IDEA
International Dialogue for the Evaluation of Allergens

IDEA Workshop on the

replacement of animal testing in QRA for skin sensitisation

May 16-17, 2018

Hotel Amigo
Rue de l'Amigo 1, 1000 Brussels (Belgium)
+32 2 547 47 47

<u>Final Report from *the Rapporteur*, Prof. James Bridges</u>

The primary aim of the workshop was to enable an expert discussion on non-animal based strategies being used / under development for the identification of human dermal sensitisation inducing substances and their potency (numerically expressed not simply their classification) to derive the NESIL to be used in the QRA. These would in essence provide a replacement for the in vivo LLNA test. The presenters were asked to address in particular the:

- Types of data relied on, the sources and how the data is used.
- Basis for the choice of the benchmark substances and the rationale for the dose selection for each.
- Basis for the methodology used to determine potency
- How these estimated potency values are used for risk assessment
- Uncertainties in the assessment and how they are evaluated.

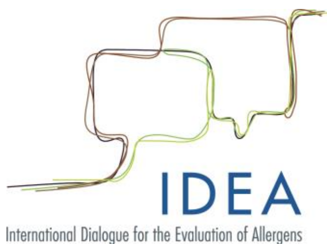
1. Outline of relevant OECD and other international activities

The development of a Test Guideline on defined approaches for skin sensitization were presented by Silvia Casati (JRC) who is heading the project along with USA experts. Although the scope of the OECD project is primarily hazard evaluation for the purpose of classification and labelling, some of the defined approaches under consideration may provide a point of departure to be used in a weight-of-evidence assessment based Integrated Approach to Testing and Assessment (IATA) for dermal sensitisers. She provided a comprehensive update on the OECD activity and the current priorities. It was confirmed that the defined approaches are focused on the first three key events (KE1, KE2 and KE3) in the single, well studied adverse outcome pathway for the induction of dermal sensitisation. This is appropriate as the widely used in vivo test (LLNA) was designed to detect and quantify changes in the 4th of the critical stages in the same pathway (KE4).

OECD recommends that in order to replace the LLNA test, data from more than one non-animal information source is likely to be necessary. The findings from the selected in vitro tests should be evaluated in combination with other types of information (e.g. in silico predictions, structural alerts, physicochemical properties) which can be garnered for the fragrance material of interest, including data banks. The evaluation should be based on a transparent methodology which could be in the form of an algorithm. The pros and cons of this aspect of the strategy were not discussed. The

IDEA Management Team
Avenue des Arts, 6
1210 Brussels, Belgium
Tel: +32-2 214 20 61
Fax: +32-2 214 20 69

www.ideaproject.info



need to identify and address uncertainties such as in vitro test variability, human variability and data gaps was also identified.

2. Approaches discussed at the workshop

This presentation was followed by six presentations on how the OECD IATA framework was being utilised in practice to identify and determine the potency for induction by different companies / organisations. Turning to the questions asked to each presenter:

a) Types of data relied on, its sources and how it is used.

The presentations showed a very good general agreement on the sources of data used but there were some differences in the order in which they are applied and their relative importance.

In vivo reference values for potency.

The majority of the presentations relied on previously produced LLNA test findings. This is appropriate as the LLNA test is the one that must be replaced. However, the assumption is therefore made that the LLNA test for each substance investigated predicts accurately the sensitisation potency in man. In one presentation mention was made of supplementing the LLNA results with human data for reference purposes whereas in another one HRIPT data was used instead as the main source for potency.

Structural and metabolism data.

Structural considerations allow a read across from fragrance materials of comparable structures. Prediction of metabolic activation of a potential pro-hapten is also important as the in vitro tests used mare unlikely to sufficiently reflect the 'xenobiotic metabolism' capability of human and mouse skin in vivo.
All the presentations identified TIMES as the main database used for structural considerations and metabolism. A few supplemented this with additional in silico tools such as the OECD tool box (2), ToxTree, Derek or ACD percepta. What was not clear, because discussion time was limited, is the degree to which the test fragrance materials were already incorporated into the databases that are relied on. If they are, then the accurate prediction of in silico properties and metabolism activation would be expected.

Other sources used for the IATA.

Assessment of the degree of bioavailability/skin penetration was for example noted in two presentations but did not appear to be considered significant in others This was because it was considered that the physico-chemical properties are a sufficient indicator of bioavailability.

b) Basis for the choice of the benchmark substances and the rationale for the dose selection for each.

The main basis for the choice of the bench mark substances appeared to be a combination of trust in the reliability of the relevant LLNA test data as the main basis for the comparison of the findings, and in-house interest in the substances. The rationale for dose selection in each case is assumed to be that set out in the OECD guidelines. Thus, for the DPRA and like tests this requires using predefined ratios of substance to amino acid along with visible changes that might indicate precipitation. For the cell-based tests the top dose is inevitably the level that produces a specified low-level cytotoxicity. This parallels the approach used to set the top dose in LLNA tests which is based on minimal irritancy as has been used traditionally in in vivo toxicity testing generally (maximum tolerated dose, MTD). This provides a consistent procedure from a regulatory viewpoint. There is no requirement to take into account the

estimated worst case real exposures. In two presentations chemical groupings considered challenging were also mentioned.

c) Basis for the methodology used to determine potency.

The primary source of data to identify potency was the results from the in vitro tests. Four presentations used only OECD approved tests. One also utilized the COCAT test. All used a KE1 test and in addition at least one KE2 and a KE3 test. Almost without exception these tests were DPRA (KE1), KeratinoSens (KE2) and hClat (KE3). The approach fits very well with the OECD strategy although the rationale for the selection of these specific tests, rather than others, was only briefly mentioned and may partly depend on availability in the different labs. The way each test was conducted in each organisation was based on the OECD guidelines and the protocols provided by the test supplier (i.e. standard operating procedures). The details of the methodology they used to identify potency in each test varied. Some approaches presented findings categorised into strong, moderate and weak sensitisers, which from a regulatory perspective for hazard-focus regulations but may not be sufficient for the confirmation that the use level is safe Others calculated a numerical value in the form of a 'most likely EC3 value'. Such a numerical value could be directly used, applying adequate safety factors, for quantitative risk assessment.

One approach addressed KE2 and KE3 in a co-culture system of HCaT (human keratinocyte cell line) and THP-1 (surrogate of dendritic cells). This may better reflect the cell-cell interactions that are important in vivo. As far as the COCAT has been studied, it was found to enable the detection of pro-haptens and to increase the dynamic range of the dendritic cell dose response of the set of fragrances studied.

One presentation was focussed on a single test, SENS-IS which is not yet OECD approved formally. This is based on profiling changes in various genes that are up-regulated and is considered to distinguish between sensitisation induction initiation and irritation. The results with the test fragrance materials compared well with their LLNA test findings. Of particular interest is whether the same genetic changes occur with different skin sensitisation inducers, the mechanistic basis for the identification of potency and whether the same genes or others are up-regulated following sensitisation in human skin.

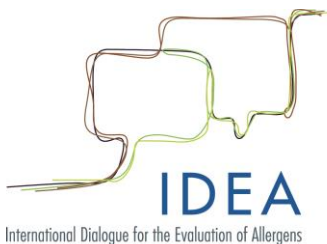
d) How the estimated potency values are used for risk assessment

All the presentations referred to the importance of using the findings from the in vitro tests along with that from the databases identified in section 2.a. above and other sources of information to estimate potency. Weight of evidence was identified in each case as the tool for analysing the data to determine potency. Mathematical regression models, Artificial neural networks or Bayesian networks were cited as the procedures used for this purpose. However, the details of these, including the weighting factors for different pieces of evidence, were not completely accessible to the workshop participants in all presented case studies. Formalised predetermined protocols clearly have the benefit of consistency in data interpretation. However, it may also lead to the rejection of data sources that are not included and thereby discourages innovation. It is important that whichever procedure is used to weigh the evidence the entire process is fully transparent and able to be replicated.

e) Uncertainties in the assessment and how they are evaluated.

The main concern is to avoid false negatives. The primary areas of uncertainty were identified as:

- Reliability of the databases for structural and metabolism information
- Variability in the vitro test findings, but also variability of the in vivo reference data
- Data gaps
- Not sufficient coverage of the AOP by the ITS
- Chemicals falling outside applicability domain



Databases.

Databases are clearly most likely to be a reliable source of the required information if they incorporate structures very similar to the fragrance material under study. Conversely the poorer the data on chemicals structurally related to the fragrance material under study, the greater the uncertainty about the value of the prediction. How this type of uncertainty should be expressed and compensated for was not discussed in any detail although read across was mentioned as a means of reducing the uncertainty in the assessment.

In vitro test variability.

There are three types of uncertainty, namely variability in findings on repetition of a single test, variability of findings between tests either for the same KE or for different KE's, and uncertainty arising from gaps in the test coverage of KE4. In the first two cases a figure of over 15-20% variability has been deemed by some to be acceptable. It is important that any variability deemed acceptable e.g. based on statistical considerations although is justified to ensure transparency. This benchmark was also considered in some presentations as an acceptable agreement level between the non-animal IATA assessment and the NESIL as determined using the LLNA as the test method. It is less clear how uncertainty is addressed between different in vitro tests if there is a qualitative difference or a very major difference in findings. One approach might be to conduct additional in vitro tests. This is an important issue. It also relates to the issue of weighting of evidence and whether KE3 tests should be given a higher weighting than those based on stages earlier in the adverse outcome pathway.

Data Gaps.

The importance of any gaps is determined by the importance attached to them in the overall assessment. This in particular requires agreement on what is the minimum number and type of in vitro test results needed to characterise potency reliably. Other aspects that need to be decided on is whether knowledge of physicochemical properties is sufficient to determine bioavailability and assurance that the metabolism has .been properly considered

Expression of uncertainty.

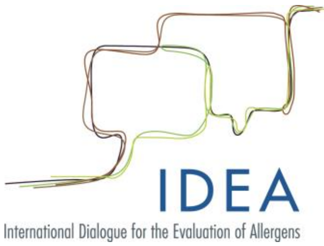
It is apparent from the workshop discussions that how uncertainties are defined and what is the maximum level of uncertainty that should be accepted without the requirement for further testing need further consideration.

3. Overall findings

All the work presented showed that a non-animal based methodology for identifying and characterising potential fragrance dermal allergens is feasible to replace the LLNA test for dermal sensitisation induction. Inevitably at this stage some aspects were identified as needing further work. This however needs to be viewed in the context that with any major changes in methodology for risk assessment there are some challenges needing further resolution.

It was noted that structures such as aldehydes, acylating and amine reacting substances could yield misleading results unless they were picked up at the outset of their assessment as difficult and requiring additional evaluation or evaluation by specific tests.

A number of factors are identified above where further detail/work is needed. In addition, it would be helpful to identify priorities for further development of the non-animal based IATA, to produce a spreadsheet covering all the



fragrance materials identified in the presentations and the in vitro test findings and LLNA results to enable a picture of:

- The range of structures assessed.
- The reproducibility between organisations in findings and interpretation where the same fragrance material has been studied.
- The consistency in findings between the KE1, KE2 and KE3 tests for each fragrance material.
- Structures which appear to give more problematic results.
- The value of considering additional information e.g. bioavailability in the assessment of potency.
- Any particular challenges in identifying and characterising weak inducers.
- The types of structure (if any) where false positives or false negatives are most likely.

The series of presentations was complemented by an update on the Research Institute for Fragrance Materials on their Dermal Sensitization Research Program, which largely supports a number of the approaches presented by producing data to broaden the databases.

4. Additional issues to consider?

Do we know enough about the critical steps in the AOP for dermal sensitisation (induction) on which the current in vitro test design is founded to be sure that this is the sole initiating route for dermal sensitisation in humans? Moreover, if the pathway is KE1→KE2→KE3, which one is the rate limiting step and why if it is a linear process is it important to have tests for all three steps?
What can we learn from previous work on the development of non-animal tests for risk assessment purposes e.g. genotoxicity testing for carcinogens that may be of help in prioritisation for further work?
Effect of other ingredients in fragrance products were not considered at the workshop. Should this be done or is the assumption justified scientifically that other ingredients will not significantly impact on the AOP?
Are there benefits to be had in terms of the further development of the methodology in making more effort to collaborate with other academia/industrial groups who also are developing non-animal testing strategies for dermal contact allergens?
How should IDEA proceed organisationally with any/all of the above?

Jim Bridges revised September 2018