EUROPEAN COMMISSION
JOINT RESEARCH CENTRE
Institute for Health and Consumer Protection
Systems Toxicology Unit
EU Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Summary Record

Joint PARERE-ESTAF Meeting

20th-21st October 2015, Ispra, Italy

The joint PARERE-ESTAF meeting started in the afternoon of 20th October 2015 and concluded at lunchtime on 21st October 2015 (agenda included in Annex 1). All presentations and documents are available on CIRCABC.

1. EURL ECVAM updates - general

EURL ECVAM presented an overview of some of the recent developments with regard to coordinating/promoting the development, validation and use of alternative methods; facilitating exchange of information on their development; managing dissemination databases/information systems; promoting dialogue between legislators, regulators and stakeholders. Recent activities include a bilateral meeting with IVTIP; hosting the first meeting of European 3Rs Centres to explore common interests; a visit of a Brazilian government delegation to discuss joint actions for cooperation on alternative methods, and EURL ECVAM's contribution to the Commission's Communication (led by DG ENV) in response to the European Citizens' Initiative "Stop Vivisection" .

EURL ECVAM detailed its significant contribution to the OECD AOP development program, especially its active participation in (and co-chairing of, on behalf of the EU/EC) the Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST) and its role in building the Adverse Outcome Pathways Knowledge Base (AOP-KB). The major research initiative SEURAT-1, in its 5th and final year, will hold a final symposium in December 2015. Brief details were also provided on the ECHA Scientific Workshop to be held in April 2016, and on the new Horizon 2020 project EU-ToxRisk, with which JRC / EURL ECVAM will formally cooperate. Additionally, the EURL ECVAM Search Guide has been translated into Korean and there are also plans for Portuguese and Chinese versions, extending the international reach of this very useful tool that facilitates the reliable and systematic retrieval of information on alternative methods that could be used to avoid animal testing.

EURL ECVAM updates - Method validation workflow and acceptance of methods

- EURL ECVAM provided an overview of its validation workflow, with an emphasis on the coordination of validation, stakeholder engagement, regulatory acceptance and international recognition and dissemination. The EURL ECVAM Status Report was published on 14 October 2015 and provides, amongst other topics, full details of all test submissions assessed during 2014-2015; the EURL ECVAM strategies; updates of the REACH annexes (VII and VIII) and guidance as well as details on OECD projects led, or co-led by EURL ECVAM.

- In the area of skin sensitisation, EURL ECVAM described its considerable contributions to the delivery of *in vitro* method OECD Test Guidelines and to the development of the OECD conceptual framework and reporting standards for Integrated Approaches to Testing and Assessment (IATA).
- A description of the EU-NETVAL AR-CALUX validation study provided an overview of the progress made to-date. The study is nearing the completion of the transfer phase which involved three EU-NETVAL laboratories. The outcome of this study will be fed into the OECD project (led by EURL ECVAM) to establish a Performance Based Test Guideline (PBTG) for Androgen Receptor Transactivation Assays (ARTA) and associated performance standards.
- EURL ECVAM is currently developing a guidance document on validation of *in vitro* methods which is intended to serve as a practical resource primarily for test developers interested in undertaking a validation study to demonstrate the reliability and/or relevance of their method.
- Two EURL ECVAM Recommendations have been published since the last meeting and there are two upcoming Recommendations: Reconstructed human Cornea-like Epithelium (RhCE) assays and Human-based CYP induction assays.
- EURL ECVAM has continued to increase its engagement and communication with regulators, industry and academia via its official networks, PARERE, ESTAF, ESAC, ICATM and through its collaborative network of validation laboratories, EU-NETVAL. Key dissemination activities, such as DB-ALM, have also helped to facilitate the use of non-animal methods in a variety of contexts.

2. European Citizens' Initiative on "Stop Vivisection" and relevant actions put forward by the Commission Communication

DG ENV presented a summary of the Commission's response to the ECI. The main message is that whilst full replacement is the ultimate goal, a ban on *in vivo* testing is premature. It is therefore necessary to use the best tools available and to combine them in line with the 3Rs principle. Highlighting the four actions, DG ENV concluded that the Directive 2010/63/EU is currently the most advanced in the world for animal protection. However, the ECI challenges the Commission to find new ways to speed up the paradigm shift and is an incentive to encourage all to work together on all fronts. There is a need for better collaboration and dissemination of information which involves all key actors in the area. EURL ECVAM has a key role in Action 1 (improving knowledge sharing in the 3Rs) and a public survey is due to be launched in order to understand how knowledge related to the 3Rs is distributed and shared. All ESTAF and PARERE members will be invited to participate in the survey. The organisers of this ECI have requested a scientific debate about the use of animals for scientific purposes, which will take place in the form of a conference at the end of 2016.

The way in which the Directive is put into legislation in individual MS varies and it is important for competent authorities to understand that it needs to be enforced. Improving the information on availability and acceptance of alternative approaches is part of this enforcement. DG ENV has produced guidance documents to promote uniform implementation and application of the Directive.

3. 3Rs in Biomedical Research

Cruelty Free International presented the highlights of WC9 which included: Organ-on-a-chip, 3D *in vitro* models, alternative methods for batch testing of biologicals, and the redundancy in animal testing (e.g. relating to drug-testing in a second species).

Discussion

Looking across the different sectors, there is a difference in approach to using alternative methods. There needs to be more engagement with pharmaceutical companies to discuss how they use *in vitro* models and to bring their screening models forward for validation. Academics tend to go for the animal model: it is necessary to increase their confidence in the *in vitro* model. In order to demonstrate the quality of a method/model to the end-users, it is essential to develop performance standards.

The priorities which were identified at the recent 3Rs Centres meeting in April include improving communication about all aspects of the 3Rs. Suggestions for this include, amongst others, sharing intellectual resources and establishing a dedicated communication network of contact points throughout Europe.

It is important to encourage more researchers to attend meetings on alternative methods but also to get more publications on alternative methods in higher impact journals.

EURL ECVAM collaborated in the B-DEBATE monthly meeting in October on future tools for biomedical research. This produced a positive discussion, with some emphasis on the role of computational and *in vitro* tools in reduction as well as replacement of animal testing.

4. Integrated approaches to testing and assessment (IATA): method validation; testing strategy development, evaluation and acceptance.

EURL ECVAM presented four case studies (serious eye damage/eye Irritation, skin sensitisation, endocrine disruption and acute systemic toxicity) to demonstrate the concept of an IATA in relation to each area and to prepare the groups for the workshop. Please see the workshop summary record (Annex 2) for the details of this workshop and the subsequent feedback from each break-out group, discussions and conclusions.

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE
Institute for Health and Consumer Protection
Systems Toxicology Unit
EU Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Annex 1

Joint PARERE-ESTAF Meeting

20th-21st October 2015

Agenda

Tuesday 20th October

Item	Time	Description	Format
	14:00	Welcome	
1	14:00-14:30	EURL ECVAM updates - general	Presentation Discussion
	14:30-15:00	EURL ECVAM updates - Method validation workflow and acceptance of methods	Presentation Discussion
2	15:00-15:30	European Citizens' Initiative on "Stop Vivisection" and relevant actions in the Commission Communication	Presentation Discussion
3	15:30-16:00	3Rs in Biomedical Research	Presentation Discussion
	16:00-16:30	*Coffee break*	
4	16:30-18:30	Integrated approaches to testing and assessment (IATA): method validation; testing strategy development, evaluation and acceptance. Introduction to the workshop Presentations: • Serious Eye Damage/Eye Irritation • Skin Sensitisation • Endocrine Disruption • Acute Systemic Toxicity	Presentations Discussion
	18.30	Meeting closure	

Wednesday 21st October

Item	Time	Description	Format
5	8:30-9:00	Working instructions for the workshop and establishment of break-out groups	
6	9:00-11:00	Break-out group discussion addressing the following questions: 1) In cases where test methods are component parts of an IATA should the IATA be established prior to formal validation of the methods, or vice versa? 2) Which regulatory instruments and procedures offer the best means of ensuring broad acceptance and use of IATA?	Break-out group discussion
	11:00-11.30	*Coffee break*	
7	11:30-13:00	Replies/conclusions/recommendations of break-out groups	Presentations Discussion
8	13:00-13:30	Wrap-up of main conclusions/recommendations	Discussion Agreement
	13:30	Closure of meeting	Discussion

Annex 2

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE
Institute for Health and Consumer Protection
Systems Toxicology Unit
EU Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Summary Record

PARERE-ESTAF Workshop on Integrated Approaches to Assessment and Testing (IATA) - method validation, testing strategy development, evaluation and acceptance.

21st October 2015, Ispra, Italy

The aim of the workshop was to discuss key issues relating to the development, evaluation, acceptance and use of Integrated Approaches to Testing and Assessment (IATA). Four case studies were presented to give an overview of different kinds of IATA that are currently considered by regulatory bodies or are under development. These IATA are designed for the assessment of eye irritation (1), skin sensitisation (2a, 2b), endocrine disruption (3, 4) and acute systemic toxicity (5). Workshop participants discussed a series of questions, with a view to identifying the pros and cons of different options for promoting the acceptance of IATA.

The orientation paper for the workshop, which contains all of the background references, is included in Annex I. The four case studies are briefly outlined here.

Serious Eye Damage/Eye Irritation

The current REACH Guidance on Information Requirements and Chemical Safety Assessment, (Endpoint specific guidance (Chapter R.7a), Section R.7.2, serious eye damage/eye irritation) includes a general IATA. This provides guidance on how to fulfil REACH information requirements using different types of information, including data generated with alternative methods. The OECD Guidance Document on an IATA for serious eye damage/eye irritation is being developed under EC and US lead. In both cases, the IATA consists of three parts: Part One – Existing and non-testing data; Part Two – Weight-of-evidence judgement on all collected information; Part Three – Testing data. The information sources which may be considered for the WoE assessment within the REACH IATA for serious eye damage/eye irritation are: Skin corrosion data; Physico-chemical data; Existing human data; *In silico* data ((Q)SAR, RA, Grouping); *In vitro* data; *In vivo* data. In the case of serious eye damage/eye irritation, the IATA offers a conceptual framework for the gathering and assessment of endpoint-relevant information. It helps increasing the efficiency and reducing the cost of the assessment by informing how to make best use of existing data to come to a conclusion or to guide the generation of new testing data, where required.

The currently (OECD) accepted *in vitro* methods can be used individually as stand-alone methods under certain conditions for C&L (i.e., to identify UN GHS No Cat. and/or Cat. 1). None of the available individual *in vitro* methods can however fully replace the regulatory Draize rabbit eye test (OECD TG 405). It is estimated that for at least 70% of the substances, one single *in vitro* test method

will be sufficient to derive a final conclusion. For 20-30% of the substances a combination of 2 or more methods within IATA will be needed. The use of non-guideline methods which are able to detect the persistence of effects may be required to achieve full replacement.

Skin Sensitisation

The biology and chemical mechanisms of skin sensitisation are well understood and have enabled the development of a variety of non-animal methods (*in silico, in chemico, in vitro*) for skin sensitisation assessment. Toxtree and the QSAR Toolbox are examples of some of the computational tools which may be used in combination with other methods to identify and measure reactivity and to characterise a number of events in the skin sensitisation AOP. Several validated *in vitro* methods are also available. A meta-analysis of the 12 case studies (predictive models / structured approaches) for skin sensitisation show that they perform in the same range. Information generated by alternative approaches which have been developed to address specific Key Events (KE) in the skin sensitisation pathway can be incorporated within an IATA to inform a regulatory decision. A WoE assessment may be made on the basis of this information. Careful consideration must be given to the consistency of the data from the various sources during the WoE evaluation. Standardised templates are under development at the OECD for the documentation of structured approaches that are potentially useful within IATA. These can be used alongside reporting standards for other fixed IATA components (e.g. QSARs, non-guideline *in vitro* methods) to support conclusions made within IATA. At present, there is no overarching approach for the documentation and evaluation of IATA (describing also the expert judgements formed as a result of the WoE evaluation).

Endocrine Disruption

Data are taken from *in vitro* Toxcast assays that are related to key steps in the ER signalling pathway, including ER binding, dimerisation, chromatin binding, transcriptional activation and ER-dependent cell proliferation. The IATA includes official methods (ERTA) and non-guideline assays, most of which are taken from ToxCast.

The integrated prediction model can be used to directly fulfil the information requirements. The 18 different assays are integrated by a mathematical model of the ER signalling pathway. Model inputs are the efficacy for each assay across the full concentration response curve which is effectively the area under the curve (AUC). Model outputs are the integrated AUC score for agonist and antagonist activity and for various “pseudo-receptors” representing different mechanisms of assay noise. Chemicals with AUC (agonist) ≥0.1 are considered ER active, and are predictive of the positive uterotrophic assay results.

Acute Systemic Toxicity

ECHA has developed a model for prediction of low acute oral toxicity; it is based on Weight of Evidence (WoE) approach. The goal is to promote the use of alternatives to animal testing and to save costs. The essential element of this WoE is the sub-acute toxicity study. ECHA's analysis has shown that in case NOAEL oral of the 28-day study is ≥ 1000 mg/kg bw, the probability that LD_{50}oral is > 2000 mg/kg bw is very high. Thereby, the standard *in vivo* test for acute oral toxicity can be “waived”. Furthermore, the correct classification (i.e. no classification for the acute oral toxicity) can be derived. This adaptation can be applied for substances registered at 10 tons or more per year,

because for these substances the 28-day study will normally be available. In addition to the oral 28-day study, at least one more element needs to be provided within the WoE by the REACH registrants. The other possible elements are *in vitro* cytotoxicity (NRU) test, results of QSAR models, and read-across data. This WoE adaptation is described in an update of ECHA's CSR Guidance, which will be published during the first half of 2016.

The acute oral toxicity adaptation is primarily based on the prediction derived from a sub-acute toxicity study. It is meant for low toxicity substance, and requires that other additional elements are presented as a Weight of Evidence (WoE) adaptation.

Pros of the acute oral toxicity adaptation, as defined above:

- Maximises the use of in vivo data, which has to be provided anyway, i.e. sub-acute oral toxicity study above 10 tpa under REACH,
- Encourages the use of other non-animal approaches, e.g. *in vitro* study, QSAR and relevant physico-chemical information, and
- Saves animals and costs in up to 500 cases of registration, according to an estimation made by ECHA.

Cons:

- The WoE approach is not easy to communicate; therefore, it may be considered easier to directly enter to the standard *in vivo* testing for acute oral toxicity
- This approach may in some cases lead to extra use of animals: the case would be such, where first a dose range finding (DRF) study is done, then ordinary sub-acute toxicity study, and because toxicity is seen, finally acute oral toxicity is considered necessary. In that case, the animal saving approach could have been: to start with acute oral toxicity study and then **IF** it can replace the DRF, proceed to the sub-acute study.

To iterate on the last point, other type of cases (where animals and costs are saved) will form the majority of the cases, e.g.

- DRF and sub-acute toxicity study both confirm low toxicity and DRF was considered necessary; then acute toxicity can be waived (Note that about 80% of industrial chemicals are not classified for acute oral toxicity).
- DRF can be avoided (e.g. there is existing information which suggest low toxicity) sub-acute study indicates low toxicity, and other WoE element are added to obtain a valid waiver under REACH.

Obviously, whenever a registrant wants to use this WoE based adaptation, he needs to *start systemic toxicity testing with the sub-acute oral toxicity study* (or possibly with a screening study OECD TG 421) *and not* with the acute oral toxicity study.

Break-out group questions

PARERE and ESTAF members were invited to reflect on the following questions, taking into account the information provided in the case studies. Each break-out group was asked to provide the rationale behind its opinions.

1) In cases where test methods are component parts of an IATA should the IATA be established prior to formal validation of the methods, or *vice versa*?

2) Which regulatory instruments and procedures[1] offer the best means of ensuring broad acceptance and use of IATA?

The main points generated by each break out group have been summarised and included in the table below[2]. A more general summary of the plenary discussion is also included.

[1] e.g. OECD guidance documents on IATA incorporating OECD Test Guideline methods and/or non-test guideline methods; REACH guidance documents .

[2] N.B.: The table exclusively summarizes the views of the Break-out groups. It does not necessarily represent the views of EURL ECVAM.

Group	1) In cases where test methods are component parts of an IATA should the IATA be established prior to formal validation of the methods, or vice versa?	2) Which regulatory instruments and procedures offer the best means of ensuring broad acceptance and use of IATA?
1.	• There is not a single truth but many possibilities. • Post-validation is an important aspect: it is a good idea to validate the methods but as they evolve they need to be revised. • The methods should be globally available and accepted; they could form part of an OECD test guideline. • The question was raised regarding in-house validated test methods: is there a way of certifying these and how can these then be used in an IATA? • There are concerns about the animal data that is used to validate methods against, including issues relating to uncertainty and relevance.	• Guidelines and strategies are required to convince the regulators as well as being important for industry. ECVAM recommendations and strategies may help, in terms of being a label of quality. • There needs to be a shift towards defining and validating standards that fit a wide range of similar classes of methods. Standards could be developed for systems and methods. • Validation of in-house methods against standards developed by an independent body may be a move towards the inclusion of these methods within IATA. • The regulators will need to be educated in order to ensure that they are open to new products. • Guidelines are also very important for industrial users. In particular, guidance on how to perform assays to ensure between-laboratory-reproducibility (BLR).

2.	• The IATA is a helpful framework for validation, but it shouldn't be mandatory.
	• Both can happen in parallel and are needed: it is an iterative process. However, the purpose of the IATA should be clear.
	• Additional elements may be needed to e.g. address additional key events, scientific developments, new mechanisms, etc.
	• Within an IATA, one could assume to use WoE for each step (or addressed KE). There might be even 3 test methods (TMs) looking at the same KE but one can choose the one with the right applicability domain. TMs for early KEs will have a narrower applicability domain, whilst more downstream KE TMs will have a broader applicability domain.
	• It is important to ensure that IATA is adapted.
		• There is a need to develop a reference document on harmonized sufficient requirements for regulatory acceptance of IATA in each endpoint.
		• Acceptance of methods is currently covered by the regulation but regulatory acceptance of IATA is currently unclear.
		• Important to take into account in the Document:
		  ○ Limitations of animal test
		  ○ Human relevance
		  ○ Quantitative and probabilistic approaches (e.g., uncertainty)
		• Define which amount of data is good enough whilst ensuring flexibility in interpretation of data.
		• This is especially important for acceptance of negative results.
		• One suggestion is to have a reference document that outlines the minimum requirements for regulatory acceptance of IATA providing a general framework.

	• Similarly, more than one test method may be validated per IATA key event to take into account, for example, the applicability domain and scientific developments.	• To be developed by regulators together with stakeholders (industry, academia, etc.) • The approach should be prescriptive on the data needs but more flexible on the data interpretation. One example to check could be the ICH guidelines in the pharmaceutical area.
3.	• A starting point could be a scoping paper which seeks to provide an idea of the test methods available for a specific endpoint and an idea of an IATA like the OECD thyroid scoping paper and the preliminary work for non-genotoxic carcinogenicity. • This is important to select most promising methods for validation and identify areas with needs for new methods. This supports also harmonisation of reference chemicals and data over the various new methods. . • Individual methods could be validated by assessing their reproducibility and the relevance in terms of the molecular pathways underlying the cellular response. The IATA cannot be validated but assessed for being “fit for purpose” in terms of mechanistic coverage of underlying AOP knowledge and overall reproducibility. The IATA may harbour one or several ITS (defined approaches) which also may be validated in the more classical way.	• At the moment, there are at least 12 solutions for skin sensitisation, but they are not ranked and provide different results. This can pose difficulties for small companies which have limited resources and cannot afford to tailor them each time for different methods. This also presents problems for authorities and regulators to decide what is acceptable. • The high flexibility of IATAs, like read-across, is also a suboptimal situation. There may be a need to move towards more prescriptive IATAs. An OECD framework would be essential for making progress in this respect. • OECD TG methods would improve the acceptance of conclusions in IATAs. • It is also important to ensure that data generated by non-standard methods are recognised. • Convincing guidance for acceptability criteria for IATAs would help. • OECD Guidance Documents and ECHA guidance and established EURL-ECVAM processed documents are useful approaches for harmonisation. • There may be a requirement for OECD to focus on developing testing strategies as opposed to individual methods. • Development of performance standards for methods and eventually ITSs is important.
4.	The idea of an IATA as a process as opposed to a fixed entity would allow the identification of test development needs. The purpose of the IATA could be determined in order to cover regulatory requirements. In this case, there could be a "pre-IATA" as proposed in the scheme below:	• This backbone IATA is the common IATA agreed on by different regulatory bodies and contains the minimal requirements. • Specific IATAs can be developed for the specific areas. • It also has to be very clear what the IATA should be used for (e.g. classification and labeling). Once this is clarified it is easier to harmonise the IATA.

	Pre-IATA → Backbone-IATA → IATA → Condensed-IATA  • The pre-IATA can be seen as a conceptual framework including existing methods and to identify gaps. It is a working scheme that can show where new test methods are needed. It is for the scientific community and not for regulatory purpose. • In the IATA formats there should be a legislation box so that it is clear under what regulations it is applicable.	• The accepted IATA will need regular revisions and weak building blocks should be identified. It is when the IATA is used that we get experience to fully re-evaluate its performance. • Need for case studies to show that the concept is usable. • The condensed IATA may be transformed into an integrated testing strategy (ITS) and included in a test guideline.

Discussion points:

- IATA are judgement-based approaches which may contain rule-based elements. An example of a rule-based element within IATA is a defined approach to data integration (e.g. integrated testing strategies), where a combination of multiple information sources is fixed in a predefined way. In the case of Skin Corrosion and Irritation, Part III of the IATA, which is linked to the generation of new testing data, is prescriptive. For skin sensitisation, however, this part is also flexible as the way in which the *in vitro* methods may be combined is not prescribed. For this reason, a template for describing defined approaches is being developed at OECD to facilitate harmonised reporting and evaluation of such approaches.
- The knowledge from Adverse Outcome Pathways (AOPs) has informed the development of tests which target key events within the pathway. IATA, therefore, are the regulatory tools that may use AOPs to underpin decisions with mechanistic knowledge. In the case of skin sensitisation, a number of alternative methods have been developed to address specific key events in the skin sensitisation pathway. The combination of *in vitro, in silico* and *in chemico* tests can provide sufficiently reliable results to forego the *in vivo* test. The question again is how much data is enough data and is it necessary to cover all of the key events?
- IATA is an emergence of a more standardised way of documenting hazard and risk assessment. However, flexibility is required as it is a chemical and context dependent process. It provides a framework for testing and assessment which may be adapted for different regulatory requirements. As such, it is necessary to have a clear understanding of the different regulatory requirements for each sector across the regions in order to facilitate a harmonised approach.
- The WoE approach is not considered to be as strong as using validated *in vitro* methods. Combining more *in vitro* methods still raises the question of acceptance and again, what is sufficient data? This could lead back to *in vivo* testing in order to satisfy the regulatory requirements.
- Regarding the acute oral toxicity adaptation presented, some industry representatives commented on the meaning of “toxicity is seen” in real life. The meaning of a DRF was to obtain good dose levels for sub-acute studies. If it is not known if animals tolerate a certain dose level (because no acute toxicity data are available), there is a risk to end in additional dose groups because the initial choice was wrong. It is thus rather unlikely that someone would take the risk to set wrong dose levels just by avoiding DRFs or acute toxicity studies.
  It was explained that "toxicity is seen" means that the NOAEL of the sub-acute study is <1000 mg/kg and, therefore, the WoE adaptation cannot be applied and the registrant would need to perform the acute toxicity study. This scenario would indeed end up with more animals used.
- The problem with occupational health and safety and GHS was also raised. In this context limit dose for repeated-dose studies (1000 mg/kg bw/d) is not declared as comparable to limit dose in acute toxicity studies (2000 mg/kg bw). Classification and labelling are geared with these values.
- Saving animals by following the strategy might be accepted in Europe. However, many products also have to be registered in other parts of the world and other authorities ask for acute toxicity data.

- Read across could be an option, however, this is often refused by authorities (especially outside Europe).
- QSAR for acute toxicity has been tested by industry, but the experience shows that it poorly or not at all covers this endpoint.
- According to the experience in industry, cytotoxicity assays failed to predict starting doses and they would not help on their own for predicting acute toxicity of industrial chemicals (Schrage *et al*., 2011[3]).

  It was clarified that the intended use of the 3T3 NRU cytotoxicity assay as part of the proposed WoE adaptation was not for setting starting doses for the *in vivo* acute toxicity assay, but as supportive evidence of the low toxicity of the substance. In fact, the potential of the 3T3 NRU cytotoxicity assay to support the identification of chemicals that do not require classification for acute oral toxicity (LD50 > 2000 mg/kg b.w.) has been shown (Prieto *et al*., 2013[4]). However, due to the known limitations of the cytotoxicity assay, the test cannot be used in isolation and therefore, it has been recommended to always interpret the results in combination with information from other sources in order to increase the confidence in the prediction (EURL ECVAM 2013[5]).

- The validation and regulatory acceptance processes pose some issues: this must be changed in order to drive progress. The issues include:
  - Regulators have varying levels of understanding.
  - There is a significant problem relating to the rejection of submitted alternative method data: often they are rejected but there is no rationale included. Regulators need to agree on what does and what doesn't fulfil the criteria.
  - The financial burden associated with developing methods for submission for validation can act as a disincentive, particularly to smaller companies.
  - Lack of between laboratory reproducibility (BLR) – but is this important when some laboratories do not transfer their methods? GD34 states that in-house methods do not require a BLR assessment. The transferability stage of the validation process is time-consuming. If it is not necessary for in-house methods which are not used in other laboratories, then more methods could be validated. However, these could be validated through a different process (e.g. by checking compliance against an agreed set of standards for the given type of method, i.e. they won't become international test guidelines.
  - How to understand that a given tool is mechanistically sound without validating it against animal data. These animal data are not 100% accurate or reliable. Confidence is needed in order to move away from the dependence on the animal model.
- Is it possible to validate an IATA?

[3] Schrage A, Hempel K, Schulz M, Kolle SN, van Ravenzwaay B, Landsiedel R. Refinement and reduction of acute oral toxicity testing: a critical review of the use of cytotoxicity data. Altern Lab Anim. 2011 Jul;39(3):273-95. Review

[4] Prieto P, Cole T, Curren R, Gibson RM, Liebsch M, Raabe H, Tuomainen AM, Whelan M, Kinsner-Ovaskainen A. Assessment of the predictive capacity of the 3T3 Neutral Red Uptake cytotoxicity test method to identify substances not classified for acute oral toxicity (LD50>2000 mg/kg): results of an ECVAM validation study. Regul Toxicol Pharmacol. 2013 Apr;65(3):344-65.

[5] EURL ECVAM Recommendation on the 3T3 NRU Assay for Supporting the Identification of Substances Not Requiring Classification for Acute Oral Toxicity. Available at: https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/3t3-nru-recommendation.

- The benefits and feasibility of validating IATAs rather than only the single test methods was discussed: some ESTAF and PARERE members voiced the need to validate IATAs in order to achieve full regulatory use of non-animal methods (which are usually used in combination with several methods to address one endpoint). EURL ECVAM explained that some parts within an IATA are difficult to standardise (and validate), i.e. usually the parts gathering existing information and their evaluation in a judgement-based weight-of-evidence approach, but that defined approaches (DA) to data integration, where a combination of multiple information sources is fixed in a predefined way, could, on the other hand, be validated.
- EURL ECVAM is therefore leading the project within the OECD Task Force on Hazard Assessment on a Guidance Document for the reporting of defined approaches to be used within IATA.
- Regarding the 12 case studies on skin sensitisation currently discussed at the OECD Task Force on Hazard Assessment, most of those are not IATAs but defined approaches (DA). DA are to be considered in the context of an IATA and depending on the regulatory purpose, some of them may in certain cases also be fit for purpose on their own.
- The process for evaluation/validation of DA is not necessarily the same as the one that is used to validate individual methods. In addition, the validation process of the individual methods is different depending if the methods are to be used as stand-alone methods or as components of DA/IATA. This concept had already been applied in recent EURL ECVAM validation studies (e.g. skin/eye irritation versus skin sensitisation).
- The existence of the IATA framework could also guide the development of new methods as gaps could be more easily identified in strategies to address specific endpoints.

- Should the IATA be outside of TGs?
  - Yes, they should be under Guidance Documents. Acceptance of fixed IATAs under MAD is a possible goal, but if this is to be achieved then validation of the assays becomes a necessity.
- The regulatory acceptance of IATA is currently unclear.
  - A reference document on harmonised standard requirements for regulatory acceptance of IATA for each endpoint is necessary to define the sufficient amount of data whilst ensuring flexibility in the interpretation of the data. This could be developed by regulators together with stakeholders.
  - Such a reference document needs to take into account the limitations, the human relevance and the uncertainties of the animal data.
- Regulatory acceptance may be approached in the following ways:
  - if the IATA is based on either validated or OECD test methods
  - if the IATA has elements which can be omitted (e.g. if specific data are not required for the category), then if this is explained it is possible to accept the data
  - the regulators need a reference point and this could be the minimal set of components and performance requirements associated with the IATA.
  - a core element of getting acceptance is experience of using the data.

- Although there is an emergence of guidance, it is difficult to promote when the *in vitro* methods may be expensive and there is no guarantee of regulatory acceptance. However, this is not a reason to return to animal testing and everyone involved is an agent for change in this area. International principles need to be agreed upon and the reporting format and amount of data needs to be considered in order to promote mutual acceptance.
- Compliance with article 13 (REACH) needs attention. It is important to push for change using legislation and to require registrants to provide a thorough scientific justification why an *in vivo* method has been used where *in vitro* methods are available. However, the legal competence of MS and of ECHA, in terms of how to enforce compliance, is not clear. In this respect, the use of IATA could guide the decisions made by the registrant and facilitate the identification of non-compliance so that testing proposals containing avoidable *in vivo* tests may be rejected.

Conclusions

IATA Strengths

- IATA can be used as a way of targeting new methods for development and validation. As a framework for developing a testing strategy to address a particular endpoint, they can be used to identify where there are gaps to develop new test methods. In this way, they present a structured approach to the development, validation and uptake of alternative methods.
- They offer flexibility in some cases, but can also be prescriptive where needed.
- This is also a route to bringing legislators from different sectors and regions together to develop a standardised approach.

Suggestions

- There needs to be a consistent format for the reporting and description of the information sources used.
- Industry needs more information on what are the specific data requirements.
- Development of performance standards is important.
- IATAs need regular revision in order to optimise the methods which are incorporated.
- Authorities need to be more explicit in their justifications about why assessments may not be sufficient. Data requirements should be fixed and clearly stated.
- Confusion exists in terms of how to apply the results from *in vitro* test methods in a regulatory context. Again, this requires a clearer definition of what types and how much is sufficient.

Annex I

PARERE-ESTAF Workshop on Integrated Approaches to Testing and Assessment (IATA) - method validation, testing strategy development, evaluation and acceptance.

Objectives of the workshop

The aim of this workshop is to discuss key issues relating to the development, evaluation, acceptance and use of Integrated Approaches to Testing and Assessment (IATA). Four case studies will be presented to give an overview of different kinds of IATA that are currently considered by regulatory bodies or are under development. These IATA are designed for the assessment of eye irritation (1), skin sensitisation (2a, 2b), endocrine disruption (3, 4) and acute systemic toxicity (5). Workshop participants are invited to discuss a series of questions, with a view to identifying the pros and cons of different options for promoting the acceptance of IATA.

General background on IATA

IATA offer a means of integrating and translating existing and/or newly generated toxicity data, thereby serving as flexible tools for chemical safety assessment and (regulatory) decision making (6). In addition to traditional *in vitro* and *in vivo* testing methods, IATA are increasingly incorporating newly developed *in vitro* systems and measurement technologies such as high throughput screening (HTS) and high content imaging (HCI). Computational approaches are also being used, both as a means of generating data (e.g. QSARs), interpreting data (bioinformatics and chemoinformatics), and as a means of integrating multiple sources of data (e.g. *in vitro* prediction models, expert systems, physiologically based kinetic/dynamic models).

For the purpose of this workshop, the following working definition of IATA will be used:

> *An IATA is a framework for hazard identification, hazard characterisation and/or safety assessment of a chemical or group of chemicals, which strategically integrates and weights all relevant existing data and guides the targeted generation of new data where required to inform regulatory decision-making regarding potential hazard and/or risk*.

This definition was agreed by the OECD expert group charged with the development of a GD on the reporting of IATA (2a).

Within an IATA, one or more elements result in predictions and/or conclusions. In some cases, these elements are fixed or prescriptive in terms of how the predictions are derived and interpreted, referred to here as "structured approaches" to data integration, whereas in other cases, the prediction and assessment approach is flexible. For example, the interpretation of *in vitro* test data by means of one or several prediction models combined in a fixed strategy (7) is a structured approach, as is the use of a mathematical model such as a quantitative structure-activity relationship (QSAR) or a physiologically based kinetic/dynamic (PBK/D) model. In contrast, flexible approaches include grouping and read-across, non-formalised weight-of-evidence, and any subjective

interpretation typically referred to as "expert judgment". Flexibility does not necessarily imply lack of transparency, provided that the prediction and assessment approach is clearly documented. The distinction between structured and flexible approaches is important, in the sense that structured approaches can in principle be scientifically validated (8).

As illustrated by the workshop case studies, existing IATA represent very diverse kinds of solutions for toxicological assessment and decision making. At one extreme, an IATA can be explicitly described and prescriptive, leaving little or no room for expert choices. At the other extreme, IATA can be loosely described and flexible, allowing multiple options for the user/assessor. Some IATA are endpoint-specific, whereas others address multiple types of toxic effect. Some include exposure considerations, whereas others do not. Some include the option for animal testing, whereas others may be animal-free. IATA can also differ in the extent to which they apply mechanistic reasoning (e.g. Adverse Outcome Pathways; AOPs) and use mechanistically relevant data. Furthermore, IATA can be designed to directly fulfil an information need/requirement (e.g. classification of skin sensitisers), or to provide information supportive of a broader assessment (e.g. prediction of adverse estrogenic effects in the context of endocrine disruption).

The development of an IATA can be thought of as an optimisation problem, with one or more optimisation criteria that depend on the problem formulation. From a regulatory perspective, the most usual criteria considered are the ability of the IATA to generate relevant and reliable results, and to reduce or replace animal testing. From an industry perspective, additional criteria could include costs, time and likelihood of regulatory acceptance. It is likely that there will be trade-offs between some of these criteria – for example, the generation of more reliable and relevant data may require the use of more expensive test systems. Thus, different solutions will be more or less suitable depending on the decision-making context, which means that an IATA and its component parts need to be "fit-for-purpose". In other words, the optimal design of IATA is highly context dependent, making it difficult or impossible to achieve a one-size-fits-all solution. This has implications in terms of how the acceptance of IATA should be treated at the regulatory level.

Break-out group questions

PARERE and ESTAF members are invited to reflect on the following questions, taking into account the information provided in the case studies. Each break-out group should provide the rationale behind its opinions.

1) In cases where test methods are component parts of an IATA should the IATA be established prior to formal validation of the methods, or *vice versa*?
2) Which regulatory instruments and procedures[6] offer the best means of ensuring broad acceptance and use of IATA?

[6] e.g. OECD guidance documents on IATA incorporating OECD Test Guideline methods and/or non-test guideline methods; REACH guidance documents .

References

1. Worth AP et al (2014). Alternative methods for regulatory toxicology – a state-of-the-art review. JRC Science and Policy Reports. Report EUR 26797 EN. Publications Office of the European Union, Luxembourg. ISBN: 978-92-79-39651-9.

2a. OECD (2015a). Draft guidance document on reporting IATA. (ENV/JM/HA(2015)7). Under review, not yet published.

2b. OECD (2015b). Draft guidance document on the reporting of structured approaches to data integration and individual information sources used within IATA for skin sensitisation (ENV/JM/HA(2015)8). Under review not yet published

3. Judson et al (2015). Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 *In vitro* High Throughput Screening Assays for the Estrogen Receptor. Toxicological Sciences, in press. Early version available from: http://toxsci.oxfordjournals.org/content/early/2015/08/12/toxsci.kfv168.full.pdf+html

4. Browne P et al (2015). Screening chemicals for estrogen receptor bioactivity using a computational model. Environmental Science and Technology. In press. http://ntp.niehs.nih.gov/ntp/about_ntp/sacatm/2015/september/screeningchemicals_draft.pdf

5. ECHA (2015). Draft update to guidance on information requirements for acute systemic toxicity. Under review, not yet published.

6. Report of the workshop on a framework for the development and use of Integrated Approaches to Testing and Assessment. ENV/JM/HA(2015)22. Available from: http://www.oecd.org/env/ehs/testing/reportoftheworkshoponaframeworkforthedevelopmentanduseofintegratedapproachestotestingandassessment.htm

7. Worth AP & Balls M (2001). The importance of the prediction model in the validation of alternative tests. Alternatives to Laboratory Animals 29:135-144.

8. Kinsner-Ovaskainen A, Akkan Z, Casati S, Coecke S, Corvi R, Dal Negro G, De Bruijn J, De Silva O, Gribaldo L, Griesinger C, Jaworska J, Kreysa J, Maxwell G, McNamee P, Price A, Prieto P, Schubert R, Tosti L, Worth A, Zuang V. (2009). Overcoming barriers to validation of non-animal partial replacement methods/Integrated Testing Strategies: the report of an EPAA-ECVAM workshop. Alternatives to Laboratory Animals.37:437-444.

9. Scott L et al (2010). A proposed eye irritation testing strategy to reduce and replace *in vivo* studies using Bottom-Up and Top-Down approaches. Toxicology *in vitro* 24: 1-9.

Characteristics of IATA for different endpoints

	Question	Serious Eye Damage/Eye Irritation	Skin Sensitisation	Endocrine Disruption	Acute Systemic Toxicity
1	Does the IATA result in a prediction or assessment that directly fulfils an information need or regulatory requirement?	Yes - classification as seriously damaging the eye (Cat 1), eye irritant (Cat 2) or not-classified	Yes – classification as sensitiser or non-sensitiser	Partially – prediction of ER agonism, antagonism, and activity in the uterotrophic assay	Partially – identification of chemicals that are not classifiable under GHS (GHS Cat 5), but no discrimination between classifiable chemicals (GHS Cats 1-5). CLP only implements cat1-4 of the GHS.
2	To what extent is the IATA mechanistically based?	Most of the available *in vitro* methods cover one or several toxicological modes-of-action associated with the development of eye irritation *in vivo*, such as (i) cell membrane lysis, (ii) protein coagulation/denaturation, (iii) saponification, (iv) chemical reactivity (8). An AOP has nevertheless not been developed.	Information elements typically refer to key events in the OECD-adopted AOP for skin sensitisation. However, key event information is not necessarily used according to the AOP sequence, and the quantitative relationships between key events is not known.	Data are taken from *in vitro* Toxcast assays that are related to key steps in the ER signalling pathway, including ER binding, dimerisation, chromatin binding, transcriptional activation and ER-dependent cell proliferation	Data from *in vitro* basal cytotoxicity assays, which are related to general mechanisms of toxicity common to most cell types, that are indicative of low cytotoxicity. Basal cytotoxicity is a key event in many prevalent toxicological modes-of-action associated with acute health effects. AOP is not yet available.
3	Does the IATA include test systems that are official methods (EU test methods or equivalent)?	Yes – BCOP (TG 437), ICE (TG 438), FL (TG 460), STE (TG 491), EpiOcular™ EIT (TG 492), Draize eye test (TG 405), as well as all of the skin corrosion test methods (TGs 430, 431, 435 and 404)	Yes – DPRA, KeratinoSens, hClat (almost)	Yes in case of ERTA?	Yes in the case of TG407, TG422

	Question	Serious Eye Damage/Eye Irritation	Skin Sensitisation	Endocrine Disruption	Acute Systemic Toxicity
4	Does the IATA include test systems that are non-standard (non-guideline) methods?	Yes – e.g. CM, HET-CAM, IRE, Ocular Irritection, PorCORA, EVEIT and others – some even to predict potential effects not covered by the current guideline methods.	Yes – in the case of some IATA solutions.	Mostly yes – Toxcast assays	Yes - in the case of *in vitro* basal cytotoxicity, QSAR, read across
5	Is the output of each test system on its own associated with its own prediction model? If so, what form does this take (equation, decision rule, etc)? Can any of the prediction models be used to directly fulfil the information need/requirement?	Yes – most approaches are/will be based on sequential testing strategies, where each individual method and its validated prediction model is a decision point for classification/no classification or for further testing (as described in individual OECD TGs). Most prediction models are based on categorical decisions taken from one or several endpoints, some of these being qualitative and others quantitative.	Official and validated test methods are associated with prediction models, but these are not always used directly in the IATA. None of these PMs can be used to directly fulfil the information need (according to the test guidelines) since they are considered to be "mechanistic building blocks"	No	In the case of sub-acute studies decision rule (if NOAEL ≥1000 mg/kg); Validated *in vitro* cytotoxicity assay is associated with a prediction model (equation). The final decision, however, needs to take into consideration several pieces of information to build the WoE argumentation. Therefore, the PMs cannot be used directly to fulfil the information need.

	Question	Serious Eye Damage/Eye Irritation	Skin Sensitisation	Endocrine Disruption	Acute Systemic Toxicity
6	Are the outputs of multiple test systems associated with an "integrated prediction model"? If so, what form does this take (equation, set of equations, formalised weight of evidence, etc.). Can the integrated prediction model be used to directly fulfil the information need/requirement?	(Yes?) – Some structured data integration solutions may use the data obtained with the individual methods in a different way than prescribed by their validated prediction models, using statistical methods such as discriminant analysis, flexible discriminant analysis, support vector machines, classification trees, k-nearest neighbours, etc.	Yes – in the case of some IATA solutions, e.g. decision trees based on multiple test method outcomes. Examples are given in ref 2b.	Yes, the outputs from 18 different assays are integrated by a mathematical model of the ER signalling pathway. Outputs are the efficacy for each assay, and the (agonist and antagonist) activity (AUC) of each type of receptor in the pathway.	No
7	Does the IATA include one or more predictions from mathematical models (e.g. QSARs)? If so, what is the intended use of these predictions (contribute to weight of evidence, trigger/inform testing, etc)	Yes – for supporting the decision to initiate either a bottom-up or top-down approach when new testing is required, as well as to contribute to WoE-based decisions.	Yes – in the case of some IATA solutions, e.g Bayesian model, QSAR, PBK, JRC decision tree classifier. The predictions contribute to the final assessment.	Yes, the integrated prediction model described in 6 directly predicts ER agonistic and antagonistic activity. Chemicals with AUC (agonist) ≥0.1 are considered ER active, and are predictive of the positive uterotrophic assay results	Yes – QSARs can be part of the WoE approach.
8	Does the IATA include grouping and read-across? If so, is the approach defined?	Yes – the approach to follow is not specifically defined	Yes – in the case of some IATA solutions, read-across may directly fill the data gap	No	No – the approach is not defined?

	Question	Serious Eye Damage/Eye Irritation	Skin Sensitisation	Endocrine Disruption	Acute Systemic Toxicity
9	Does the IATA include a non-formalised Weight of Evidence assessment?	Yes – on the basis of all collected existing information to decide if a sound conclusion can be taken or to otherwise inform the generation of new data, where required.	Yes – in the case of some IATA solutions	No	Yes
10	Does the IATA include the use of exposure/bioavailability information? If so, how is this information used (in risk characterisation characterisation, exposure-based waiving, exposure-based testing)?	No – The IATA only deals with classification and labelling.	Yes – in the case of some IATA solutions, the absence of dermal penetration may be considered in the WoE	No	Yes – bioavailability information (as indicated by certain physicochemical properties) can be used as one element of the WoE approach, i.e. relevant data on low bioavailability as indicator of low toxicity?
11	To what extent does the IATA explicitly identify and characterise sources of uncertainty? → Not at all; qualitative, semi-quantitative; quantitative	In general, the sources are not explicitly identified, but each individual information source is described using the same template, where quantitative measures of predictive capacity and reproducibility are given, and applicability domain, limitations, strengths and weaknesses are described.	In general, the sources are not explicitly identified, but quantitative measures of predictivity are generally given	Cooper performance statistics are available for prediction of ER active chemicals, and for prediction of uterotrophic assay results	The quality (e.g. Klimisch scores) and consistency of the data should be taken into account when weighing each piece of available information.