (amide) and O (carbonyl) probes represent hydrogen bond donor and acceptor groups, respectively, and the TIP probe represents a shape-field.

The five most relevant GRIND variables were:

- Variable No. 311, correlogram DRY-N1
- Variable No. 558, correlogram N1-TIP
- Variable No. 183, correlogram TIP-TIP
- Variable No. 373, correlogram DRY-TIP
- Variable No. 540, correlogram N1-TIP, PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predict efficiency of methylene blue (MB) adsorption onto CuO-NP-AC
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predict efficiency of methylene blue (MB) adsorption onto CuO-NP-AC by PCA-MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

M. Ghaedi

m_ghaedi@mail.yu.ac.ir

ghaedims@yahoo.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Ghaedi, M., Ghaedi, A. M., Hossainpour, M., Ansari, A., Habibi, M.

H., & Asghari, A. R. (2014). Least square-support vector (LS-

SVM) method for modeling of methylene blue dye adsorption

using copper oxide loaded on activated carbon: Kinetic and

isotherm s

<http://doi.org/10.1016/j.jiec.2013.08.011>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Adsorption / removal (%)

3.3.Comment on endpoint:

Methylene blue (MB) (chemical formula of $C_{27}H_{34}N_2O_4S$) adsorption was tested in CuO-NP-AC (Copper Oxide Nanoparticle loaded on Activated Carbon). The stock solution (100 mg L^{-1}) was prepared by dissolving 50.0 mg of MB in double distilled water and the test solutions were prepared by diluting stock solution to the desired concentrations daily. The pH was adjusted by addition of dilute HCl and/or KOH using pH/Ion meter model-686 and absorption studies were carried out using Jasco UV-Visible spectrophotometer model V-570. Chemicals with the highest purity available are purchased from Merck, Darmstadt, Germany. X-ray diffraction (XRD) patterns were recorded on a Bruker D8 advance X-ray diffractometer. Morphology and film thickness were measured by Philips XL-30 scanning electron microscopy.

The dye adsorption capacity of adsorbent were determined at the time intervals in the range of 1–35 min for 15 and 20 mg L^{-1} at room temperatures and it was found that equilibrium was established after 12 and 27 min for 15 and 20 mg L^{-1} .

The output is normalized between 0 and 1 to avoid numerical overflows due to very large or very small weights.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSPR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- pH
- Dye concentration
- Amount of NPs
- Time
- Amount of Carbon Active; 5

4.4.Descriptor selection:

Principal Components scores were used as inputs for the final model.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

32/5

Descriptor: Chemical ratio :5:32 ~ 1:6

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of CuO-NP-AC within the range of

parameters (descriptors) of the training set and for the specific dye, methylene blue (MB) .

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

32 Metal

List: CuO

Shape: Coarse porous surface with irregular pores

Coating: NA

Size (nm): 45

Other info: Isopropanol solvent and monoethanolamine (MEA) was used to dissolve $(\text{CH}_3\text{COO})_2\text{Cu}\cdot\text{H}_2\text{O}$. The solution was heated under magnetic stirring to 75 °C temperature for 1 h to form a homogeneous sol solution. When formed sol was obtained stable after 1 day aging at room temperature, the obtained stable sol was slowly heated under magnetic stirring up to 82 °C temperature until evaporate the solvent and form a hard homogeneous gel. After 1 day aging of gel at room temperature, the pyrolysis of the final gel was performed at temperature of 350, 450 and 550 °C for 2 h.

The SEM photograph was recorded by using Philips Netherland (Model-SEM-EDAX XL-30). The surface texture was found rough and heterogeneous porous in nature after treatment. CuO has considerable number of pores where is suitable for trapping and adsorption of dyes into these pores. XRD analyses as powerful tools was used to study the crystal structures of the

CuO nanoparticles and result is shown in Figure 4 in the publication, as synthesized and annealed at 350, 450 and 550 °C for 2 h. The XRD patterns show that annealing causes an increase in the intensities of the peaks at planes (1 1 1) and (2 0 0).

The FT-IR spectrum of present adsorbent show high intensity of OH vibrations and asymmetric and symmetric stretching vibrations correspond to CH₂ and CH₃.

The BET surface area measurement of AC prepared from was made by nitrogen adsorption at 196 °C using Sorptomatic 1990 (Thermo Fisher Scientific, USA). Before the measurement, the carbon sample was out gassed under a reduced atmosphere for 4 h at room temperature, 8 h at 110 °C and finally 12 h at 200 °C. Cumulative pore volume and area for mesopores were calculated using Barret–Joyner–Halenda method.

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

$R^2 = 0.90$

MSE = 0.0068

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2 = 0.85$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

9 M Metal

List

CuO

Shape: Coarse porous surface with irregular pores

Coating: NA

Size(nm): 45

Other properties:

Isopropanol solvent and monoethanolamine (MEA) was used to dissolve $(\text{CH}_3\text{COO})_2\text{Cu}\cdot\text{H}_2\text{O}$. The solution was heated under magnetic stirring to 75 °C temperature for 1 h to form a homogeneous sol solution. When formed sol was obtained stable after 1 day aging at room temperature, the obtained stable sol was slowly heated under magnetic stirring up to 82 °C temperature until evaporate the solvent and form a hard homogeneous gel. After 1 day aging of gel at room temperature, the pyrolysis of the final gel was performed at temperature of 350, 450 and 550 °C for 2 h.

The SEM photograph was recorded by using Philips Netherland (Model-SEM-EDAX XL-30). The surface texture was found rough and heterogeneous porous in nature after treatment. CuO has considerable number of pores where is suitable for trapping and adsorption of dyes into these pores. XRD analyses as powerful tools was used to study the crystal structures of the CuO nanoparticles and result is shown in Figure 4 in the publication, as synthesized and annealed at 350, 450 and 550 °C for 2 h. The XRD patterns show that annealing causes an increase in the intensities of the peaks at planes (1 1 1) and (2 0 0).

The FT-IR spectrum of present adsorbent show high intensity of OH vibrations and asymmetric and symmetric stretching vibrations correspond to CH₂ and CH₃.

The BET surface area measurement of AC prepared from was made by nitrogen adsorption at 196 °C using Sorptomatic 1990 (Thermo Fisher Scientific, USA). Before the measurement, the carbon sample was out gassed under a reduced atmosphere for 4 h at room temperature, 8 h at 110 °C and finally 12 h at 200 °C. Cumulative pore volume and area for mesopores were calculated using Barret–Joyner–Halenda method.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.88$

MSE = 0.0047

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

The results of examination of the time on experimental adsorption data and fitting the data to conventional kinetic model show the suitability of pseudo-second order and intraparticle diffusion model. Evaluation of the experimental equilibrium data by Langmuir, Tempkin, Freundlich and Dubinin Radushkevich (D-R) isotherm explore that Langmuir is superior to other model for fitting the experimental data in term of higher correlation coefficient and lower error analysis.

PCA: Principal Component Analysis

MLR: Multiple Linear Regression

CuO-NP-AC: Copper Oxide Nanoparticle load on Activated Carbon

MB: methylene blue

R^2 : correlation coefficient

Q^2 : cross-validated correlation coefficient

MSE: Mean Squared Error

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


NA, NA, QSPR, - pH

- Dye concentration

- Amount of NPs

- Time
- Amount of Carbon Active, MLR: Multiple Linear Regression

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predict efficiency of methylene blue (MB) adsorption onto CuO-NP-AC
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predict efficiency of methylene blue (MB) adsorption onto CuO-NP-AC by PCA-LSSVM

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

M. Ghaedi

m_ghaedi@mail.yu.ac.ir

ghaedims@yahoo.com

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Ghaedi, M., Ghaedi, A. M., Hossainpour, M., Ansari, A., Habibi, M. H., & Asghari, A. R. (2014). Least square-support vector (LS-SVM) method for modeling of methylene blue dye adsorption using copper oxide loaded on activated carbon: Kinetic and isotherm s

<http://doi.org/10.1016/j.jiec.2013.08.011>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Adsorption / removal (%)

3.3.Comment on endpoint:

Methylene blue (MB) (chemical formula of $C_{27}H_{34}N_2O_4S$) adsorption was tested in CuO-NP-AC (Copper Oxide Nanoparticle loaded on Activated Carbon). The stock solution (100 mg L^{-1}) was prepared by dissolving 50.0 mg of MB in double distilled water and the test solutions were prepared by diluting stock solution to the desired concentrations daily. The pH was adjusted by addition of dilute HCl and/or KOH using pH/Ion meter model-686 and absorption studies were carried out using Jasco UV-Visible spectrophotometer model V-570. Chemicals with the highest purity available are purchased from Merck, Darmstadt, Germany. X-ray diffraction (XRD) patterns were recorded on a Bruker D8 advance X-ray diffractometer. Morphology and film thickness were measured by Philips XL-30 scanning electron microscopy.

The dye adsorption capacity of adsorbent were determined at the time intervals in the range of 1–35 min for 15 and 20 mg L^{-1} at room temperatures and it was found that equilibrium was established after 12 and 27 min for 15 and 20 mg L^{-1} .

The output is normalized between 0 and 1 to avoid numerical overflows due to very large or very small weights.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSPR

4.2.Explicit algorithm:

LSSVM: Least-Squares Support Vector Machine

4.3.Descriptors in the model:

- pH
- Dye concentration
- Amount of NPs
- Time
- Amount of Carbon Active; 5

4.4.Descriptor selection:

Principal Components scores were used as inputs for the final model.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

32/5

Descriptor: Chemical ratio :5:32 ~ 1:6

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of CuO-NP-AC within the range of

parameters (descriptors) of the training set and for the specific dye, methylene blue (MB) .

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

32 Metal

List: CuO

Shape: Coarse porous surface with irregular pores

Coating: NA

Size (nm): 45

Other info: Isopropanol solvent and monoethanolamine (MEA) was used to dissolve $(\text{CH}_3\text{COO})_2\text{Cu}\cdot\text{H}_2\text{O}$. The solution was heated under magnetic stirring to 75 °C temperature for 1 h to form a homogeneous sol solution. When formed sol was obtained stable after 1 day aging at room temperature, the obtained stable sol was slowly heated under magnetic stirring up to 82 °C temperature until evaporate the solvent and form a hard homogeneous gel. After 1 day aging of gel at room temperature, the pyrolysis of the final gel was performed at temperature of 350, 450 and 550 °C for 2 h.

The SEM photograph was recorded by using Philips Netherland (Model-SEM-EDAX XL-30). The surface texture was found rough and heterogeneous porous in nature after treatment. CuO has considerable number of pores where is suitable for trapping and adsorption of dyes into these pores. XRD analyses as powerful tools was used to study the crystal structures of the

CuO nanoparticles and result is shown in Figure 4 in the publication, as synthesized and annealed at 350, 450 and 550 °C for 2 h. The XRD patterns show that annealing causes an increase in the intensities of the peaks at planes (1 1 1) and (2 0 0).

The FT-IR spectrum of present adsorbent show high intensity of OH vibrations and asymmetric and symmetric stretching vibrations correspond to CH₂ and CH₃.

The BET surface area measurement of AC prepared from was made by nitrogen adsorption at 196 °C using Sorptomatic 1990 (Thermo Fisher Scientific, USA). Before the measurement, the carbon sample was out gassed under a reduced atmosphere for 4 h at room temperature, 8 h at 110 °C and finally 12 h at 200 °C. Cumulative pore volume and area for mesopores were calculated using Barret–Joyner–Halenda method.

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

$R^2 = 0.97$

MSE = 0.0031

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2 = 0.87$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

9 M Metal

List

CuO

Shape: Coarse porous surface with irregular pores

Coating: NA

Size(nm): 45

Other properties:

Isopropanol solvent and monoethanolamine (MEA) was used to dissolve $(\text{CH}_3\text{COO})_2\text{Cu}\cdot\text{H}_2\text{O}$. The solution was heated under magnetic stirring to 75 °C temperature for 1 h to form a homogeneous sol solution. When formed sol was obtained stable after 1 day aging at room temperature, the obtained stable sol was slowly heated under magnetic stirring up to 82 °C temperature until evaporate the solvent and form a hard homogeneous gel. After 1 day aging of gel at room temperature, the pyrolysis of the final gel was performed at temperature of 350, 450 and 550 °C for 2 h.

The SEM photograph was recorded by using Philips Netherland (Model-SEM-EDAX XL-30). The surface texture was found rough and heterogeneous porous in nature after treatment. CuO has considerable number of pores where is suitable for trapping and adsorption of dyes into these pores. XRD analyses as powerful tools was used to study the crystal structures of the CuO nanoparticles and result is shown in Figure 4 in the publication, as synthesized and annealed at 350, 450 and 550 °C for 2 h. The XRD patterns show that annealing causes an increase in the intensities of the peaks at planes (1 1 1) and (2 0 0).

The FT-IR spectrum of present adsorbent show high intensity of OH vibrations and asymmetric and symmetric stretching vibrations correspond to CH₂ and CH₃.

The BET surface area measurement of AC prepared from was made by nitrogen adsorption at 196 °C using Sorptomatic 1990 (Thermo Fisher Scientific, USA). Before the measurement, the carbon sample was out gassed under a reduced atmosphere for 4 h at room temperature, 8 h at 110 °C and finally 12 h at 200 °C. Cumulative pore volume and area for mesopores were calculated using Barret–Joyner–Halenda method.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.92$

MSE = 0.0043

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

The results of examination of the time on experimental adsorption data and fitting the data to conventional kinetic model show the suitability of pseudo-second order and intraparticle diffusion model. Evaluation of the experimental equilibrium data by Langmuir, Tempkin, Freundlich and Dubinin Radushkevich (D-R) isotherm explore that Langmuir is superior to other model for fitting the experimental data in term of higher correlation coefficient and lower error analysis.

PCA: Principal Component Analysis

LS-SVM: Least Square Support Vector Machine

CuO-NP-AC: Copper Oxide Nanoparticle load on Activated Carbon

MB: methylene blue

R^2 : correlation coefficient

Q^2 : cross-validated correlation coefficient

MSE: Mean Square

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


NA, NA, QSPR, - pH

- Dye concentration

- Amount of NPs

- Time
- Amount of Carbon Active, LSSVM: Least-Squares Support Vector Machine

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins by PLSR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins by PLSR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Warren C. W. Chan

warren.chan@utoronto.ca

2.6. Date of model development and/or publication:

2014

2.7. Reference(s) to main scientific papers and/or software package:

Walkey, C. D., Olsen, J. B., Song, F., Liu, R., Guo, H., Olsen, D. W. H., ... Chan, W. C. W. (2014). Protein corona fingerprinting predicts the cellular interaction of gold and silver nanoparticles. ACS Nano, 8(3), 2439–2455.

<http://doi.org/10.1021/nn406018q>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log₂ transformed (Cell association [mL/μg(Mg)])

3.3. Comment on endpoint:

The ratio mcell/mwell divided by mcells quantified the total cell association, where mcell is the total

atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling.

Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

PLSR: Partial Least Square Regression

4.3.Descriptors in the model:

6 components obtained via Principal Component Analysis from the 64 selected protein corona fingerprints.

The list of abbreviations of selected proteins for the training set and the full name of those proteins can be obtained in the supplementary material of this publication (tables S7 and S5 respectively); 0

4.4.Descriptor selection:

Initial number of descriptors was reduced by the identification of the most predictive ones applying a sequential forward selection (SFS). Then PCA and 6 principal components were used to training the final model based on PLSR.

In Supplementary Material of the publication there are the Figures S13-14 which illustrate the evaluation of the Accuracy (R^2) and predictivity (Q^2_{LOO}) as a function of the number of principal components (PCs) and parameters (descriptors).

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

84/0

Descriptor: Chemical ratio :6:84 ~ 1 :14

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure S15 in Supplementary Material of publication)

$h^* = 0.25$

No outliers were detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

84 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid

11-Mercaptoundecanoic acid

(11-Mercaptoundecyl)tetra(ethylene glycol)

(11-Mercaptoundecyl)-N,N,N-trimethylammonium

Amino-poly(ethylene glycol)-thiol (5kDa)

Amino-poly(ethylene glycol)-thiol (5kDa) (low density)

2-Napthalenethiol @ deoxycholic acid

2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrenecomaleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Transmission electron microscopy (TEM) confirmed that the nanoparticle cores were monodisperse and had uniform morphology. After surface modification, it was used dynamic light scattering (DLS) to measure the hydrodynamic diameter (HD) of each formulation and absorbance spectrophotometry (AS) to measure the localized surface plasmon resonance index (LSPRi) and LSPR peak wavelength (LSPRpeak). The electrophoretic mobility and zeta potential (ZP) were characterized using light scattering and agarose gel electrophoresis.

The composition of the protein corona around each formulation was characterized qualitatively using poly(acrylamide) gel electrophoresis (PAGE) and semiquantitatively using high-resolution label-free shotgun tandem mass spectrometry (LC-MS/MS). The abundance of several key adsorbed serum proteins was further confirmed by western blotting.

Further experimental details allocated in the Material and Methods section.

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

The applied splitting was in order to perform Leave-One-Out and Leave-Many-Out cross-validation techniques.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.93$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{LOO} = 0.81$

$Q^2_{LMO25\%} = 0.61 \pm 0.18$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA MMetal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyldimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)

Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Transmission electron microscopy (TEM) confirmed that the nanoparticle cores were monodisperse and had uniform morphology. After surface modification, it was used dynamic light scattering (DLS) to measure the hydrodynamic diameter (HD) of each formulation and absorbance spectrophotometry (AS) to measure the localized surface plasmon resonance index (LSPRi) and LSPR peak wavelength (LSPRpeak). The electrophoretic mobility and zeta potential (ZP) were characterized using light scattering and agarose gel electrophoresis.

The composition of the protein corona around each formulation was characterized qualitatively using poly(acrylamide) gel electrophoresis (PAGE) and semiquantitatively using high-resolution label-free shotgun tandem mass spectrometry (LC-MS/MS). The abundance of several key adsorbed serum proteins was further confirmed by western blotting.

Further experimental details allocated in the Material and Methods section.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The relative abundance of different protein corona were used as

descriptors. Although they are not physicochemical descriptors it was concluded that protein corona encodes relevant biological information regarding cell association to the target NP.

In addition to protein corona fingerprint, physicochemical parameter were used to develop other models with the same procedure. Interestingly, a combined model that uses DLS, TEM, AS, and ZP along with the serum protein fingerprint

predicts cell association with a Q^2_{LOO} of 0.86, which is

only marginally more accurate than the model that uses the serum protein fingerprint alone, suggesting that the serum protein fingerprint encodes information about the size, surface charge, and aggregation state of the nanoparticles. Hence, only the fingerprint descriptor model was classified in this table.

A set of silver NPs was used to test the model for Au NP, but it indicated that a model derived using Au NPs cannot be applied to accurately predict cell association of Ag.

A mechanistic interpretation of the most relevant descriptors and general conclusion of the whole descriptors was developed in the publication.

NP: Nanoparticle

PLSR: Partial Least Square Regression

R^2 : Correlation coefficient

Q^2_{LOO} : Leave-one-out cross-validation correlation coefficient

$Q^2_{LMO25\%}$: Leave-many-out (25% of training data) cross-validation correlation coefficient

AD: Application

9.2. Bibliography:

NA

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, 6 components obtained via Principal Component Analysis from the 64 selected protein corona fingerprints.

The list of abbreviations of selected proteins for the training set and the full name of those proteins can be obtained in the supplementary material of this publication (tables S7 and S5 respectively), PLSR: Partial Least Square Regression

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting dispersion of SWNTs in different organic solvents by GA-
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting dispersion of SWNTs in different organic solvents by GA-MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Hayriye Yilmaz

hayriyey@erciyes.edu.tr

hayriyey@9cnanotox.org

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Yilmaz, H., Rasulev, B., & Leszczynski, J. (2015). Modeling the dispersibility of single walled carbon nanotubes in organic solvents by quantitative structure-activity relationship approach.

Nanomaterials, 5(2), 778–791.

<http://doi.org/10.3390/nano5020778>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Dispersion in organic solvents

3.3. Comment on endpoint:

Dispersion and solubilization are different phenomena but they are used interchangeably in the literature. Based on diameter and length of CNTs (Carbon Nanotubes) and their ability to form colloidal solutions, Geckeler and Premkumar recommended to use the dispersion term. The dispersion of CNTs in different organic solvents was determined measuring the dispersion absorbance as concentration (C_max) of SWNTs (Single-Walled Nanotubes) after sonication and mild centrifugation. Detailed description of the method reported as reference in the publication (Giordani et al. 2006; Bergin et al. 2008)
The Log C_max was used as endpoint value in this model.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

GA-MLR: Genetic Algorithm and Multivariate Linear Regression

Using BuildQSAR software

4.3.Descriptors in the model:

- SRW09: self-returning walk count of the 9th order and represents the surroundings of odd membered rings (five-membered in our case).
- ATS6m: atomic mass weighted term.
- X0Av: the connectivity index descriptor, which is an average valence connectivity index chi-0.
- Dipole Z: indicates the strength and orientation behavior of a molecule in an electrostatic field. Both the magnitude and the components (X, Y, Z) of the dipole moment are calculated. The descriptor is estimated by utilizing partial atomic charges and atomic coordinates.; 4

4.4.Descriptor selection:

The semiempirical quantum chemical descriptors were calculated by the RM1 method implemented in HyperChem. An initial set of 258 DRAGON software generated theoretical descriptors was selected from the entire set of generated descriptors and used to describe the chemical diversity of the compounds.

In addition, the density functional theory (DFT) with the hybrid meta exchange-correlation functional M06-2X/6-311G(d,p) calculations were applied to obtain another set of quantum-chemically generated physico-chemical parameters of studied SWCNTs solvents-including dipole moments (total dipole moment, X, Y, and Z components); orbital energies, E_HOMO, E_LUMO and heats of formation. All DFT calculations were performed using the Gaussian 09 software.

The selection of descriptors were done within the model generation by Genetic Algorithm technique, for a different number of descriptors. The plot of correlation coefficient of training set and test set against the number of variables of each model (see publication's Figure 2) reveals 4 as the optimal number of descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

22/4

Descriptor: Chemical ratio :4:22 ~ 1:5

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

Not specified in the paper.

Expected an applicability domain of CNTs in organic solvents within the range of parameters (descriptors) of the training set solvents.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

22 Carbon-based

List: SWNTs: Single-walled nanotubesShape: FiberCoating: NASize (nm): NA

Other info: The dispersibility of SWNTs in a series of 29 different organic solvents was extracted from Bergin et al. 2009.

The initial structures of investigated organic solvents for SWCNTs were built using HyperChem

7.5 package. The structures of compounds were firstly preoptimized with the Molecular Mechanics Force Field (MM+) procedure included in the HyperChem.

6.6.Pre-processing of data before modelling:

A modified Kennard-Stone algorithm, where the response vector was replicated k (number of descriptors) times to enhance the influence of the response on the splitting results (Puzyn, T., Mostrag-Szlichtyng, A., Gajewicz, A., Skrzyński, M., & Worth, A. P. (2011). Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. *Structural Chemistry*, 22(4), 795–804. <http://doi.org/10.1007/s11224-011-9757-4>) was employed to split the data set into a training and a prediction set.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.797$$

$$s = 0.456$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$$Q^2 = 0.666$$

$$SDEP = 0.527$$

$$SPRESS = 0.585$$

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

7 MCarbon-based

List

SWNTs: Single-walled nanotubes

Shape:Fiber

Coating:NA

Size(nm): NA

Other properties:

The dispersibility of SWNTs in a series of 29 different organic solvents was extracted from Bergin et al. 2009.

The initial structures of investigated organic solvents for SWCNTs were built using HyperChem

7.5 package. The structures of compounds were firstly preoptimized with the Molecular Mechanics Force Field (MM+) procedure included in the HyperChem.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.807$

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

r_m^2 statistical values presented in the publication, which were in doubt to be taken into account due this statistical values have not yet established in the whole community.

CNTs: Carbon Nanotubes

SWNTs: Single walled Nanotubes

GA-MLR: Genetic Algorithm Multiple Linear Regression

R^2 : correlation coefficient

Q^2 : leave-one-out cross-validation correlation coefficient

LOO: Leave-one-out cross-validation

SDEP: Standard De

9.2.Bibliography:

NA

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

NA, NA, QSPR, - SRW09: self-returning walk count of the 9th order and represents the surroundings of odd membered rings (five-membered in our case).


- ATS6m: atomic mass weighted term.

- X0Av: the connectivity index descriptor, which is an average valence connectivity index chi-0.

- Dipole Z: indicates the strength and orientation behavior of a molecule in an electrostatic field. Both the magnitude and the components (X, Y, Z) of the dipole moment are calculated. The descriptor is estimated by utilizing partial atomic charges and atomic coordinates.,GA-MLR: Genetic Algorithm and Multivariate Linear Regression

Using BuildQSAR software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on physicochemical properties of protein
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on physicochemical properties of protein corona by MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Robert Rallo

robert.rallo@urv.cat

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Kamath, P., Fernández, A., Giralt, F., & Rallo, R. (2015).

Predicting cell association of surface-modified nanoparticles using protein corona structure - activity relationships (PCSAR). Current Topics in Medicinal Chemistry, 15(18), 175–182.

<http://doi.org/10.2174/1568026615666150506152808>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3. Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands)

with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio mcell/mwell divided by mcells quantified the total cell association, where mcell is the total atomic gold (or silver) content associated with cells, mwell is the total atomic gold (or silver) content in well (associated with cells and free in solution) and mcells is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO2 atmin 175 cm2 tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression
using R statistical software.

4.3.Descriptors in the model:

- Probability of expression in inclusion bodies
- Tiny amino acids percentage
- Basic amino acids percentage
- Aspartic acid DayhoffStat
- Molecular weight
- Polar amino acids percentage
- Acidic amino acids percentage; 7

4.4.Descriptor selection:

For protein corona, 129 out of 785 initial data set proteins were quantifiable for developing the model. Based on physicochemical properties, protein properties were computed using the EMBOSS Pepstats program from their amino acid sequences.Redundant descriptors (i.e., highly correlated) were removed from the Pepstats output, and a total of 35 non-redundant descriptors were finally selected for prediction of net cell association. The final descriptors were computed averaging the physicochemical properties of the proteins weighted by the relative abundance of the corresponding protein.

Finally, the normalized physicochemical descriptors were used to form a new AP-based fingerprint.

The selection of the best set of features (i.e., averaged protein descriptors) for the final model was based on the adjusted correlation coefficient, which measures if the addition of a new descriptor increases the explanatory power of the resulting model.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

84/7

Descriptor: Chemical ratio :7:84 ~ 1:12

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

Expected an applicability domain of surface modified Au NPs within the range of physicochemical parameters (descriptors) of the protein corona fingerprints training set .

Due the use of physicochemical descriptors weighted by the relative abundance results, the model can be used to predict cell association of nanoparticles with protein coronas containing proteins different than those used for training the model. The only information needed by the model is the spectral counts and the primary sequence of the protein.

After test the model with a data set of Ag NPs it was concluded that in those models based on fingerprints the nanoparticle core is a key factor that determines the structure and composition of the protein corona. Hence, the model will be only able to be applied onto Au NPs with (not specific) protein corona.

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

84 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid

11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Transmission electron microscopy (TEM) confirmed that the nanoparticle cores were monodisperse and had uniform morphology. After surface modification, it was used dynamic light scattering (DLS) to measure the hydrodynamic diameter (HD) of each formulation and absorbance spectrophotometry (AS) to measure the localized surface plasmon resonance index (LSPRi) and LSPR peak wavelength (LSPRpeak). The electrophoretic mobility and zeta potential (ZP) were characterized using light scattering and agarose gel electrophoresis.

The composition of the protein corona around each formulation was characterized qualitatively using poly(acrylamide) gel electrophoresis (PAGE) and semiquantitatively using high-resolution label-free shotgun tandem mass spectrometry (LC-MS/MS). The abundance of several key adsorbed serum proteins was further confirmed by western blotting.

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

Leave-One-Out and Leave-Many-Out cross-validations were applied to evaluate the performance of the model.

The applied splitting was in order to perform a k-fold cross-validation test (k=10)

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.80$$

$$R^2_{E632} = 0.77 \pm 0.07$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$$R^2_{LOO} = 0.76$$

$$R^2_{LMO25\%} = 0.72 \pm 0.11$$

10-fold cross-validation:

$$R^2 = 0.77 \pm 0.14$$

Y-randomization:

$$R^2 = 0.16$$

7.External validation - OECD Principle 4**7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

NA Metal

List

Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid

11-Mercaptoundecanoic acid

(11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Transmission electron microscopy (TEM) confirmed that the nanoparticle cores were monodisperse and had uniform morphology. After surface modification, it was used dynamic light scattering (DLS) to measure the hydrodynamic diameter (HD) of each formulation and absorbance spectrophotometry (AS) to measure the localized surface plasmon resonance index (LSPR_i) and LSPR peak wavelength (LSPR_{peak}). The electrophoretic mobility and zeta potential (ZP) were characterized using light scattering and agarose gel electrophoresis.

The composition of the protein corona around each formulation was characterized qualitatively using poly(acrylamide) gel electrophoresis (PAGE) and semiquantitatively using high-resolution label-free shotgun tandem mass

spectrometry (LC-MS/MS). The abundance of several key adsorbed serum proteins was further confirmed by western blotting.

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Ag NPs data set was used to evaluate the predictivity of the model in other NPs. It was concluded that the model and the interaction is core specific. In order to check if the methodology could be applied to generate a model for different core nanoparticles, the Ag NPs set was used, and relative good results were obtained despite of the small size of the data set.

Not external validation test was applied, hence we cannot say that the obtained results will be totally reliable.

Close related with previous already reported publications in this table (C. D. Walkey et al., 2014 and Liu, R. et al., 2015) with a key difference in the way of how were treated the protein corona descriptors. The two fingerprint techniques have been compared by calculating the normalized mutual information index between the partitions obtained after clustering the nanoparticles represented in terms of each fingerprint approach. Clustering results indicate that the information conveyed by both fingerprints is essentially the same. In addition, the modelling approach proposed here for gold and silver nanoparticles outperforms models based only on relative abundances in terms of applicability, size and stability.

There is a mechanistic interpretation for the final descriptors applied

in the model.

NP: Nanoparticle

MLR: Multiple Linear Regression

R^2 : correlation coefficient

R^2_{LOO} : Leave-one-out corss-validation correlation coefficient

$R^2_{LMO25\%}$: Leave-many-out (25% of training data) cross-validation correlation coefficient

R^2_{E632} : The 0.6

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - Probability of expression in inclusion bodies

- Tiny amino acids percentage
- Basic amino acids percentage
- Aspartic acid DayhoffStat
- Molecular weight
- Polar amino acids percentage
- Acidic amino acids percentage,MLR: Multiple Linear Regression using R statistical software.

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 - C70 solubility in chlorobenzene by GA-MLR
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 - C70 solubility in chlorobenzene by GA-MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Eslam Pourbasheer

pourbasheer@ut.ac.ir

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Pourbasheer, E., Aalizadeh, R., Ardabili, J. S., & Ganjali, M. R. (2015). QSPR study on solubility of some fullerenes derivatives using the genetic algorithms - Multiple linear regression. Journal of Molecular Liquids, 204, 162–169.

<http://doi.org/10.1016/j.molliq.2015.01.028>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in chlorobenzene

3.3. Comment on endpoint:

NA

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSPR

4.2.Explicit algorithm:

GA-MLR: Genetic Algorithm and Multivariate Linear Regression

4.3.Descriptors in the model:

- D1: Max nucleoph. react. index for a C atom.
- D2: RNCS Relative negative charged SA (SAMNEG*RNCG) [Quantum-Chemical PC].
- D3: Totmolecular 2-center resonance energy.
- D4: Highest normalmode vib frequency.; 4

4.4.Descriptor selection:

The MOPAC output files are introduced to CODESSA program to calculate several classes of the descriptors.

Descriptors with constant or near constant values were detected and then eliminated. Also, pairs of variables with a correlation coefficient greater than 0.9 were classified as inter-correlated, and only one of them, which is presented higher correlation values with solubility, was considered in developing the model. A total of 97 descriptors were considered for further investigations.

The genetic algorithm (GA) variable subset selection method was used for the selection of the most relevant descriptors. Here the fitness function is the correlation coefficient of leave-one-out cross-validation (Q^2_{LOO}). The GA program was written in Matlab 6.5.

In addition, variance inflation factor (VIF) and the value of mean effect (MF) were computed to check the final descriptors.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

21/4

Descriptor: Chemical ratio :4:21 ~ 1:5

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

$h^* = 0.714$

No outliers were detected.

Also, Euclidean based AD was performed. A compound (molecule 16)

belonging to the test set was detected as an outlier

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

21 Carbon-based

List: Methanofullerenes

(Fullerenes C60 and C70)

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: Molecular structure of the molecules are given in Table 1 of the publication.

All of the molecules were drawn into a HyperChem 7.5 package and preoptimized using the MM+ molecular mechanics force field. Then, a more precise optimization is done with the semiempirical AM1 method in MOPAC. The molecular structures are optimized using the Polak–Ribiere algorithm until the root mean square gradient reaches 0.01 kcal mol⁻¹.

6.6.Pre-processing of data before modelling:

In order to demonstrate the structural diversity of this data set, it was performed a hierarchical cluster analyses (CA) of these chemicals. According to the results of hierarchical cluster analyses, all the data were divided into a training set of 21 compounds to develop the model and a test set of 6

compounds to evaluate the model based on two rules:

- 1) the range of the solubility values of both the training set and the test set should be covered from the lowest to the highest;
- 2) the linking distances corresponding to the training set in the dendrogram, should not be out of the main clusters.

6.7. Statistics for goodness-of-fit:

$R^2 = 0.801$

RMSE = 0.198

CCC = 0.858

MAE = 0.128

6.8. Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:

$Q^2_{\text{LOO}} = 0.716$

$Q_{\text{BOOT}} = 0.674$

Y-scrambling (500 times):

no correlation between descriptors and solubility. See publication's Fig. 5

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: NA

NP surface chemistry: NA

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

6 MCarbon-based

List

Methanofullerenes

(Fullerenes C60 and C70)

Shape:SphericalCoating:NASize(nm): NAOther properties:

Molecular structure of the molecules are given in Table 1 of the publication.

All of the molecules were drawn into a HyperChem 7.5 package and preoptimized using the MM+ molecular mechanics force field. Then, a more precise optimization is done with the semiempirical AM1 method in MOPAC. The molecular structures are optimized using the Polak–Ribiere algorithm until the root mean square gradient reaches 0.01 kcal mol⁻¹.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:
 $R^2 = 0.792$
 $CCC = 0.756$
 $RMSE = 0.277$
 $MAE = 0.218$
7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

In addition to the reported model another model with different splitting compounds from the hierarchical clustering was performed (it was known as Test set 2, while our reported model was the Test set 1). Also from the Genetic Algorithm procedure, different sets of descriptors for the Test set 1 were evaluated and the one which had the best performing was selected as the main model.

Q^2_{LGO} : Leave-group-out cross-validation correlation coefficient was computed but the size of group was not reported, hence the value was omitted in this report.

Also the modified r^2 ($(r_m)^2$) was computed, after a revision of whole reported publications, it will be discussed if it will be reported or not.

A widely explanation of Mechanistic Intepretation for each of the final descriptors was developed in the publication.

NP: Nanoparticle

GA-MLR: Genetic Algorithm Multiple Linear Regression

R^2 : correlation coefficient

Q^2_{LOO} : Leave-one-out corss-validation correlation coefficient

Q_{BOOT} : correlation coefficient for bootstraping procedure

RMS: Root Mean Square Error

9.2.Bibliography:

Troshin PA, Hoppe H, Renz J et al (2009) Material solubility- photovoltaic performance relationship in the design of novel fullerene derivatives for bulk heterojunction solar cells. Adv Funct Mater 19:779–788. doi:10.1002/adfm.200801189

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:


NA, NA, QSPR, - D1: Max nucleoph. react. index for a C atom.

- D2: RNCS Relative negative charged SA (SAMNEG*RNCG) [Quantum-Chemical PC].

- D3: Totmolecular 2-center resonance energy.

- D4: Highest normalmode vib frequency.,GA-MLR: Genetic Algorithm and Multivariate Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Consensus model to predict Zeta potential of metal oxide nanoparticle
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Consensus model to predict Zeta potential of metal oxide nanoparticle by GA-MLR

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Tomasz Puzyn

t.puzyn@qsar.eu.org

2.6. Date of model development and/or publication:

2015

2.7. Reference(s) to main scientific papers and/or software package:

Mikolajczyk, A., Gajewicz, A., Rasulev, B., Schaeublin, N., Maurer-Gardner, E., Hussain, S., ... Puzyn, T. (2015). Zeta potential for metal oxide nanoparticles: A predictive model developed by a nano-quantitative structure-property relationship approach. Ch

<http://doi.org/10.1021/cm504406a>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Zeta potential (ζ)

3.3. Comment on endpoint:

The agglomeration phenomenon largely influences toxicity of nanoparticles. Such properties of nanoparticles that describe their behavior are known as extrinsic properties. The ease of formation of agglomerates strongly depends on the surface charge that stabilizes dispersed nanoparticles and prevents them from agglomeration. However, the available experimental techniques are unable to measure surface charge directly; its value can only be estimated by measuring zeta potential (ζ) in a given medium.

Dynamic light scattering (DLS) was used zeta potential (ζ) in cell culture media (serum-free) was done using a Malvern Instruments zeta-sizer Nano-ZS instrument.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

Consensus GA-MLR: Genetic Algorithm Multiple Linear Regression
using QSARINS software

4.3.Descriptors in the model:

- ψ : spherical size
- $\epsilon_{\text{HOMO/nMe}}$: the weighted energy of the highest occupied molecular orbital; 2

4.4.Descriptor selection:

11 descriptors has been derived from images obtained with Hitachi H-7600 TEM capable of 0.35 nm point-to-point resolution. Each image has been converted to numerical format, by converting pixels to certain values. In the 8-bit monochrome image (called gray scale image), each pixel has been assigned a value from 0 to 255. The assigned values depend on the image gray levels.

17 theoretical quantum-mechanical descriptors were calculated at the semiempirical level of theory with the use of PM6 method implemented in MOPAC2009 software.

The quantum-mechanical calculations included two steps: (i) optimization of the cluster's geometry with respect to the decreasing energy gradient and (ii) calculation of the descriptors on the basis of the optimized geometry

The final descriptors were obtained by Genetic Algorithm implemented in the QSARINS software.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

10/2

Descriptor: Chemical ratio :2:15 ~ 1:8

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 4)

$h^* = 0.6$

No outliers were detected

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

10 Metal Oxide

List:

Al₂O₃

Bi₂O₃

CoO

Fe₂O₃

In₂O₃

La₂O₃

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: Spherical

Coating: NA

Size (nm): 15-210

Other info: To verify morphology and size, one drop of a 100 µg/mL solution (RPMI-1640 media, at pH 7.5 (ATCC, Manassas, VA) supplemented with 10% (v/v) fetal bovine serum (FBS, ATCC) and 1% (w/v) penicillin/streptomycin (Sigma, St. Louis, MO) was spotted on a forever/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Hitachi H-7600.

TEM (Schaumburg, IL) at 120 kV. Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ζ) in cell culture media (serum-free) was done using a Malvern Instruments zeta-sizer Nano-ZS instrument according to the procedure described by Murdock et al.

All necessary crystallographic data have been collected from Cambridge Crystallographic Data Center (CCDC)

6.6.Pre-processing of data before modelling:

The data related to the 15 MeOx-NPs were split into two sets: a training set (to be used to develop a nano-QSPR model) and a validation set (to be used only for validating the model's predictive ability). To perform a splitting, the nanoparticles were sorted along with the increasing values of zeta potential. Then, every third NP was included in the validation set (V), whereas the remaining NPs formed the training set (T). Because models developed on the basis of very small data sets might not be robust enough, they have carried out multiple splittings to investigate the potential influence of the splitting procedure on the modeling results. Because two metal oxides (TiO₂ and In₂O₃) were characterized by the same values of zeta potential (−9.6 mV), two additional combinations of the prepared splits were carried out (publication's Table 2)

6.7.Statistics for goodness-of-fit:

$R^2 = 0.82$

CCC = 0.96

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

NA

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

5 M Metal Oxide

List

Al₂O₃

Bi₂O₃

CoO

Fe₂O₃

In₂O₃

La₂O₃

Sb₂O₃

SiO₂

SnO₂

TiO₂

V₂O₃

WO₃

Y₂O₃

ZnO

ZrO₂

Shape: Spherical

Coating: NA

Size(nm): 15-210

Other properties:

To verify morphology and size, one drop of a 100 µg/mL solution (RPMI-1640 media, at pH 7.5 (ATCC, Manassas, VA) supplemented with 10% (v/v) fetal bovine serum (FBS, ATCC) and 1% (w/v) penicillin/streptomycin (Sigma, St. Louis, MO) was spotted on a forever/carbon-coated TEM grid (EMS Diasum, Hatfield, PA) and allowed to dry. Once dried, the nanoparticles were viewed using a Hitachi H-7600.

TEM (Schaumburg, IL) at 120 kV. Dynamic light scattering (DLS) for characterization of nanoparticle size and zeta potential (ζ) in cell culture media (serum-free) was done using a Malvern Instruments zeta-sizer Nano-ZS

instrument according to the procedure described by Murdock et al.

All necessary crystallographic data have been collected from Cambridge Crystallographic Data Center (CCDC)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$Q^2_{EXT} = 0.87$

RMSEP = 1.25

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Consensus nano-QSPR model based on seven different splits of data was performed. The idea to perform a consensus model it was to face the small size of data set.

The goodness-of-fit and prediction ability of the consensus model were estimated on a single training and test set (for split 2, where both of NPs , TiO₂ and In₂O₃, are in training data set, which also had the best individual performance).

For individual models (each splitting) cross-validation leave-one-out was computed.

A good Mechanistic Interpretation is presented.

NP: Nanoparticle

GA-MLR: Genetic Algorithm Multiple Linear Regression

R^2 : correlation coefficient

Q^2_{EXT} : correlation coefficient for external validation data set

CCC: Concordance Correlation Coefficient

RMSEP: Root Mean Square Error of Prediction

9.2.Bibliography:

(already reported in this table)

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., & Leszczynski, J. (2015). Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), 313–325.

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC


10.3.Keywords:

NA, NA, QSPR, - ψ : spherical size

- $\epsilon_{\text{HOMO/nMe}}$: the weighted energy of the highest occupied molecular orbital,Consensus GA-MLR:
Genetic Algorithm Multiple Linear Regression

using QSARINS software

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: Predicting C60 - C70 solubility in chlorobenzene by PLS
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

Predicting C60 - C70 solubility in chlorobenzene by PLS

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Natalia Sizochenko

sizochenko@icnanotox.org

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Sizochenko, N., Kuz'min, V., Ognichenko, L., & Leszczynski, J. (2016). Introduction of simplex-informational descriptors for QSPR analysis of fullerene derivatives. Journal of Mathematical Chemistry, 54(3), 698–706.

<http://doi.org/10.1007/s10910-015-0581-8>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

NA

NA

3.2. Endpoint:

Solubility in chlorobenzene

3.3. Comment on endpoint:

Original solubility (mg/ml) was converted to molar (mmol/ml) and expressed as log(S) values.

Molecular structures, solubility and data transformation are summarized in publication's Table 1.

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSPR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- S1: Atomic weight
- S2 and S3: Partial charges
- S4: Lipophilicity
- S5 and S6: Polarization; 6

4.4.Descriptor selection:

Simplex approach and the informational field theory were simultaneously applied describing structure features of fullerene derivatives. Fullerene's molecular graphs were differentiated using informational potentials of the influence of near and far surroundings. Due to this fact the set of descriptors becomes more diverse.

Simplex-informational descriptors were generated by using HiT QSAR software.

The descriptors that highly correlated with each other were eliminated. When the squared correlation coefficient between descriptors in a pair was higher than a given limit (set here as 0.85), one of variables was deleted. The descriptors having higher sum of squared correlation coefficients calculated in relation to all other descriptors were excluded. In addition, descriptors with no or with very little variance were also eliminated.

The final 6 descriptors were combined into 3 latent variables within the model building, PLS.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

22/6

Descriptor: Chemical ratio :6:22 ~ 1:4

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Not specified in the paper. Expected applicability domain of nanomaterials within the range of experimental solubility and final descriptors of the training set.

One of the fullerenes was not soluble, then, one insoluble compound was replaced by soluble fullerene C60 (M.T. Beck, G.Mándi, Fuller. Sci.

Tech. 5, 291 (1997))

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

22 Carbon-based

List: Methanofullerenes

(Fullerenes C60 and C70)

Shape: Spherical

Coating: NA

Size (nm): NA

Other info: Chemical structures were first pre-optimized with the Molecular Mechanics Force Field (MM+), and the resulting geometries were further refined by means of the semiempirical PM7method.

6.6.Pre-processing of data before modelling:

The initial dataset was splitted into training and a test sets based on random selection considering two rules:

(a) the range of the response values of both the training set and the test set should be covered from the lowest to the highest;

(b) the highest and lowest response values were included in the training set.

Thus, initial dataset was splitted into 22 compounds for training set and 5 compounds for test set.

6.7.Statistics for goodness-of-fit: $R^2 = 0.939$

RMSE = 0.120

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods: $Q^2 = 0.904$

RMSE = 0.141

Y-Scrambling:

 $R^2 = 0.026$ $Q^2 = 0.031$ **7.External validation - OECD Principle 4****7.1.Availability of the external validation set:**

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:NA

NP surface chemistry: NA

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

5 MCarbon-based

List

Methanofullerenes

(Fullerenes C60 and C70)

Shape:SphericalCoating:NASize(nm): NA

Other properties:

Chemical structures were first pre-optimized with the Molecular Mechanics Force Field (MM+), and the resulting geometries were further refined by means of the semiempirical PM7 method.

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.873$

RMSE = 0.146

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5**8.1.Mechanistic basis of the model:**

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information**9.1.Comments:**

AD was verified with leverage approach and Williams plot, but the resulting plot was not reported neither the leverage threshold nor about any detected outlier from data set.

A Mechanistic Interpretation of the final descriptors is presented.

PLS: Partial least squares

R^2 : correlation coefficient

Q^2 : cross-validation correlation coefficient

RMSE: Root Mean Square Error

9.2.Bibliography:

Troshin PA, Hoppe H, Renz J et al (2009) Material solubility- photovoltaic performance relationship in the design of novel fullerene derivatives for bulk heterojunction solar cells. Adv Funct Mater 19:779–788. doi:10.1002/adfm.200801189

10.Summary (JRC QSAR Model Database)**10.1.QMRF number:**

To be entered by JRC


10.2.Publication date:

To be entered by JRC

10.3.Keywords:

- NA, NA, QSPR, - S1: Atomic weight
- S2 and S3: Partial charges
- S4: Lipophilicity
- S5 and S6: Polarization, PLS: Partial Least Squares

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (MLR case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

MLR: Multiple Linear Regression

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

$$h^* = 0.350$$

This analysis confirmed the large applicability of the model, which covers 96% of the AD calculated for the 84 Au-NPs, with the exclusion of G60 Phe-SH (ID 80), G15 DDT@SDS (ID 14) and G15 HDA (ID 18).

This result is in agreement with results from PCA.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.87$

RMSE = 0.81

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods: $Q^2_{loo} = 0.85$ RMSE_{CV} = 0.89**7. External validation - OECD Principle 4****7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

24 M Metal

List

Au

Shape: NACoating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Naphthalenethiol @ deoxycholic acid
 2-Naphthalenethiol @ Pluronic F-127
 2-Naphthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ poly(vinyl alcohol)
 Octadecylamine

Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.76$

RMSE = 1.07

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-validation

RMSE: Root Mean Square Error

AD: Ap

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B, MLR: Multiple Linear Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (k-NN case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio mcell/mwell divided by mcells quantified the total cell association, where mcell is the total atomic gold (or silver) content associated with cells, mwell is the total atomic gold (or silver) content in well (associated with cells and free in solution) and mcells is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO2 atmin 175 cm2 tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

kNN: k-Nearest Neighbour

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol
 Thiolated L-alanine
 Thiolated L-asparagine
 11-Amino-1-undecanethiol
 Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)

Thiolated poly(ethyleneimine)

L-Phenylalanine

Thiolated L-phenylalanine

Thiolated poly(L-lysine)

Poly(vinyl alcohol)

Poly(vinylpyrrolidone)

Stearic acid

Thiolated L-serine

Bis(p-sulfonatophenyl)phenylphosphine

TWEEN20

Thiolated L-threonine

N-(2-Mercaptopropionyl)glycine

Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.88$$

$$RMSE = 0.81$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{\text{loo}} = 0.79$

RMSE_{CV} = 1.17

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

24 MMetal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)

Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.73$

RMSE = 1.12

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

kNN: k-Nearest Neighbour

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-validation

RMSE: Roo

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:


To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B,kNN: k-Nearest Neighbour

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (GRegNN case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio mcell/mwell divided by mcells quantified the total cell association, where mcell is the total atomic gold (or silver) content associated with cells, mwell is the total atomic gold (or silver) content in well (associated with cells and free in solution) and mcells is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO2 atmin 175 cm2 tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

GRegNN: General Regression Neural Network

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol
 Thiolated L-alanine
 Thiolated L-asparagine
 11-Amino-1-undecanethiol
 Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)

Thiolated poly(ethyleneimine)

L-Phenylalanine

Thiolated L-phenylalanine

Thiolated poly(L-lysine)

Poly(vinyl alcohol)

Poly(vinylpyrrolidone)

Stearic acid

Thiolated L-serine

Bis(p-sulfonatophenyl)phenylphosphine

TWEEN20

Thiolated L-threonine

N-(2-Mercaptopropionyl)glycine

Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.93$$

$$RMSE = 0.63$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{\text{loo}} = 0.75$

RMSE_{CV} = 1.18

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

24 Metal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)

Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.74$

RMSE = 1.09

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

GRegNN: General Regression Neural Network

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-valid

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B, GRegNN: General Regression Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (RBFNN case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RBFNN: Radial Basis Function Neural Networks

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol
 Thiolated L-alanine
 Thiolated L-asparagine
 11-Amino-1-undecanethiol
 Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)

Thiolated poly(ethyleneimine)

L-Phenylalanine

Thiolated L-phenylalanine

Thiolated poly(L-lysine)

Poly(vinyl alcohol)

Poly(vinylpyrrolidone)

Stearic acid

Thiolated L-serine

Bis(p-sulfonatophenyl)phenylphosphine

TWEEN20

Thiolated L-threonine

N-(2-Mercaptopropionyl)glycine

Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.87$$

$$RMSE = 0.82$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{\text{loo}} = 0.86$

RMSE_{CV} = 0.86

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

24 Metal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)

Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.74$

RMSE = 1.10

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

RBFNN: Radial Basis Function Neural Networks

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-validation

9.2. Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B, RBFNN: Radial Basis Function Neural Networks

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (CPANN case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio $\frac{m_{cell}}{m_{well}}$ divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

CPANN: Counter Propagation Neural Network

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol
 Thiolated L-alanine
 Thiolated L-asparagine
 11-Amino-1-undecanethiol
 Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)

Thiolated poly(ethyleneimine)

L-Phenylalanine

Thiolated L-phenylalanine

Thiolated poly(L-lysine)

Poly(vinyl alcohol)

Poly(vinylpyrrolidone)

Stearic acid

Thiolated L-serine

Bis(p-sulfonatophenyl)phenylphosphine

TWEEN20

Thiolated L-threonine

N-(2-Mercaptopropionyl)glycine

Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.92$$

$$RMSE = 0.66$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{\text{loo}} = 0.72$

RMSE_{CV} = 1.25

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

24 Metal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)

Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.82$

RMSE = 0.96

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

CPANN: Counter Propagation Neural Network

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-valid

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B, CPANN: Counter Propagation Neural Network

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (SVM-radial case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio mcell/mwell divided by mcells quantified the total cell association, where mcell is the total atomic gold (or silver) content associated with cells, mwell is the total atomic gold (or silver) content in well (associated with cells and free in solution) and mcells is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO2 atmin 175 cm2 tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM-radial : Support Vector Machine radial

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

60 Metal

List: AuShape: NACoating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol
 Thiolated L-alanine
 Thiolated L-asparagine
 11-Amino-1-undecanethiol
 Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)

Thiolated poly(ethyleneimine)

L-Phenylalanine

Thiolated L-phenylalanine

Thiolated poly(L-lysine)

Poly(vinyl alcohol)

Poly(vinylpyrrolidone)

Stearic acid

Thiolated L-serine

Bis(p-sulfonatophenyl)phenylphosphine

TWEEN20

Thiolated L-threonine

N-(2-Mercaptopropionyl)glycine

Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.94$$

$$RMSE = 0.59$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{\text{loo}} = 0.82$

RMSE_{CV} = 0.99

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

24 Metal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)

Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.76$

RMSE = 1.09

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

SVM-radial : Support Vector Machine radial

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-validated

9.2. Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B, SVM-radial : Support Vector Machine radial

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (SVM-linear case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

SVM-linear: Support Vector Machine linear

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol
 Thiolated L-alanine
 Thiolated L-asparagine
 11-Amino-1-undecanethiol
 Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)

Thiolated poly(ethyleneimine)

L-Phenylalanine

Thiolated L-phenylalanine

Thiolated poly(L-lysine)

Poly(vinyl alcohol)

Poly(vinylpyrrolidone)

Stearic acid

Thiolated L-serine

Bis(p-sulfonatophenyl)phenylphosphine

TWEEN20

Thiolated L-threonine

N-(2-Mercaptopropionyl)glycine

Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.87$$

$$RMSE = 0.82$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{\text{loo}} = 0.85$

RMSE_{CV} = 0.88

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

24 MMetal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)

Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.78$

RMSE = 1.04

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

SVM-linear: Support Vector Machine linear

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-valid

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B,SVM-linear: Support Vector Machine linear

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (PLS case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

PLS: Partial Least Squares

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4. Data for the dependent variable for the training set:

Yes

6.5. Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol
 Thiolated L-alanine
 Thiolated L-asparagine
 11-Amino-1-undecanethiol
 Peptide sequence 'CALNN'
 Peptide sequence 'CFGAILS'
 Citrate
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)
 (low density)
 Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127

2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)

2-Napthalenethiol @ poly(vinyl alcohol)

Octadecylamine

Thiolated poly(allylamine)

Thiolated amino-poly(ethylene glycol) (3kDa)

Thiolated poly(ethyleneimine)

L-Phenylalanine

Thiolated L-phenylalanine

Thiolated poly(L-lysine)

Poly(vinyl alcohol)

Poly(vinylpyrrolidone)

Stearic acid

Thiolated L-serine

Bis(p-sulfonatophenyl)phenylphosphine

TWEEN20

Thiolated L-threonine

N-(2-Mercaptopropionyl)glycine

Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$$R^2 = 0.87$$

$$RMSE = 0.81$$

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

$Q^2_{\text{loo}} = 0.85$

RMSE_{CV} = 0.84

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NA

7.2.Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size:Yes

NP surface chemistry: Yes

7.3.Data for each descriptor variable for the external validation set:

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

24 Metal

List

Au

Shape:NA

Coating:N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)

Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.76$

RMSE = 1.07

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

PLS: Partial Least Squares

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-validation

RMSE: R_o

9.2. Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:


To be entered by JRC

10.3. Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B, PLS: Partial Least Squares

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (PPR case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

PPR: Projection Pursuit Regression

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

$h^* = 0.350$

This analysis confirmed the large applicability of the model, which covers 96% of the AD calculated for the 84 Au-NPs, with the exclusion of G60 Phe-SH (ID 80), G15 DDT@SDS (ID 14) and G15 HDA (ID 18).

This result is in agreement with results from PCA.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.91$

RMSE = 0.69

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods: $Q^2_{loo} = 0.81$ RMSE_{CV} = 1.01**7. External validation - OECD Principle 4****7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

24 M Metal

List

Au

Shape: NACoating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Naphthalenethiol @ deoxycholic acid
 2-Naphthalenethiol @ Pluronic F-127
 2-Naphthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ poly(vinyl alcohol)
 Octadecylamine

Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.79$

RMSE = 1.01

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

PPR: Projection Pursuit Regression

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-validation

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B, PPR: Projection Pursuit Regression

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (EARTH case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

EARTH: Multivariate adaptive regression splines (MARS or EARTH)

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

$$h^* = 0.350$$

This analysis confirmed the large applicability of the model, which covers 96% of the AD calculated for the 84 Au-NPs, with the exclusion of G60 Phe-SH (ID 80), G15 DDT@SDS (ID 14) and G15 HDA (ID 18).

This result is in agreement with results from PCA.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.90$

RMSE = 0.73

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods: $Q^2_{\text{loo}} = 0.77$ RMSE_{CV} = 1.10**7. External validation - OECD Principle 4****7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

24 M Metal

List

Au

Shape: NACoating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine

Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.80$

RMSE = 0.97

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

EARTH: Multivariate adaptive regression splines (MARS or EARTH)

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

9.2. Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B, EARTH: Multivariate adaptive regression splines (MARS or EARTH)

10.4. Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (RF-6 case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RF-6 : Random Forest using 6 descriptors

4.3.Descriptors in the model:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 6

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/6

Descriptor: Chemical ratio :6:60 ~ 1:10

5. Defining the applicability domain - OECD Principle 3**5.1. Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

AD was verified with leverage approach and Williams plot. (For specific details see the publication's Figure 3)

$$h^* = 0.350$$

This analysis confirmed the large applicability of the model, which covers 96% of the AD calculated for the 84 Au-NPs, with the exclusion of G60 Phe-SH (ID 80), G15 DDT@SDS (ID 14) and G15 HDA (ID 18).

This result is in agreement with results from PCA.

5.2. Method used to assess the applicability domain:

Not applicable

5.3. Software name and version for applicability domain assessment:

Not applicable

5.4. Limits of applicability:

No information available

6. Internal validation - OECD Principle 4**6.1. Availability of the training set:**

Yes

6.2. Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3. Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

60 Metal

List: Au

Shape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.95$

RMSE = 0.62

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9. Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10. Robustness - Statistics obtained by Y-scrambling:

No information available

6.11. Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods: $Q^2_{loo} = 0.71$ RMSE_{CV} = 1.29**7. External validation - OECD Principle 4****7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

24 M Metal

List

Au

Shape: NACoating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide
 Peptide sequence 'CVVIT'
 1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide
 1-Dodecanethiol @hexadecyltrimethylammonium bromide
 1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane
 1-Dodecanethiol @ hexadecylamine
 1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Naphthalenethiol @ deoxycholic acid
 2-Naphthalenethiol @ Pluronic F-127
 2-Naphthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Naphthalenethiol @ poly(vinyl alcohol)
 Octadecylamine

Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.80$

RMSE = 1.01

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

RF-6 : Random Forest using 6 descriptors

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-validation

9.2.Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, - zav_serum: hydrodynamic diameter measured after exposure to serum

- A1AT: Alpha-1-antitrypsin


- KNG1: Kininogen-1

- GFAP: Glial fibrillary acidic protein

- VTNC: Vitronectin

- CO4B: Complement C4-B, RF-6 : Random Forest using 6 descriptors

10.4.Comments:

	QMRF identifier (JRC Inventory): To be entered by JRC
	QMRF Title: NP-cell association based on corona proteins and physicochemical by
	Printing Date: 30/03/2017

1. QSAR identifier

1.1. QSAR identifier (title):

NP-cell association based on corona proteins and physicochemical by machine learning approaches (RF-150 case)

1.2. Other related models:

NA

1.3. Software coding the model:

NA

2. General information

2.1. Date of QMRF:

30/03/2017

2.2. QMRF author(s) and contact details:

LEITAT

2.3. Date of QMRF update(s):

2.4. QMRF update(s):

2.5. Model developer(s) and contact details:

Ester Papa

ester.papa@uninsubria.it

2.6. Date of model development and/or publication:

2016

2.7. Reference(s) to main scientific papers and/or software package:

Papa, E., Doucet, J. P., Sangion, A., & Doucet-Panaye, A. (2016). Investigation of the influence of protein corona composition on gold nanoparticle bioactivity using machine learning approaches. SAR and QSAR in Environmental Research. QSAR Research Unit

i

<http://doi.org/10.1080/1062936X.2016.1197310>

2.8. Availability of information about the model:

No information available

2.9. Availability of another QMRF for exactly the same model:

No information available

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Cell

A549 human lung epithelial carcinoma cells

3.2. Endpoint:

In vitro - log2 transformed (Cell association [mL/μg(Mg)])

3.3.Comment on endpoint:

The correlation of cell association of Au NPs (modified with different ionic/cationic surface ligands) with corona proteins and physicochemical properties was investigated via QSAR analysis of a recently published dataset (C. D. Walkey, et al. et al., 2015 already reported in this table). The ratio m_{cell}/m_{well} divided by m_{cells} quantified the total cell association, where m_{cell} is the total atomic gold (or silver) content associated with cells, m_{well} is the total atomic gold (or silver) content in well (associated with cells and free in solution) and m_{cells} is the total mass of magnesium per sample. The related net cell association data were log transformed (log2 transformation) prior to modelling. Cell association was chosen as a model biological interaction because of its relevance to inflammatory responses, biodistribution, and toxicity in vivo.

A549 human lung epithelial carcinoma cells (ATCC) were maintained in RPMI1640 (Wisent, cat#: 350-000-CL) supplement with 10%(v/v) fetal bovine serum (FBS) (Gibco, cat#: 12483-020) and 1% (v/v) penicillin-streptomycin (P-S) (Gibco, cat#: 15140-122) in a sterile 5% CO₂ atmin 175 cm² tissue culture flasks (NEST, cat#: 709003)

3.4.Endpoint units:

See 3.2

3.5.Dependent variable:

See 3.2

3.6.Experimental protocol:

No information available

3.7.Endpoint data quality and variability:

No information available

4.Defining the algorithm - OECD Principle 2**4.1.Type of model:**

QSAR

4.2.Explicit algorithm:

RF-150 : Random Forest using 6 descriptors

4.3.Descriptors in the model:

Most relevant descriptors:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B; 150

4.4.Descriptor selection:

785 proteins detected in the protein corona and 21 properties extracted from the physico-chemical characterization of the NPs were used to generate the initial set of descriptors. Among the 785 proteins measured in the corona, once the proteins measured in only one NP were excluded, a further selection excluding all the proteins expressed in fewer than 10 NPs was performed. Proteins were then sorted according to decreasing number of sum of spectral counts (over all the NPs) and the first 129 were selected as the final fingerprint for the further development of the models.

Using genetic algorithm (GA) optimized for MLR models based on ordinary least squares (MLR-OLS) and by support vector machines (SVMs) applied for selection of variables. This double selection was performed in order to evaluate possible differences in the results of procedures optimized for linear (i.e. MLR-OLS) and ML (SVM) approaches.

4.5.Algorithm and descriptor generation:

No information available

4.6.Software name and version for descriptor generation:

No information available

4.7.Chemicals/Descriptors ratio:

60/150

Descriptor: Chemical ratio :150:60

5.Defining the applicability domain - OECD Principle 3**5.1.Description of the applicability domain of the model:**

The analysis of the distribution of the 84 Au-NPs by PCA in publication's Figure 2, and in supplementary material of publication Figure S3(a–c), shows that training and prediction set objects are distributed quite homogeneously in the descriptors space. However, four NPs in the prediction set were detected as slightly isolated points in at least two out of the three scatterplots, i.e. G15 DDT@DOTAP (ID 11), G15 DDT@SDS (ID 14), G60 HDA (ID 74) and G60 Phe-SH (ID 80)

5.2.Method used to assess the applicability domain:

Not applicable

5.3.Software name and version for applicability domain assessment:

Not applicable

5.4.Limits of applicability:

No information available

6.Internal validation - OECD Principle 4**6.1.Availability of the training set:**

Yes

6.2.Available information for the training set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Nanoparticle: Yes

Part extended for NPs.

NP composition: Yes

NP size: Yes

NP surface chemistry: Yes

6.3.Data for each descriptor variable for the training set:

Yes

6.4.Data for the dependent variable for the training set:

Yes

6.5.Other information about the training set:

60 Metal

List: AuShape: NA

Coating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine

1-Dodecanethiol @ stearic acid

1-Dodecanethiol @ sodium dodecyl sulfate

5,5'-Dithiobis(2-nitrobenzoic acid)

Pluronic F-127

Thiolated L-glycine

Hexadecylamine

α -Lipoic acid

Mercaptoacetic acid

4-Mercaptobenzoic acid

2-Mercaptoethanesulfonate

Thiolated L-methionine

6-Mercaptohexanoic acid

16-Mercaptohexadecanoic acid

3-Mercaptopropionic acid

Methoxy-poly(ethylene glycol)-thiol (1kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa)

Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)

Methoxy-poly(ethylene glycol)-thiol (2kDa)

Methoxy-poly(ethylene glycol)-thiol (5kDa)

Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*

Mercaptosuccinic acid

11-Mercaptoundecanoic acid

(11-Mercaptoundecyl)tetra(ethylene glycol)

(11-Mercaptoundecyl)-N,N,N-trimethylammonium

Amino-poly(ethylene glycol)-thiol (5kDa)

Amino-poly(ethylene glycol)-thiol (5kDa) (low density)

2-Napthalenethiol @ deoxycholic acid

2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrenecomaleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)
 Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size (nm): 15, 30, and 60

Other info: Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

6.6.Pre-processing of data before modelling:

From the initial 105 Au modified NP, those ones (21) with neutral ligands were dropped due to their negligible adsorption of serum proteins.

6.7.Statistics for goodness-of-fit:

$R^2 = 0.95$

RMSE = 0.63

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

No information available

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12. Robustness - Statistics obtained by other methods:
 $Q^2_{\text{loo}} = 0.64$
 $\text{RMSE}_{\text{CV}} = 1.43$
7. External validation - OECD Principle 4**7.1. Availability of the external validation set:**

NA

7.2. Available information for the external validation set:

CAS: No

Chemical Name: not applicable

SMILES: not applicable

Formula: not applicable

INChI: not applicable

MOL file: not applicable

Part extended for NPs.

NP composition: NA

NP size: Yes

NP surface chemistry: Yes

7.3. Data for each descriptor variable for the external validation set:

NA

7.4. Data for the dependent variable for the external validation set:

NA

7.5. Other information about the external validation set:

24 Metal

List

Au

Shape: NACoating: N-Acetyl-L-cysteine

6-Amino-1-hexanethiol

Thiolated L-alanine

Thiolated L-asparagine

11-Amino-1-undecanethiol

Peptide sequence 'CALNN'

Peptide sequence 'CFGAILS'

Citrate

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

Carboxymethyl-poly(ethylene glycol)-thiol (5kDa)

(low density)

Hexadecyltrimethylammonium bromide

Peptide sequence 'CVVIT'

1-Dodecanethiol @benzyltrimethylhexadecylammonium bromide

1-Dodecanethiol @hexadecyltrimethylammonium bromide

1-Dodecanethiol @ 1,2-dioleoyl-3-trimethylammonium-propane

1-Dodecanethiol @ hexadecylamine

1-Dodecanethiol @ octadecylamine
 1-Dodecanethiol @ stearic acid
 1-Dodecanethiol @ sodium dodecyl sulfate
 5,5'-Dithiobis(2-nitrobenzoic acid)
 Pluronic F-127
 Thiolated L-glycine
 Hexadecylamine
 α -Lipoic acid
 Mercaptoacetic acid
 4-Mercaptobenzoic acid
 2-Mercaptoethanesulfonate
 Thiolated L-methionine
 6-Mercaptohexanoic acid
 16-Mercaptohexadecanoic acid
 3-Mercaptopropionic acid
 Methoxy-poly(ethylene glycol)-thiol (1kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa)
 Methoxy-poly(ethylene glycol)-thiol (20kDa) (low density)
 Methoxy-poly(ethylene glycol)-thiol (2kDa)
 Methoxy-poly(ethylene glycol)-thiol (5kDa)
 Thiolated amino-poly(ethylene glycol) (methoxyterminated)(5kDa)*
 Mercaptosuccinic acid
 11-Mercaptoundecanoic acid
 (11-Mercaptoundecyl)tetra(ethylene glycol)
 (11-Mercaptoundecyl)-N,N,N-trimethylammonium
 Amino-poly(ethylene glycol)-thiol (5kDa)
 Amino-poly(ethylene glycol)-thiol (5kDa) (low density)
 2-Napthalenethiol @ deoxycholic acid
 2-Napthalenethiol @ Pluronic F-127
 2-Napthalenethiol @ (4'-aminoacetophenone)-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ aminopropanol-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethanolamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ ethylenediamine-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ urea-modified poly(styrene-co-maleic anhydride)
 2-Napthalenethiol @ poly(vinyl alcohol)
 Octadecylamine
 Thiolated poly(allylamine)
 Thiolated amino-poly(ethylene glycol) (3kDa)
 Thiolated poly(ethyleneimine)
 L-Phenylalanine
 Thiolated L-phenylalanine
 Thiolated poly(L-lysine)

Poly(vinyl alcohol)
 Poly(vinylpyrrolidone)
 Stearic acid
 Thiolated L-serine
 Bis(p-sulfonatophenyl)phenylphosphine
 TWEEN20
 Thiolated L-threonine
 N-(2-Mercaptopropionyl)glycine
 Thiolated L-tryptophan

Size(nm): 15, 30, and 60

Other properties:

Further experimental details allocated in the Material and Methods section from source publication (Walkey et al., 2014)

7.6.Experimental design of test set:

No information available

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.80$

RMSE = 1.06

7.8.Predictivity - Assessment of the external validation set:

No information available

7.9.Comments on the external validation of the model:

No information available

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

No information available

8.2.A priori or a posteriori mechanistic interpretation:

No information available

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

Multiple machine learning approaches compared in this publication, brief explanation of the different techniques are present.

Mechanistic Interpretation of results was provided.

NP:Nanoparticle

GA: Genetic Algorithm

MLR: Multiple Linear Regression

OLS: Ordinary Least Squares

RF-150 : Random Forest using 6 descriptors

R^2 : correlation coefficient

Q^2_{loo} : leave-one-out cross-validation correlation coefficient

CV: Cross-validation

9.2. Bibliography:

(already reported in this table)

C. D. Walkey, et al., Protein corona fingerprinting predicts the cell association of gold nanoparticles, ACS Nano, 2014, 8, 2439–2455

10. Summary (JRC QSAR Model Database)

10.1. QMRF number:

To be entered by JRC

10.2. Publication date:

To be entered by JRC

10.3. Keywords:

Cell, A549 human lung epithelial carcinoma cells, QSAR, Most relevant descriptors:

- zav_serum: hydrodynamic diameter measured after exposure to serum
- A1AT: Alpha-1-antitrypsin
- KNG1: Kininogen-1
- GFAP: Glial fibrillary acidic protein
- VTNC: Vitronectin
- CO4B: Complement C4-B, RF-150 : Random Forest using 6 descriptors

10.4. Comments: