PUBLIC CONSULTATION “DRAFT REPORT ON ALTERNATIVE (NON-ANIMAL) METHODS FOR COSMETICS TESTING: CURRENT STATUS AND FUTURE PROSPECTS -2010”

COMMENT TEMPLATE

CHAPTER OF REPORT* (PLEASE TICK THE BOX OF THE CHAPTER YOUR ARE COMMENTING ON AND PROVIDE ONE TEMPLATE PER CHAPTER):

☐ Chapter 1 - Repeated dose – including chronic, sub-chronic and sub-acute exposure
X Chapter 2 - Skin sensitisation
☐ Chapter 3 - Carcinogenicity
☐ Chapter 4 - Toxicokinetics
☐ Chapter 5 - Reproductive toxicity

SUBMITTED BY (NAME): Dr Katy Taylor

AFFILIATED TO: ECEAE (European Coalition to End Animal Experiments)

Type of organization:
☐ Industry
☐ Government
X Animal Welfare Organisation
☐ Academia
☐ Other –please specify

ON: 15/10/2010 TO BE TREATED CONFIDENTIAL: X NO ☐ YES
**General comments**

These comments are prepared by Dr Katy Taylor, Dr Carlotta Casalegno, Dr Wolfgang Stengel and Dr Irmela Ruhdel of the European Coalition to End Animal Experiments (ECEAE). Our comments have also been reviewed by Dr Dave Roberts and Dr Steve Enoch, Liverpool John Moores University, UK.

The ECEAE is Europe's leading alliance peacefully campaigning on behalf of laboratory animals. Formed in 1990 by animal organisations across Europe to successfully campaign to ban cosmetics testing on animals, the Coalition now leads campaigning on all animal experimentation issues in Europe. The ECEAE is made up of the following organisations **ADDA (Spain), Animal (Portugal), Animal Friends Croatia (Croatia), Animal Rights Sweden (Sweden), Animalia (Finland), BUAV (UK), Deutscher Tierschutzbund (Germany), Dyrevernalliansen (Norway), EDEV (Netherlands), Forsøgsdyrenes Værn (Denmark), GAIA (Belgium), Irish Anti-Vivisection Society (Ireland), LAV (Italy), One Voice (France), People for Animal Rights (Germany), Doctors Against Animal Experiments (Germany), Svoboda zvírat (Czech Republic), SSPA (Switzerland).**

It is our understanding that the purpose of the public consultation on the expert reports is to establish whether the terms of reference (ToR) given to the experts is consistent with the legal text of the Cosmetic Directive, whether the experts have properly adhered to these ToR and whether the resulting reports represent a complete and accurate analysis. In other words, has the question ‘what is the status of alternative (non-animal) methods for cosmetics testing?’ been
properly addressed and can we add any further information? The consultation is not on whether we support any extension to the 2013 deadline, which naturally, the ECEAE on ethical grounds does not.

In reviewing these expert reports we would like you to be aware of the four cross-cutting points. These are;

**1. The terms of reference given to the experts.**

As we wrote to the Commission at the time of call for experts (letter available on request), the terms of reference given to the experts should be consistent with the legal text of the Cosmetics Directive. What is crucial is that an evaluation is made of whether alternatives can be used for the purposes of the Cosmetics Directive, i.e. for regulatory purposes. It is not clear in the terms of reference given to the experts that this should be the basis of their reports. As a consequence it appears that the approach taken throughout the reports is to examine what is required to be scientifically understood before the entire mechanism of action of a particular toxic reaction is understood and modelled. This falls into the ‘high-fidelity fallacy’ trap; recognised by Russell and Burch in 1959 (also known as the ‘uncertainty paradox’ Schaafsma et al. 2009). This is the assumption that in vivo models are automatically superior models of the human response and that in order to be useful, non-animal methods must replicate the in vivo model in full. Not only does this ignore the fact that animal models may be poor predictors of human effects but it fails to recognise that not all aspects of the mechanism of action are needed to be covered by a model in order for it to be highly predictive (and therefore useful for regulatory purposes). If the requirement is to fully understand and replicate the in vivo response then this will obviously result in inordinately long timescales. However this is not the
correct or fair legal test. The test is whether alternative methods are sufficiently well developed and predictive of human responses to the same (or better) extent than animal models.


Assuming the purpose of the reports is to address “the current status of alternative (non-animal) methods for cosmetics testing”, we have reviewed each report along the following basis. We request that in your revision you also consider the following;

a. **neutrality** – is the report neutral in its tone and its evaluation of non-animal methods in comparison with the existing animal tests? In assessing adequacy do the reports expect the alternative methods to outperform the animal test where the adequacy of the animal test is known?

b. **completeness** – is the report complete and have all appropriate studies evaluating non-animal methods been considered? Studies that have been omitted or not fully discussed are included in specific comments.

c. **quantitative** - In assessing the regulatory suitability of these methods it is vital that the report covers the key aspects of this which are; the reliability of the method, accuracy (including concordance with in vivo or gold standard methods, sensitivity and specificity), applicability domain (based on the known mechanism of detection, range of substances used in any evaluation or known physical or biological limitations of the test) and, finally, availability of the method. The point at which methods would be in all likelihood considered
adequate by the regulators (i.e. the SCCS), if they applied the correct legal test as above, should be clearly given and methods rated against this. Whilst we are pleased to see a consistent approach to presentation of the methods in the report through the use of Table 1, this does not cover these important aspects nor in a quantitative manner. Authors are currently able to dismiss (in many cases) methods without apparent evaluation of the data. We suggest Table 1 is enlarged to cover these details or a separate table is produced to cover more details for those that are considered to have reached the stage of optimisation, pre-validation and validation and therefore are more likely to have this kind of information.

d. timescales – do the reports adequately tackle –as requested- the timescale to full replacement and if full replacement is not imminently possible then the steps needed to achieve this? Given the considerable public opposition to the use of animals to test cosmetics the experts should have aimed to keep the timescales short (but realistic) and provided a clear steer on what needs to be done to achieve these. It is our general impression that this aspect has not been adequately answered by the experts.

e. applicability – do the reports properly address the adequacy of alternatives in relation to cosmetics (as opposed to other substances)? i.e. are all specifics relating to the cosmetic use of these substances addressed including opportunities for waving animal tests due to (low) exposure (including the TTC concept) and clarity over whether animal and non-animal tests have been properly assessed as to their applicability to cosmetics?
3. Consistency across reports

Upon reviewing the reports it is clear that much could be gained from a consistent approach. We list some of the areas that should be considered.

- **Consistency in headings** and numbering of headings- this will help ensure all reports cover all relevant aspects. Examples of good practice include the sections on Inventroy of Non animal Methods in Chapter 3.

- **Limitations of in vivo studies** –this is well covered by some reports (e.g. Chapter 1) but not at all in others (e.g. Chapter 5, Chapter 3)

- **Company strategies**- Chapter 1 appears to have consulted with companies on their approach but this has not been attempted by all (e.g. not in Chapter 5). This is a good approach that can help not only assess the impact of the deadline on development (i.e. is the test really needed?) but also the strategy that they currently employ in instances where it is.

- **Consideration of regulatory requirement** (see Point 2). The extent to which the endpoint is required by the SCCS is an important point when evaluating the impact of deadlines and the need to replace. This was well covered by Chapters 3 (carcinogenicity) and 5 (reproductive toxicity) but not in Chapter 1 (repeat dose toxicity)

- It would be reader friendly to keep the **format of referencing** consistent across the reports
• **Methods to be included in Table 1** – Chapter 1 includes QSAR models whereas Chapter 5 does not.

• **Summary of ITS studies** and their outcomes and potential - this is covered in some Chapters, albeit not completely such as Chapter 1, but is not mentioned in others such as Chapter 2. This is an important tool that can lead to replacement if not reduction of animal testing and must not be overlooked.

• The **TTC concept**, this important exposure-related concept that can avoid animal testing is covered well in Chapter 1, better in Chapter 3 but not at all in Chapter 2 or 5.

4. **Skin sensitisation and Carcinogenicity.**

Finally, as we wrote to the Commission at the time of call for experts (letter available on request), we wish to remind you that there is no legal basis for any legislative proposal to extend the deadlines with respect to the endpoints Skin Sensitisation and Carcinogenicity. The possibility to extend the 2013 deadline in the text of the Cosmetic Directive (now a Regulation) only applies to Repeat Dose, Toxicokinetics and Reproductive Toxicity and does not include skin sensitisation or carcinogenicity. We will not go into our arguments but the legal text is clear. Subsequent assumption, even over the years on the part of the Commission, that the term ‘repeat dose’ includes these endpoints is not only inconsistent with all other EU regulations relating to toxicity requirements, internally inconsistent (by the Commission’s approach reproductive toxicity should be included under repeat dose), inconsistent with the opinion of experts here (see definition of repeat dose in Chapter 1) but also does not supersede the legal text. Authors of these two reports need not adjust their opinions as to timescale, as
this is a technical and not legislative question, but should bear this point in mind.

To allow a correct and balanced evaluation as well as for consistency between different chapters of consultation, scientific relevance and purpose, field of application and limitations, shall be given for each animal and non-animal test method. For non-animal methods information on status of validation and/or standardization together with ongoing developments shall be also given.

5 The need for an overall testing strategy

A general chapter is missing summarizing how to integrate the individual tests in a testing strategy covering all of the five health effects. This would have been important because it might not be necessary to test cosmetic ingredients on some of these endpoints if they do not show considerable oral intake or a significant systemic absorption for example. It could have been expected that the five expert groups would come up with proposals on the minimum testing requirements and how to fulfill them with in vitro methods. This would be similar to the strategies within REACH and new Biocides Directive, where endpoints or tests within endpoints can be waived based on physicochemical properties, exposure, previously conducted tests or by ordering the tests so as to avoid some tests later on depending on the result. This has been mentioned in some chapters such as reproductive toxicity and carcinogenicity where the link with results from repeat dose studies is made. See also Schaafsma et al. 2009.

Russell W. and Burch R (1959) The principles of humane experimental

2 50-52 We disagree with this statement. Based on the text within the chapter itself this is too negative. Many studies show extremely high concordance with in vivo data, in the realm of 80% accuracy which is considered sufficient for ECVAM validation purposes (see ECVAM 2009). With this in mind it is really not clear why the authors conclude we do not yet have alternatives for this endpoint.

Several non-animal methods are being developed and validated for hazard identification to support hazard classification and labeling, it will not be long before they will also be acceptable however data from these test methods alone will not be sufficient for risk assessment decision-making.

3 Table 1 We disagree with the timescales for the steps to ENTER pre-validation, since many of these tests have entered pre-validation it is not correct to include date ranges (without explanation).

Since Table 1 is very similar to Table 2 consider deletion of Table 1.

See Comment 39

4 72-90 The meaning of the expected ‘partial’ replacement provided under lines 88-90 is not clear.
As is stated here, there is no need to have a full characterization of skin sensitisation mechanism for risk assessment decisions.
Since the ToR for this report is the replacement of animals for the purposes of the Cosmetics Directive, i.e. regulatory testing then replacement should be considered ‘complete’ when RA decisions can be made for all ingredients.

The possibility that by 2013 the partial replacement from non testing methods will fulfill the requirements for risk assessment should be considered.

5 110-118 It would be useful to have experts’ opinion on the relative predictivity

Give a predictivity value (based on the results
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>of each step involved in the skin sensitisation mechanism. Where possible this should be based on the results of in vivo-in vitro comparisons for methods that model each step (see comments for these steps below)</td>
<td>of in vivo-in vitro studies where these have been conducted) to each step involved in skin sensitization mechanism.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 180-184</td>
<td>On 22 July 2010 the rLLNA was included in the updated OECD TG 429 and the non-radioactive modifications of the LLNA was accepted by the OECD as new Test Guidelines (OECD TG 442A, OECD TG 442B).</td>
<td>The rLLNA has been included in the updated OECD TG 429, adopted 22 July 2010. Two new test guidelines for the non-radioactive modifications of the LLNA – the LLNA: BrdU-ELISA assay (OECD TG 442B) and the LLNA: DA method (OECD TG 442A) have been adopted by the OECD on 22 July 2010.</td>
<td></td>
</tr>
<tr>
<td>7 191-193</td>
<td>No evidence is given to suggest that the GPMT is reliable. Please provide or delete.</td>
<td>For decades the standard guinea pig tests (Buehler test, GPMT) developed for the identification of skin sensitisers have been used as reliable predictors of allergic hazard.</td>
<td></td>
</tr>
<tr>
<td>8 216</td>
<td>It is important that these in vivo methods are assessed as to their reliability, applicability domain and validity, otherwise there is an unfair comparison for the subsequent in vitro methods. Please provide this data, including for the guinea pig tests. For example data on guinea-pig sensitization procedures from e.g. Goodwin et al., 1981, Andersen et al., 1995, and on LLNA from e.g. Gerberick et al., 1992, 2000 and 2005, Schneider et al, 2004 (for further references see IPCS, 2008). It shall be also highlighted that the guinea-pig tests have never been formally validated and limited data sets have been published for these tests (IPCS, 2008). The LLNA has been formally validated for hazard identification for regulatory purposes by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM, 1999).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 225-228</td>
<td>Please briefly summarise the findings of the predictivity of the LLNA and the chemicals assessed. This information shall be provided for the guinea pig tests as well.</td>
<td>The LLNA was compared to the guinea pig assays in terms of specificity,</td>
<td></td>
</tr>
</tbody>
</table>
sensitivity, positive predictivity and negative predictivity by a peer review panel, assembled by ICCVAM, and the results are available on EU web site (Opinion on the Murine Local Lymph Node Assay (LLNA) adopted by the SCCNFP during the 12th plenary meeting of 3 May 2000’,
In order to make the comparison, the guinea pig assay would have to undergo the same rigorous evaluation as the LLNA, however, such an evaluation has not been performed. Therefore it was decided to compare the performance of LLNA to available guinea pig data, and also including available sources of human data that were viewed as the "gold standard".
The comparisons between GPMT/BA versus human maximization test (57 comparisons) gave a positive predictivity on 100% and a negative predictivity of 16%, and an accuracy of 72%.
The comparison between LLNA versus human maximization test (74 comparisons) gave a positive predictivity of 96%, negative predictivity on 17% and accuracy on 72%. The comparison between LLNA versus GPMT/BA (97 comparisons) gave a positive predictivity on 93%, and negative predictivity on 80% and accuracy on 89%.
The LLNA resulted to be predictive for strong and moderate human contact allergens as with the guinea pig test, with risk of both false positive and false negative results.
The local lymph node assay resulted deficient in detecting sensitisation by low grade to moderate contact sensitisers, some metals, and organometal compounds and it was less sensitive compared to the GPMT with types of agents as metals, benzocaine, 4-chloroaniline, neomycin sulfate, streptomycin sulfate, and sulfanilic acid.
The peer review panel was also concerned about some strong irritants like SLS giving false positive results and recommended not use animal assays for contact sensitisation as stand-alone.
Performance standards for murine LLNA are also given by ECVAM
Please list of what these uncertainty factors are as they add to the information on the validity and accuracy of the in vivo models.

Not only is this statement in conflict with other statements (e.g. 85-88; 284-287, 291-301, 626-628) but, for completeness alternative hypotheses must be provided. The report underplays the importance of the covalent binding step in the determination of skin sensitisation (as measured in the LLNA). Too much emphasis is placed on the need to understand the additional mechanistic steps that lead to skin sensitisation when the literature evidence clearly shows that if we can predict the level of haptenation then we can predict potency in the LLNA. The other downstream biological steps do not affect potency. There are numerous papers by Roberts, Aptula, Patlewicz that demonstrate this. For example Roberts and Patlewicz (2010) argue that haptenation (the reaction with protein, step 1) is the “single most important and possibly the only important step” in the prediction of skin sensitization. See also Alenius et al. (2008) and Roberts and Aptula (2008) [ref 41 of chapter 2]. The premise that haptenation is the key step to model skin and respiratory sensitisation has been used also under Enoch et al. (2008a and 2010). These two articles, in conjunction with the extensive work of Roberts, Aptula and Patlewicz, help to strengthen the argument that the formation of the covalent adduct (haptenation) is the key step that lead to sensitisation (either in the skin or the lung).

In addition, the need for knowledge of portioning is mis-leading. Evidence from SAR studies shows that it is only important for certain mechanistic domains. For example, several papers have shown that the Michael domain does not require a logP term – it is modelled well by reactivity alone (e.g. Roberts and Natsch, 2009).

The report needs to discuss the fact the all we need to model sensitisation is data about the haptenation step. These data are as
follows and can be gathered from chemistry experiments:

1. Rate data for the formation of chemical-adducts. Model nucleophiles such as cysteine and lysine are sufficient. Complex peptides are not really required.

2. Hydrophobicity measurements / calculation for chemicals. Important for the harder domains where the nucleophiles are membrane bound.

3. Information about relative oxidation / metabolisms rates. This data is more difficult to gather. This data is important because this step might become the rate limiting step in the adduct formation.

<table>
<thead>
<tr>
<th>Page</th>
<th>Start Page</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>247-251</td>
<td>It should be clarified that the measure of skin sensitizer potency yielded by non-animal tools will allow a Quantitative Risk Assessment to determine whether a skin sensitizing chemical can be used safely.</td>
</tr>
<tr>
<td>13</td>
<td>255</td>
<td>To give a complete overview of non-animal tools suitable for cosmetic risk assessment, and for consistence between different chapters (see e.g. Chapter 1, Repeated Dose Toxicity, line 630), reference shall be made to the Threshold of Toxicological Concern concept (TTC) as a pragmatic risk assessment tool that establishes a human exposure threshold for chemicals without specific toxicity data (Kroes and Kozianowski, 2002). The use of the TTC in safety evaluation of cosmetic ingredients has been reviewed by Kroes et al. (2007) and the SCCS (Scientific Committee on Consumer Safety) is currently reviewing the use of the TTC for cosmetic ingredients. The outcome of this review may have implications for use of the TTC in quantitative risk assessment for cosmetics. Reference shall be also made to the Threshold of Sensitization Concern (TSC) concept recently proposed by Keller et al. (2009). In order to establish the concept with human data, a meta-analysis on fragrance ingredients from the IFRA/RIFM dataset accounting for interindividual variability and different exposure conditions has been conducted. TSC values of 0.91 or 0.30 μg/cm² were derived in terms of amount per... allowing for a Quantitative Risk Assessment.</td>
</tr>
</tbody>
</table>

... The Threshold of Toxicological Concern concept (TTC) is a pragmatic risk assessment tool that establishes a human exposure threshold for chemicals without specific toxicity data (Kroes and Kozianowski, 2002). The use of the TTC in safety evaluation of cosmetic ingredients has been reviewed by Kroes et al. (2007) and the SCCS (Scientific Committee on Consumer Safety) is currently reviewing the use of the TTC for cosmetic ingredients. The outcome of this review may have implications for use of the TTC in quantitative risk assessment for cosmetics. Recently a Threshold of Sensitization Concern (TSC) concept based on an evaluation of animal data, in analogy to the Threshold of Toxicological Concern concept, has been...
skin area. TSC values are compared with typical exposure levels of cosmetic products. A substance can be considered to be virtually safe if the quotient of exposure level and TSC is <1. The findings derived from human data include several conservative assumptions and largely support the dermal sensitization thresholds previously derived from animal data.

| 14 | 259-262 | Please insert references supporting similarity/differences on how chemicals are metabolized and penetrate through the skin between human and guinea pig/mice |
| 15 | 265-267 | The assumption that the use of in vitro methods will be associated with markedly lower bioavailability and metabolic considerations is not well explained. The models for in vitro skin metabolism may not yet have been evaluated. For cosmetics risk assessment purposes it should be considered that dermal absorption data come from an in vitro test (in vitro dermal absorption, OECD 428, adopted February 2004). This statement should be taken in context with the species differences from in vivo models for ADME. |
The text of the guideline stating that:
“However, it is reasonable to anticipate that the use of in vitro methods will be associated with a markedly lower, or even absent, impact of bioavailability and metabolic considerations on the prediction. A recent review examined this topic in detail, without identifying how the problem could be resolved (34). “- implicitly assumes that this is a “problem”.
We must not forget that the goal of risk assessment is to assess the risk when very large numbers of humans are exposed. Since very few people have a perfectly intact skin barrier all the time, the impact of “bioavailability considerations” will be to underestimate the risk (Roberts and Patlewicz, 2010). Consequently non-animal methods that do not include a model for bioavailability are likely to give a better estimate of risk to human exposed populations.
It shall also be considered that different opinions exist on the need to assess penetration potential and include it in an integrated approach since, with the exception of a few specific special cases, the ability to penetrate can be taken for granted and treated as independent of the identity of the chemical (Roberts and Aptula, 2008).

| 16 | 284-287 | Indeed chemical reactivity may be the key determining step (Roberts and Patlewicz 2010) | In brief, the ability of a chemical (either as a direct acting or after autoxidation/metabolism) to react covalently with a carrier protein is a major determinant factor in its ability to act as a skin sensitizer (40, 41), indeed this may be the key determining step (Roberts and Patlewicz 2010) |
| 17 | 280-301 | It is not clear why this section is separated from 1.5.4 reactivity assays as it is largely the same thing. Perhaps make the connection clearer. |
| 18 | 280-301 | Please summarise the key pre-validation/validation studies – either here or in section 1.5.4, including predictivity and applicability domains. |
| 19 | 280-301 | Please include reference to the results from INCHEMICOTOX project, The ability to make predictions on the skin |
focused on skin sensitisation potential and in chemico reactivity, (funded by the United Kingdom Department for Environment, Food and Rural Affairs).
See comments to lines 245-247 for further details on in chemico reactivity.

sensitisation potential based on in chemico reactivity has been recently evaluated under the UK Defra LINK project (funded by the United Kingdom Department for Environment, Food and Rural Affairs). QSARs, Quality assured databases together with Integrated testing strategies decision tools for skin sensitisation have been developed and are available under http://www.inchemicotox.org/

Reference shall be made to the work of Roberts and Aptula (2008) in the use of mechanistic domains. It is only within these domains that one can use simple and interpretable descriptors (logP and rate data) to model the formation of the haptan and thus, in turn skin sensitisation.

A mechanistically based paper that makes use of an in silico descriptor that is useful in modelling reactivity (and thus the LLNA) within the Michael domain is given under Enoch et al (2008a). The same descriptor has also been used to model respiratory sensitisation based on the same premise that haptentation is the key step that needs to be understood (the rest of the biology does not affect the sensitisation outcome) (Enoch et al, 2010).

Additionally, the report shall quote the QSAR model Toxtree which can be used to predict potential skin sensitisation mechanisms (i.e. if a chemical will be a Michael acceptor, Schiff base former etc) based on the Enoch encoding (Enoch et al., 2008b) of the Roberts rules for reaction mechanistic domains (Aptula and Roberts, 2006). The rule base is intended to allow a user to group chemicals into a mechanistic category. The data for the chemicals within the category can then be used in trend analysis / read across to fill any skin sensitisation data gaps within the category.

Relevant work on the use of mechanistic domains and of descriptors as logP and rate data to model the formation of the haptan and thus, in turn skin sensitisation, has been done by e.g. Roberts and Aptula (2008). A mechanistically based paper that makes use of an in silico descriptor useful in modelling reactivity (and thus the LLNA) within the Michael domain is given under Enoch et al (2008a). The same descriptor has also been used to model respiratory sensitisation based on the same premise that haptentation is the key step that needs to be understood (Enoch et al, 2010).

The QSAR model Toxtree allows to predict the potential skin sensitisation mechanisms (i.e. if a chemical will be a Michael acceptor, Schiff base former etc) based on the Enoch encoding (Enoch et al., 2008b) of the Roberts rules for reaction mechanistic domains (Aptula and Roberts, 2006). The rule base is intended to allow a user to group chemicals into a mechanistic category. The data for the
The report also lacks a detailed discussion of a number of expert systems – specifically Derek for Windows which has an extensive rule base able to identify skin sensitisers. The rule base within Derek for Windows is mechanistically based taking the premise that haptenation is the key event that leads to skin sensitisation. Such systems are of great benefit in supporting other predictions (for example predictions made by read across using the category approach of The (Q)SAR application Toolbox). The use of multiple in silico tools can lead to weight of evidence approaches for the prediction of skin sensitisation. A detailed discussion of such approaches is crucial and shall be added to the report.

Finally, reference shall be made to statistically-based models developed within the FP6 project CAESAR for skin sensitization and to the possibility to use them for regulatory purposes (Chaudhry et al., 2010).

Several QSARs for skin sensitisation are commercially available. For e.g. Derek for Windows has an extensive rule base able to identify skin sensitisers. The rule base within Derek for Windows is mechanistically based taking the premise that haptenation is the key event that leads to skin sensitisation. Such systems are of great benefit in supporting other predictions (for example predictions made by read across using the category approach of The (Q)SAR application Toolbox). The use of multiple in silico tools in a weight of evidence approach can be of great help for the prediction of skin sensitisation.

Statistically-based models for skin sensitisation have been developed within EU-funded CAESAR project (http://www.caesar-project.eu), implemented into open-source software and made available for online use via the web.

These models have been developed and
tested under stringent quality criteria to fulfill the principles laid down by the OECD and the final models, accessible from CAESAR website, offer a robust and reliable method of assessing skin sensitisation for regulatory use (Chaudhry et al., 2010). A new version of the CAESAR models (CAESAR v. 2.0 will be made freely available soon), will include some new features to obtain more reliable predictions and will allow to assess the applicability domain (AD), through quantitative and visual ways.

<p>| 21 | 327-358 | See Comment 18 please insert details of key studies of these reactivity assays - including predictivity and applicability domains. |
| 22 | 377 | Is the observed concentration-dependent increase of intracellular IL-18 an indication of the potential ability of the assay to estimate potency? Details from a recent publication by Mitjans et al, 2010 (developed under Sens-it-iv project) on the possibility to use IL-8 production and p38 MAPK activation to classify allergens according to their potency shall be considered. According to the authors, a significant correlation between IL-8 release and the LLNA EC(3) was found (Pearson correlation r=0.743, p=0.0036, n=12). On the contrary, the activation of p38 MAPK showed no significant correlation between LLNA data and vigor of p38 MAPK activation. Overall, data presented reveal IL-8 as potential tool not only to identify sensitizers, with the exception of pro-haptens, but also to classify them according to their potency, while p38 MAPK activation allows the identification of all sensitizers, including pro-haptens, but was not useful for potency classification. |
| 23 | 382-385 | Are there any results available from Sens-it-iv yet? If yes, please provide them. |</p>
<table>
<thead>
<tr>
<th>Page</th>
<th>393-397</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>This is the kind of information that should be provided for all studies on all assays. The current status of inter-laboratory evaluation to assess its transferability and reproducibility in discriminating between sensitising and non-sensitising chemicals should be also given.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>402-404</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Please give the results of the study. For example, the method and processing algorithm were developed from a training set of 39 chemicals, including three cationic metals, chromium (Cr), nickel (Ni), and silver (Ag). A predicted toxicity index (PTI) was determined for each test chemical. Since direct extrapolation from PTI to LLNA was difficult due to the large variation in LLNA data for compounds in the same potency category, 58 additional compounds were submitted in a blinded manner. Accuracy was approximately 84%, with a sensitivity of 81% and a specificity of 92%. The model correctly identified 2 of 3 cationic metals as positive (McKim et al., 2010).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>424-426</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>Please provide a summary of the results from these studies so the reader can evaluate the predictivity of this strategy. There is also no section specifically on ITS (Cf other chapters and Comment 1).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page</th>
<th>467-471</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>A 84% correlation between h-CLAT results with LLNA results has been observed (Ashikaga et al., 2010). One hundred chemicals were selected according to their LLNA categorization as being extreme, strong, moderate and weak sensitisers and non-sensitisers. The h-CLAT could positively predict not only extreme and strong sensitisers, but also moderate and weak sensitisers, though the detection rates of weak and non-sensitisers were low. The accuracy of the test was improved by using the combination of CD86 and CD54. The h-CLAT is expected to be a useful in vitro method for predicting skin sensitization potential since a 84% correlation between h-CLAT results with LLNA results has been recently observed when testing one hundred chemicals selected according to their LLNA categorization as being extreme, strong, moderate and weak sensitisers and non-</td>
</tr>
</tbody>
</table>
Therefore the h-CLAT is expected to be a useful in vitro method for predicting skin sensitization potential and future improvements shall focus on limitations with water insoluble chemicals, metabolic activation and sensitivity.

The h-CLAT could positively predict not only extreme and strong sensitisers, but also moderate and weak sensitisers, though the detection rates of weak and non-sensitisers were low. The accuracy of the test was improved by using the combination of CD86 and CD54. Future improvements shall focus on limitations with water insoluble chemicals, metabolic activation and sensitivity.

<table>
<thead>
<tr>
<th>Page</th>
<th>References/Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>480-483 This is the kind of information that should be provided for all studies on all assays.</td>
</tr>
<tr>
<td>29</td>
<td>490-498 According to jaCVAM and ECVAM updates (see below for links) the hCLAT using THP-1 cell lines is under validation (Validation study beginning 2009; JaCVAM lead, ECVAM collaboration) <a href="http://ntp.niehs.nih.gov/Ntp/about_NTP/SACATM/2009/June/Presentations/JaCVAM_Update.pdf">http://ntp.niehs.nih.gov/Ntp/about_NTP/SACATM/2009/June/Presentations/JaCVAM_Update.pdf</a> <a href="http://www.alttox.org/spotlight/040.html">http://www.alttox.org/spotlight/040.html</a> Please add this information both in the text and under Tables 1 and 2.</td>
</tr>
<tr>
<td>30</td>
<td>521-523 According to Wanner et al. (2010), the results from LCSA assay were to a large extent analogous to results from LLNA (analogous grouping of allergens into categories like weak-moderate-strong). This could be regarded as an indication of the potential of LCSA to assess potency and further details (quantitative data) of the results shall be provided. It is important to note that the LCSA has the improved capacity to distinguish sensitizers from non-sensitizers and irritants, which is another example where non-animal models outperform in vivo models.</td>
</tr>
<tr>
<td>31</td>
<td>538-539 Please provide a summary of the results. For example, When tested on a panel of five sensitizers and three non-sensitizers sensitisers (Ashikaga et al., 2010). The h-CLAT using THP-1 cell lines is under validation (Validation study beginning 2009; JaCVAM lead, ECVAM collaboration).</td>
</tr>
</tbody>
</table>
the novel MUTZ-LC, based on the in vitro migratory behaviour of LC, and the analysis of CXCL8 secretion proved to be more successful than the analysis of CD86 in predicting sensitizers from non-sensitizers and warrant further investigation (Ouwehand et al., 2010).

<table>
<thead>
<tr>
<th>32</th>
<th>557-560</th>
<th>Please provide a summary of the results.</th>
</tr>
</thead>
</table>
| 33 | 570 | There is no section devoted to Integrated Testing Strategies (ITS). Integrated Testing Strategies for skin sensitisation developed under OSIRIS project (EU Sixth Framework Programme), should be quoted. Data from the review of ITS for Skin sensitisation provided in the JRC Scientific and Technical report by Patlewicz and Worth (2008), and prepared as a contribution to OSIRIS project, shall be given or the document quoted. Several literature-reported ITSs for skin sensitisation are reviewed here. For example an ITS for skin sensitisation customised by Grindon et al. (2006) within a research project sponsored by Defra on the status of alternative approaches to animal testing and the development of integrated testing strategies for assessing the toxicity of chemicals under REACH, focused on regulatory hazard assessment. A non-animal strategy exploiting QSAR approaches which involve applying mechanistic principles is also presented (Roberts and Aptula, 2007). The authors describe the determinants of skin sensitisation potential from a chemical perspective arguing that the rate limiting step involves the covalent binding of a chemical electrophile to a protein nucleophile. Such electrophiles can be classified into a limited number of reaction mechanistic domains within which QMM (Quantitative Mechanistic Model) may be derived using the RAI (Relative Alkylation Index) approach which related reactivity and hydrophobicity to sensitisation potential. The development of Integrated Testing Strategies (ITS) for human health end points, including skin sensitization, is one of the focus of OSIRIS project included under the Sixth Framework Programme funded by European Commission. In addition, an important objective of the OSIRIS project is to develop a generic strategy for ITS including quantitative estimates of certainty. A review of ITS for Skin sensitisation is provided in the JRC Scientific and Technical report by Patlewicz and Worth (2008), prepared as a contribution to OSIRIS project. Several literature-reported ITSs for skin sensitisation are reviewed here as for e.g. the ITS for skin sensitisation customised by Grindon et al. (2006) highlighting as, before conducting new in vivo test, in vitro and in silico data shall be considered. A non-animal strategy exploiting QSAR approaches which involve applying mechanistic principles is also presented (Roberts and Aptula, 2007). The authors describe the determinants of skin sensitisation potential from a chemical perspective arguing that the rate limiting step involves the covalent binding of a chemical electrophile to a protein nucleophile. Such electrophiles can be classified into a limited
<table>
<thead>
<tr>
<th>Page</th>
<th>Lines</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>575-581</td>
<td>This paragraph is very negative and not correct if the proceeding information is to be taken into account. Not only is there evidence that some of these in vitro assays can provide potency information but potency is not available from the guinea pig assay. For risk assessment purposes the approach used with animal testing methods not giving information on potency (GPMT and Buehler test) could be used with in vitro methods and should be highlighted here. E.g. physicochemical data and chemical reactivity, comparison of estimated potency with general use conditions, comparison with well characterized benchmark allergens and the use of a safety factor. Combined use of in vitro, in silico and in chemico to examine whether a response may be a false positive.</td>
</tr>
<tr>
<td>35</td>
<td>601-604</td>
<td>This statement is confused and at odds with statements elsewhere (e.g. 85-88; 626-628). If single methods are adequate for risk assessment purposes then this endpoint, with respect to regulatory purposes, can be considered replaced.</td>
</tr>
<tr>
<td>36</td>
<td>613-615</td>
<td>This comment implies that in vitro methods are not used by companies to assess the hazard of chemicals. It would be more complete to ask companies if this is the case, see Chapter 1 on repeat dose where they ask companies for their strategies.</td>
</tr>
<tr>
<td>37</td>
<td>628-631</td>
<td>It is not clear how the 2017-2019 timescale has been decided upon. We are of the opinion that based on the haptenation step alone we are extremely close to ‘replacement’ of skin sensitization for regulatory purposes (i.e. predictivity of human skin allergy) Given that haptenation is the key determining step in skin sensitization, and in chemico methods are already in use for this, it can be considered that we are very close to regulatory replacement of skin sensitization.</td>
</tr>
</tbody>
</table>


Gerberick GF, House RV, Fletcher ER, Ryan


CAESAR - Computer Assisted Evaluation of chemical Substances According to Regulations. [http://www.caesar-project.eu/].


Enoch S J., M T. Cronin, T.W. Schultz, and J C.
Madden. (2008a) Quantitative and mechanistic read across for predicting skin sensitisation potential of alkenes acting via Michael addition. Chemical Research in Toxicology, 21, p513-520


OSIRIS - Optimized Strategies for Risk Assessment of Industrial Chemicals through
Integration of Non-Test and Test Information, EU Integrated Research Project, Sixth Framework program, Contract no. GOCE-CT-2007-037017 (http://www.osiris-reach.eu/).


Please complete, to the best of your ability, based on published studies, the fields ‘Areas of application’ and ‘Comments’.

Please also include whether the method can assess potency, actual or hypothetically.

Please include more quantitative analysis of the methods, where available (see Comment 1).
Please include more details about the status of the method, when it entered pre-validation, the name of the method, etc.

These suggestions are given to:
a). allow a proper evaluation of the status of alternatives for skin sensitisation (availability, pro/cons) and
b). for consistency with the same Table in Chapter 3, Carcinogenicity and Chapter 4 Toxicokinetics, which is more complete.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>We disagree with the ‘estimated time to have the method ready for entering pre-validation’, since several methods have already entered this stage. According to ECVAM guidance; “a Prevalidation study: is a small-scale inter-laboratory study, carried out to ensure that the protocol of a test method is sufficiently optimised and standardised for inclusion in a formal validation study and to obtain a preliminary assessment of its relevance and reliability (Curren et al. 1995). The prevalidation process is divided into three consecutive phases namely, protocol refinement (Phase I), protocol transfer (Phase II), and protocol performance (Phase III). It involves the testing of a limited number of coded substances in at least three laboratories.” For clarity we suggest the date of entry is given, as Table 1.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide reactivity: 2009 (or 0 years) Keratinocyte cultures: 2009 (or 0 years) Keratinocyte: dendritic cells co-cultures: Optimisation R&amp;D 3D reconstituted skin models: 2009 (or 0 years) 3D reconstituted skin models + dendritic cells: Dendritic cells activation: 2009 (or 0 years) (hCLAT using THP-1 cells validation study beginning 2009).</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>The criteria used to define the status of a method as R&amp;D, Optimised, under pre-validation, etc, as given in Table 1 of Chapter 5, must be provided for clarity.</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>40</td>
</tr>
</tbody>
</table>