t⁴ Report

An Expert Consortium Review of the EC-commissioned Report "Alternative (Non-Animal) Methods for Cosmetics Testing: Current Status and Future Prospects – 2010"*

Thomas Hartung [1], Bas J. Blaauboer [2], Sieto Bosgra [3], Edward Carney [4], Joachim Coenen [5], Rory B. Conolly [6], Emanuela Corsini [7], Sidney Green [8], Elaine M. Faustman [9], Anthony Gaspari [10], Makoto Hayashi [11], A. Wallace Hayes [12], Jan G. Hengstler [13], Lisbeth E. Knudsen [14], Thomas B. Knudsen [15], James M. McKim [16], Walter Pfaller [17], and Erwin L. Roggen [18]

[1]Johns Hopkins University, Bloomberg School of Public Health, CAAT, Baltimore, and University of Konstanz, CAAT-Europe, Germany; [2]Doerenkamp-Zbinden Chair on "Alternatives to Animal Testing in Toxicological Risk Assessment", Institute for Risk Assessment Sciences, Division of Toxicology, Utrecht University, The Netherlands; [3]TNO, Pharmacokinetics, Utrecht, The Netherlands; [4]The Dow Chemical Company, Predictive Toxicology, Midland, MI, USA; [5]Merck Serono, Animal Science & Welfare, R&D Quality Assurance Darmstadt, Germany; [6]U.S. Environmental Protection Agency, Integrated Systems Toxicology Division, Research Triangle Park, NC, USA; [7]School of Pharmacy at the Università degli Studi di Milano, Toxicology, Italy; [8]Howard University College of Medicine, Department of Pharmacology, Washington, DC, USA; [9]University of Washington, Department of Environmental and Occupational Health Sciences, Institute for Risk Analysis and Risk Communication, Seattle, WA, USA; [10]University of Maryland, School of Medicine, Department of Dermatology, Baltimore, MD, USA; [11]Biosafety Research Center, Foods, Drugs and Pesticides, Shizuoka, Japan; [12]Harvard School of Public Health, Boston, MA, USA; [13]Leibniz Research Centre (IfADo) at the TU Dortmund, Germany; [14]University of Copenhagen, Faculty of Health Sciences, Department of Public Health, Copenhagen, Denmark; [15]U.S. Environmental Protection Agency, National Center for Computational Toxicology, Research Triangle Park, NC, USA; [16]CeeTox, Inc., Kalamazoo, MI, USA; [17]Innsbruck Medical University, Dept. of Physiology and Medical Physics, Division Physiology, Innsbruck, Austria; [18]Novozymes A/S, Bagsvaerd, Denmark

Summary

The European cosmetics legislation foresees a review in 2011 and possible postponement of the 2013 marketing ban to enforce the testing ban for systemic and repeated-dose animal tests. For this purpose, a 119-page report commissioned by the European Commission was published recently. Here, a group of 17 independent experts from the US, Europe, and Japan was brought together to evaluate the report. The expert panel strongly endorsed the report and its conclusions. A number of important options not considered were identified; these do not, however, affect the overall conclusions regarding the current lack of availability of a full replacement, especially for the areas of repeated dose toxicity, carcinogenicity testing, and reproductive toxicity, though a roadmap for change is emerging. However, some of these options may provide adequate data for replacement of some animal studies in the near future pending validation. Various recommendations expand the original report. The reviewers agree with the report that there is greater promise in the short term for the areas of sensitization and toxicokinetics. Additional opportunities lie in more global collaborations and the inclusion of other industry sectors.

Keywords: alternatives to animal testing, 3Rs, toxicology, novel approaches, product safety

1 Introduction

The 7th amendment (European Commission, 2003) of the cosmetics legislation (Cosmetics Directive 76/768/EEC) provided *inter alia* for more detailed provisions on the phasing out of animal testing. This recently was transformed into a regulation (European Commission, 2009), which did not, however, affect these provisions: A general testing ban on cosmetic ingredients from 11 March 2009, reinforced for 10 animal test requirements by an instant marketing ban, as well as a marketing ban for the more complex endpoints (those requiring repeated substance application, e.g., repeated dose toxicity, sensitization, reproductive toxicity, carcinogenicity and toxicokinetics), to go into effect 11 March 2013. It is important to note here that the legislation provides for a review in 2011 of the feasibility of the 2013 deadline, and it can be further postponed in a co-decision procedure.

To support this process, a large EU expert consortium recently reviewed the current state of the art concerning alternative methods for cosmetics testing (Adler et al., 2011). The review includes five toxicological key areas: toxicokinetics, skin sensitization, repeated dose toxicity, carcinogenicity, and reproductive toxicity. Based on a detailed analysis of the state of the art, the consortium estimates that full replacement of animal testing is possible for aspects of skin sensitization and toxicokinetics in approximately five to nine years, but more than 10 years will be required for the others. These time frames do not include formal validation and acceptance, i.e., a multi-year process.

The 7th amendment and the resulting engagement of the cosmetic industry has become an engine for progress in the field of alternative methods over the last few years. However, this has a potential influence on other economic areas in the world and on other industrial fields as well. As one of our US reviewers noted: *"The EU Cosmetics Directive ban on animal testing for cosmetic ingredients is of major significance for chemical safety testing, as it fundamentally challenges animal-based testing practices, which have been in place for decades. The impact of the EU Cosmetics Directive is already being felt in sectors well beyond cosmetics, with the 2013 proposed ban looming large across the industry and regulatory agencies. On a sociopolitical level, the Directive represents a clash between society's wish for an end to animal testing on products such as cosmetics and the current state of science surrounding alternatives to animal testing."*

In a similar way, the status of alternatives had been assessed, under the lead of ECVAM, after the 7th amendment was issued (Eskes and Zuang, 2005). Such stocktaking can and must also be the starting point for the initiation or reshaping of activities (Hartung et al., 2003; Zuang and Hartung, 2005; Hartung, 2008). The Centers for Alternative to Animal Testing (CAAT) on both sides of the Atlantic hope that this will result in a new roadmap for scientific activities to further the replacement of animal testing. As a first step, the expert report was subjected to an independent review, presented here. The selection of experts aimed to include non-Europeans while avoiding those who work extensively with the cosmetic industry or who were involved in the generation of the report[1]. The expansion of views was meant to broaden and consolidate the basis for planning for future research opportunities. It is important to note that this is not a consensus report but reflects a spectrum of individual views and recommendations. The individual reviewers do not represent any institutions for the purpose of this review. The originators of the different comments are not identified, because each and every contribution was part of a general brainstorming process that lay the ground for further steps. Note that CAAT staff members were not involved in these comments, other than moderating the process and compiling the individual reviews. Further steps to promote a roadmap for research into alternatives for systemic endpoints will include strategic workshops based on currently commissioned white papers.

2 General assessment of the report

The general judgment on the report was very positive, as shown by some pertinent remarks:

– *"The overall conclusions reflect the scientific truth of the matter regarding the current state of non-animal alternatives to animal testing."*

– *"This status/prospect paper related to alternative (non animal) methods for cosmetics testing, which I find very useful, thorough, and well documented..."*

– *"This reviewer agrees with the overall conclusions of the report… In summary, this is a fantastic 'state of the science' report that clearly indicates the need for further research in almost all of the areas reviewed. Some additional credit for strong and promising efforts in the specific areas towards reduction of animal use should be emphasized and applauded."*

– *"The report is well written, timely, and covers, for the most part, the current status of the potentially available alternative methods for cosmetic testing. The individual experts involved for repeated dose toxicity, carcinogenicity, and reproductive toxicity had the necessary experience and training to evaluate the various alternative methods against the* in vivo *methods currently employed for most risk assessments. The experts, rightly, point out that no replacement alternatives are currently available to replace the test now used for repeated dose toxicity, carcinogenicity, or reproductive evaluations."*

– *"It is clear that Adler and co-authors were well aware of the importance of their assignment and conducted their work with diligence. At 119 pages in length, the report was undoubtedly thorough, not to mention objective and scientifically rigorous. The credibility of the report is very high, given that the co-authors were well-respected experts, many of whom devoted their careers to advancement of alternatives. The authors also clearly understood the regulatory process, guideline testing, and applications of the data for classification and*

---

labeling, and risk assessment. This latter point is critical as 'alternatives' are not just about individual methods but about testing strategy and end uses of the data as well."

– *"Thus, I agree with the conclusions given in the report… In conclusion I find this extensive report comprehensive, updated, and with well justified conclusions related to replacement strategies."*

– *"Overall, the analysis is realistic, concrete, and comprehensive, including all relevant techniques – briefly, a must-read for everyone interested in this field of research."*

– *"This 119 pages document provides an exhaustive overview on the alternative methods available to potentially replace animal testing in safety assessments of cosmetics, drugs, and chemicals. In this respect, the report is very well suitable to serve as a reference and compendium for industry, scientists, politicians, animal welfare groups, and consumer associations."*

– *"Overall, the list of alternative methods evaluated seems quite complete, and the quality of the evaluation is very high in all areas reviewed. However, the individual areas had some freedom in the design of their respective contributions, which makes it difficult to compare status quo, achievements over the last 5 years, gaps, and future needs and strategies."*

The reviewers unanimously agree with the overall conclusion that, currently, no replacement approaches are available for most areas:

– *"Consequently, the first part of this assessment should deal with the main conclusion, which is that for most (almost all) of the areas evaluated, a meaningful safety assessment with complete replacement of the animal studies is not possible within the envisioned time frame. Only one area (skin sensitization) could be considered as an exception. However, it may be possible to provide preliminary information that can be used to focus on the most likely hazards and provide a strategy for an optimized risk assessment, avoiding animal testing wherever possible."*

– *"It is still difficult to predict when alternative methods can fully replace animal testing for the toxicological endpoints of repeated dose toxicity, carcinogenicity, and reproductive toxicity – a courageous statement considering the pressure politics have exerted on scientists."*

– *"The strength of the review by Adler et al. (2011) is that it gives a factual analysis about where we really stand, summarizing the most promising techniques available and describing how such techniques could be combined to allow quantitative in vitro prediction of in vivo toxicity. The consortium correctly concludes that further progress in basic sciences is needed before this goal can be accomplished. For example, there is a need to better understand how mechanisms of toxicity (or biomarkers of activated/perturbed 'toxicity pathways') are linked to adverse effects."*

– *"Perhaps even more difficult will be the replacement of repeated dose and carcinogenicity testing. It certainly will not be possible in the foreseeable future to establish in vitro systems that cover all possible organ-specific mechanisms, as well as their interactions, and to test these with the help of automated high throughput systems. Therefore, the consortium is correct in their statement that the time horizon of this long-term goal is currently difficult to estimate. In my opinion, this will be work we pass on to the next two or three generations of scientists, considering that we must first understand the relevant pathways of all target cells of toxicity in an organism and translate this in vitro. Nevertheless, it is only a matter of time until this goal is achieved. Understanding the functioning of the human organism will always be a central object of human curiosity, and the benefits of in vitro systems correctly recapitulating molecular mechanisms of human cell types will go far beyond the replacement of animal experiments."*

The reviewers also recognized attempts in the report to go beyond stocktaking and make further suggestions to improve the approach:

– *"In addition, for some areas, e.g., repeated dose toxicity, reproductive toxicity, the authors present and discuss new and innovative testing strategies and overall approaches to generate high quality safety assessments for new compounds and products that expand and extend the original approach to simply replace animal studies. This promising strategy can be applied to all areas discussed in this report and, in a second step, could also be used to combine the information generated in the individual approaches, identify synergies, and provide an overall approach (as an example, bioavailability was reported crucial in many areas of this report)."*

– *"The estimation of time frames for the individual areas could benefit from the definition of specific milestones and prerequisites on which these estimates are performed. The general time estimates should be made based on harmonized assumptions in terms of success rate, availability of resources, etc. Also, a table listing these harmonized milestones and estimated over all areas could be of added value to allow for an easy comparison of what has been achieved and the work ahead."*

– *"The authors should be encouraged to evaluate the possibility of identifying the potential for synergies and developing one strategy that, ideally, can provide all respective areas with the information needed in an efficient way. The information accumulated (e.g., biokinetics, QSAR, in vitro methods) in combination with new strategies (TTC approach, sensitivity, susceptibility, and predictivity of certain methods, educated statistic-based hazard and risk assessments) could be used to develop one overall strategy (decision tree) that provides guidance for the steps necessary to generate high quality and efficient safety assessment focusing on the relevant elements while avoiding unnecessary animal studies and work."*

– *"To emphasize that only an in silico approach can direct the assessment of human hazard/risk."*

– *"However, the full replacement of repeated dose, reproductive (and neuro) toxic, as well as carcinogenetic studies in live non-human organisms for assessing the risk in humans is still to be developed; this will include the new technologies of metabonomics and toxicogenomics, as well as concise human exposure considerations."*

– *"I suggest further development of the discussion regarding inclusion of human data on exposure, toxicokinetics, and risk in the future strategy."*
– *"This reviewer felt that the discussions regarding the Threshold of Toxicological Concern added considerably to the report because of the importance of exposure in the risk assessment process. Use of toxicokinetics and other methods to rule out systemic effects (no absorption via the skin) again speaks to the issue of exposure (internal dose)."*

A clearly identified need is to move now to a forward-thinking strategy and roadmap:
– *"In all areas evaluated, the statement recurs that currently available alternative methods and strategies are not sufficient for a complete replacement of all animal studies and that additional investigations and time are required to achieve the final goal. However, it is not clear what exactly these next steps could be, which specific areas and questions need to be addressed, and which milestones are proposed to reach this goal."*
– *"As we consider new paradigms, we must also be willing to relinquish some long-held default practices, such as testing at high doses that far exceed human exposures, and testing for endpoints that rarely, if ever, drive risk assessment or classification and labeling. Similarly, we can't continue to add more and more apical endpoints of reproductive function to the test guidelines just because a chemical might be capable of affecting that endpoint at some dose. Testing strategies of the future should focus on distinguishing between safe and potentially unsafe exposures, rather than on characterizing all possible hazards, even those that only occur at very high doses. A more selective approach to choice of endpoints is warranted."*
– *"Funding and additional legislation will be needed to encourage industry to move forward. The 7th amendment moved the cosmetic industry – maybe a similar approach is needed to encourage big pharma and the chemical industry. REACH seems to be a problem."*
– *"As a last thought, in my opinion, a global perspective should be considered, beyond the European area, including a strategy addressing how the European view could be combined with other strategies, particularly from the US (Tox21)."*
– *"Without encouragement from governments, the process will continue to move slowly. It seems to this reviewer that greater cooperation between the US and the EU would help move the process forward."*
– *"In redesigning an integrated system for the 21st century, a systems biology approach that cuts across existing toxicology disciplines would seem the most promising, as key regulatory pathways tend to be involved in a wide range of toxicities, such as development, cancer, and organ toxicity. This allows us to probe the root causes of toxicity, rather than the apical or descriptive end. With computational tools, we now have the ability to achieve systems-level integration, as well as to anchor new approaches to existing animal-based toxicity data, which are housed in some new publically available databases (e.g., US EPA's ToxRefDB). There are several new assay platforms that are amenable to a systems- or pathways-based approach. These range from panels of human cells in culture to model organisms such as zebrafish embryo and the roundworm, C. elegans."*

3 Executive Summary

Table I aims to summarize the 29 comments received on the five toxicological areas from 17 reviewers. The first author (TH) summarized the comments received on a scale from (---, i.e. very strong disagreement) to (+++, i.e. very strong agreement) based on personal judgment only.

4 Toxicokinetics

4.1 Overall assessment of the adequacy of this status report

The overall assessment is very positive, e.g.:
– *"The section is in general comprehensive, well written, and gives account of all the most relevant methods and developments."*
– *"This is a very comprehensive overview of the needs, possibilities, and limitations of kinetic considerations in the context of non-animal studies in the risk assessment of cosmetics."*
– *"The report does a comprehensive job of considering factors that determine the ADME of cosmetics ingredients (and of potentially toxic chemicals in general)."*
– *"This is a very useful and complete review on the state of art for the area of biokinetics and the possibilities for using non-animal (i.e. in silico and in vitro) information in this field for a proper risk assessment. The conclusion that kinetic information is a (or THE) essential element in the interpretation of in vitro-derived toxicodynamic data is fully supported. It also stresses the fact that biokinetics is not an 'endpoint'; it is an indispensable tool in integrated strategies for risk assessment."*
– *"This reviewer was extraordinarily pleased to see the prominence and importance that this section had in the overall review. This reviewer is wholeheartedly supportive of this visibility and the thoroughness that this review gave to this area. This section was very well done. This reviewer especially liked the details on what kinetic analysis can do and how it informs risk assessment. Of particular importance were the decision trees (Fig. 4, 5) included in this section, as well as the explanation of how kinetic information would fit within a tiered analysis (Fig. 7)."*
– *"The paper rightly mentions the absolute necessity of the use of biokinetic considerations in the interpretation of in vitro toxicodynamic data."*
– *"The knowledge gaps are accurately established."*

Some more critical remarks:
– *"Again, parallel to efforts to find alternatives to fill these gaps, their importance and the uncertainty in the existing methods should be addressed."*
– *"A number of not-very-well-supported statements are made about the maturity of in silico methods that can or might be*

Tab. I: Executive summary of the 29 comments received on the five toxicological areas from 17 reviewers

	Toxicokinetics	Sensitization	Repeated dose toxicity	Carcinogenicity	Reproductive Toxicity
Number of reviews	*6*	*5*	*6*	*6*	*6*
Overall assessment	*++*	*++*	*++*	*+++*	*+++*
Time lines suggested in report	*5-7 years*	*7-9 years*	*10+ years (not clearly stated)*	*10+ years*	*10+ years*
Time line according to reviewer	*1 more optimistic, 1 more pessimistic*	*1 more optimistic, 2 more pessimistic*	*2 more optimistic, 1 more pessimistic*	*1 more pessimistic*	
Specific comments (selected)	*Sensitivity analysis; TTC important; bioavailability cannot be excluded*	*Questioned potency issue; need for validation*	*Not focused on cosmetics; kidney more important*	*TTC very valuable; more on QSAR*	*Extended 1-generation study important*
Omitted methods	*Dynamic models; variation of bioavailability; virtual tissues; plasma protein binding in vitro; bioaccessibility models; placental models; human microdosing; internal TTC; NTP data*	*Transgenic mice; mouse ear swelling assay; tissue from ACD patients; NKT-cells, NK cells and mast cells; KeratinoSens; SenCeeTox; NCTC2544/IL-18 test; GARD; DotScan; MUSST; h-CLAT; DC migration; VITOSENS; photo-sensitization; bio-availability; myelotoxicity; fluorescent cell chip; human whole blood cell culture; lymphocyte proli-feration; human T cell activation assay; immunotoxicoge-nomics; Mishell-Dutton assay*	*Dog MDCK cell line; GRB-MAL1; Mdct; immortalised glo-merular mesangial and epithelial cells; liver cell co-cultures (with endothelial cells, monocytes) and 3-dimensional cultures; intestinal epithelia; brain slices; CeeTox Panel; Hurel organ specific toxicity panels; cardiomyocytes derived from stem cells; more omics technologies*	*Weight of evidence approach; progress carcinoGENOMICS; toxicogenomics; categorical approach based activity (AAR)*	*Cross-references to the 'repeated dose'; neuro endpoints; daphnia, sea urchin and C. elegans; 3-D germ cell assays; placental transport; WEC and other assays using transcriptomics; Toxicity Testing in the 21st Century approaches; newer ToxCast studies*
Agreement with report	*+++*	*+*	*++*	*+++*	*+++*
Further advice	*Filling data gaps; generic model framework; Tox-21c*	*Develop integrated testing strategies; main challenge mixtures; immuno-toxicity important; discuss potency issue; in vitro kinetics; use without validation; lack of forward-thinking*	*No concepts, ideas and suggestions on future strategies; preparation of an overall approach; integration with biokinetics and internal exposure; realistic expectations of contribution from in vitro; encouragement from governments needed; co-operation between the US and the EU*	*Mechanistic understanding; lessons learnt from the Sensitization area*	*More visionary evaluation of potential testing strategies; a rethinking of testing strategies; improved human exposure models; testing strategies which are fit for 21st century purposes; a systems- or pathways-based approach*

able to replace in vivo *animal-based approaches to obtaining needed data. Not addressed at all is the metric by which a procedure or tool is judged to be sufficiently mature, so the predictions of, for example, availability in 7 or 9 years come across as arm-waving."*

– *"A critical note on the report is related to the fact that it is not always necessary to 'know' all aspects of the kinetics to do a proper risk assessment. This is one of the strong points of PBBK modeling approaches: one can perform sensitivity analysis for the different parameters that need to be taken into account in the models. Sensitivity analysis not only (as mentioned in the report) is imperative for assessing the validity of the model; it also indicates the importance of parameters. If such an analysis shows that a certain parameter (e.g., the exact numbers for the renal clearance of the compound) is not a sensitive one, the need to dig deep into this issue does not have a high priority. Such approaches also increase the confidence one has in the model outcomes and decreases the need for having large uncertainty factors. This will improve the quality of a risk assessment."*

– *P. 392: "Sentence that says '…assuming that such models are available for the chemical class of interest.' This is a very big assumption and it's not at all clear when this capability will be available for most classes of chemicals. In the same paragraph, the statement that the PBTK model enables prediction of the most sensitive endpoint is problematic. By itself, toxicokinetics does not enable prediction of toxicity."*

4.2 Is the judgment on promise and timelines adequate, overly optimistic, overly pessimistic?

Altogether, the suggested timelines of 5 to 7 years are endorsed with slightly more pessimistic and more optimistic notions, e.g.:

– *"It is estimated that the development of alternatives for these parameters (namely renal and biliary clearance, pulmonary absorption) will take 5 to 7 years, and integration in PBTK model framework even longer. I agree with the first part, but I don't see why integration should take so much longer. The processes by which compounds are cleared or absorbed are known. They may be complex, but models can be and are being built while alternatives to provide the necessary input are being developed."*

– *"This reviewer was in concurrence with the review's statements that, in general, 5 to 7 years are needed before significant models are available for all routes of exposure (see comments on lung) and that for overall integrated systems more time will be needed. Since many of the* in vitro *models are not yet fully developed, kinetics for those systems can be 'premodeled,' although experience and detailed assessments will be needed under the standardized model protocols. Hence it is essential that these models develop in parallel with the* in vitro *systems and not just as separate exercises. It also implies that, as there are delays in the other methodologies, then parallel delays in the kinetics could occur."*

– *"A key insight of this section, with which this reviewer agrees, is that a future for toxicity testing based on data collection* in

vitro *will require and depend on a combination of* in vitro *and* in silico *tools for the characterization of toxicokinetics and, most importantly, that this capability will need additional time beyond 2013."*

– *"I am more optimistic than the authors of the report about the possibilities for introducing kinetics that are not based on animal studies into the process of risk assessment, particularly for the area of cosmetics."*

– *"The document as it stands might be well received by toxicologists who have a good understanding of the technical issues, but to the degree that the document will be read by regulators or EU rule-makers, it has some significant shortcomings. The document does not provide clear support for the claims that at least some components of what will be needed will be available in seven years, and there is real potential for setting up expectations that will not be fulfilled. It needs to be emphasized, though, that there should be a more focused discussion of the criteria that will be used to determine when the needed capability actually does exist. I would expect this discussion to include consideration of quantitative metrics."*

Specific comments:
PBTK model simulations

– *"As mentioned above, there are methods available, imperfect as they may be. Priority for improvement should be those parameters that not only have relevance and lack precise methods, but in addition model simulations show sensitivity towards that parameter."*

– *"As mentioned, development of modeling strategies and integration do not have to be on hold until* in vitro *methods become available. Good pulmonary models are available. Biliary efflux clearance can be measured* in vitro *in hepatocytes or transfected cell-lines, and scaled up to* in vivo*, correcting for transporter abundance. Building an accurate model of processes involved in renal clearance should not take as long as the development of the QSARs and/or* in vitro *methods needed to provide the input. The model performance can already be assessed with the existing methods (despite their limited quality) and with available animal data."*

– *"Another key area, also well covered in the report, is the emphasis on the role of specific barrier functions in the* in vivo *situation and the possibilities and limitations of the use of* in vitro *and* in silico *models to address these. The role of specific transporters and the way in which these can be incorporated in PBBK modeling needs to be stimulated also."*

– *"The chapter very strongly emphasizes the possibilities of stepwise approaches, as illustrated by the decision-trees presented in the different figures."*

– *"Toxicokinetics (TK) and the integrating potential of PBTK modeling are envisioned as a pivotal step in the alternative approach. Major gaps in TK input are identified as renal and biliary clearance and lung permeability. I agree with these statements, with the remark that it should be investigated how far off the current, imprecise* in vitro *models,* in silico *models, and default assumptions are:*

 - *Renal clearance: product of fraction unbound in blood and glomerular filtration rate.*

- *Biliary clearance: scaling of efflux clearance of the parent compound or major phase I metabolite, ignoring biliary clearance of their phase II conjugates. Further, if the conjugate is neither active nor deconjugated, it may be considered as cleared anyway, and efflux is of less importance.*
- *Pulmonary permeability: no barrier, instant equilibrium between blood and alveolar air.*

This could be done by retrospective analysis of available data.

This remark applies not only to these major information gaps but to the model input in general: how well do we do with the existing methods? It should also be borne in mind that in the current animal-based approach, rather large uncertainties exist in the extrapolation steps from animal to human, human intraspecies, and possibly time and route-to-route, which are all dealt with by conservative estimates. Many of these uncertainties are explicitly addressed (and thereby reduced) by the in vitro + *PBTK approach, but some* in vitro-in vivo *uncertainties are faced instead. These can in turn be dealt with in the same way: by conservative estimates."*

– *"How much do we gain in terms of precision with extra efforts in development of* in vitro/in silico *models? This question, preferably, should not be answered for each parameter but in systematic sensitivity analyses for model parameters in generic PBTK model simulations."*

– *"It is mentioned that TK is indispensible to assessing whether the compound will be bioavailable, so that toxicity testing can be waived for non-available compounds. In a similar effort in the EU FP7 project ChemScreen, focused on REACH chemicals, it was concluded that waiving tests on the basis of bioavailability probably will not significantly reduce the number of compounds to be tested. Zero absorption is hard to prove, both by QSPR methods and* in vitro *assays, so the absorbed fraction must be compared to a toxicological reference, using PBTK and possibly internal TTC concentrations."*

– *P. 391: "Statement 'When* in vivo *studies cannot be performed or when inadequate* in vivo *data are available, the toxicokinetics of a substance can be predicted on the basis of* in vitro *or in silico studies.' This just isn't accurate. This is a research goal, not a current reality. In the same paragraph, last sentence saying that (PBTK) models '…should be able to significantly reduce or even replace animals in many research and toxicity studies' is also inaccurate. The sentence appears to confuse toxicokinetics with toxicity. Also, it's not at all clear what 'should' means in this context."*

– *"An accurate account is given of the methods and capabilities of PBTK models. However, in the general description the issue of determining the appropriate dose-metric is somewhat underestimated. Often, the exposure scenario is a given, and thus so is the concentration-time profile that will result from this exposure. It is not always possible to mimic that profile in* in vitro *studies. A good understanding of the dependence of the effects, not only on the concentration but also on the concentration-time profile, is needed in interpreting the* in vivo *relevance of findings in* in vitro *toxicodynamic studies. This relationship deserves more attention, preferably both from toxicologists involved in TK and in TD."*

– *"The report provides many examples in one section that are relevant for other areas of discussion in the report. This is done with little redundancy and with significant details. However, there are a few examples where some more cross-referencing is needed. For example, in the section on kinetics there is a discussion of kinetic models for women (see Figure 6) and a discussion on specific barriers such as the blood-placenta barrier. There is limited discussion, however, not just on transport across the placenta but on the need for determining and sometimes modeling the embryo-fetal components. This is especially relevant when this chapter is considered alongside the chapter on reproduction. Additional details and cross referencing is needed with sufficient detail to understand why a model with a temporal component would be needed across gestation in order to understand potential for exposure as well as determination of differential metabolism by both the mother and fetus, as well as differential ADME overall. There is a base of models for this, but there is a need for additional application and appreciation within the risk assessment frameworks presented. Greater use of such knowledge could significantly affect the number and types of tests that are needed for extrapolation."*

The TTC concept
– *"The emphasis on the concept of internal TTC, especially in the context of cosmetic risk evaluation, is also key to the possibilities of avoiding ALL testing, also* in vitro, *and therefore is a tool in priority setting."*

– *"The TTC concept could be an important lower tier in alternative risk assessment. Currently, the TTC is mainly based on oral toxicity studies. If no route-specific (inhalatory or dermal) TTC is available, the external TTC should be converted to internal TTC: without account of the oral bioavailability and target organ concentrations, it cannot be inferred what dermal or inhalatory doses could be considered 'of no concern.' With the proper attention and resources spent, I estimate that this matter could be solved within two years time."*

In vitro TK as a key for 1R
– *"Not much seems to be gained from bioavailability estimates as exclusion criteria. Zero bioavailability is hard to establish and will not apply to many compounds, at least in REACH. However, bioavailability is a crucial parameter in combination with the internal TTC concept."*

Importance of analytical methods
– *"It is stated that a sensitive, specific and validated analytical and quantitative method is a pre-requisite. However, one could question what an acceptable degree of uncertainty is in view of total uncertainties in the* in vitro-in silico *approach. The same applies to analytical methods to determine physico-chemical properties as input to QSAR/QPPR. Uncertainties may be negligible in comparison with those in biological processes, and current methods may be sufficiently precise."*

Importance of actual rather than nominal concentrations
– *"I support the importance of the actual rather than nominal concentration in the test system."*

– *"It is stated that EU project PredictIV is the first attempt at combining toxicodynamics, TK, and modeling to link real exposure to effects. Although this project may take this a step further, the correction for free concentration between* in vitro *and* in vivo *was already addressed in ReProTect (Verwei et al., 2006)."*

– *"Your points are well taken and accurate. A couple of comments. Protein binding, while important, is not nearly as important as Kd, or off kinetics of a molecule. This is absolutely true in the pharma industry where a protein-associated drug with high Kds creates a kinetic situation that enhances solubility and releases drug as equilibrium demands. Drugs with high affinity or low Kds are more affected at concentrations below saturation.*

I have not seen a discussion of steady-state conditions. Typically, compounds dosed repeatedly reach steady state, and this generally can happen in just a few days. How do the assumptions put forth change if steady state conditions prevail? Under these conditions, many of the tenets put forth may have less impact."

Strategic considerations

– *"It is proposed here that the validation of PBTK modeling be done in a stepwise approach where, in pre-screening, validation of model prediction is done case-by-case instead of striving for high precision in each input parameter. Method validation follows in a higher tier. This seems like a very sensible strategy."*

Absorption and bioavailability

– *"The prediction of absorption by* in vitro/ex vivo *studies (under OECD TG 428) gives reasonably accurate results. Dermal bioavailability could be somewhat more problematic, as CYP activity, in general, drops in these systems (while phase II enzymes remain active). Focus should be on methods to scale up skin metabolic activity in the same way as is done for hepatic clearance."*

– *"There is a lack of* in vitro *models or QSPRs for pulmonary permeability. It is stated that years of intensive research are required. Be that as it may, it would also be worthwhile to assess by retrospective analysis the predictive value of the default assumption of instant equilibrium between alveolar air and blood. This could be a reasonable worst case."*

– P. 386-387: *"The discussion of 'Inhalatory exposure' (should be Inhalation exposure) is not accurate. The statement that 'In the upper respiratory airways, the absorption is low, and it mostly occurs in the lower part' is just plain wrong. Think of formaldehyde. Regional dosimetry in the respiratory tract is reasonably well understood, and sophisticated models are available."*

– *"Dermal absorption may be estimated by* in vitro *tests, but ways to estimate dermal bioavailability (dermal metabolism) should be further developed."*

– *"It is stated that a particular challenge in modeling itself is a description of dermal exposure. Increasingly complex physical and physiological models emerge to describe dermal absorption, but it seems reasonable to ask whether sufficient approximations can be made by compartmental models."*

Distribution

– *"Several efforts are being undertaken to estimate active transport. The strategy of measuring transport in transfected cell-lines or membrane vesicles (K_m and V_{max}) and scaling up the activity (V_{max}) from* in vitro *to* in vivo *by correcting for the difference in transporter abundance seems most promising. Prof. Sugiyama's lab is advanced in this work, as are several labs in Europe (including TNO). Efforts are underway to determine the actual abundance of the most prominent transporter proteins in various tissue membranes (as poor correlations are observed at mRNA expression level). This strategy could, in principle, be applied to any tissue and could lead to good estimates of various barriers. A particular challenge is the scaling up of transporter activity at the BBB, as the blood brain barrier is hard to sample representatively."*

– P. 387: *"The statement that 'The fate of a compound in the body is determined by partitioning into human tissues' is a big over-simplification."*

Excretion

– *"As mentioned, while alternatives for renal and biliary excretion are not in advanced stages of development, the issue should be addressed in terms of how well current methods perform in relation to the overall sensitivity of the model to these parameters. In case of renal clearance, the importance may be foreseen, as the default assumption of glomerular filtration without active resorption or secretion may give poor predictions and is not worst-case. In case of biliary clearance, measuring efflux transporter activity towards the parent compound (or its major phase I metabolite)* in vitro *and scaling up to whole liver may be sufficiently accurate."*

Metabolism

– *"Methods to predict hepatic metabolic clearance by determining intrinsic clearance in hepatocytes or hepatocyte-like cell-lines are in a well advanced stage and seem to give reasonable estimates. Extrahepatic metabolism could, in principle, be addressed in the same way, although good cell-lines expressing the relevant enzymes may not (yet) be available. CYP activity in ex vivo material in general seems to decline rather rapidly."*

– *"Even in cryopreserved hepatocytes or HepaRG cell lines, the activity of CYP enzymes may differ from* in vivo. In vitro-in vivo *scaling factors were traditionally determined in animal studies. Microdosing studies may provide the same information in humans. This strategy could be very useful to establish scaling factors and to validate not only* in vitro *assays to predict metabolism but a range of* in vitro *TK parameters, including bioavailability and renal and biliary clearance."*

– P. 392, first paragraph: *"The model depicted in Fig. 6 assumes that the liver is the only metabolizing organ and that excretion only happens in the kidney. It's difficult for this reviewer to be comfortable with the idea that any model de-*

veloped for generic application in the context of cosmetics ADME would be limited in this way."

– *"It is stated that ideal* in vitro *methods are metabolically competent. This may in practice be qualitative: identifying the major metabolites and pathways involved. But more effort may be needed to quantitatively scale the* in vitro *metabolic activity to* in vivo. *This applies in principle to all metabolically active tissues."*

Non-validated human *in vivo* approaches

– *"Microdosing may not be sufficient in a 1R approach unless the TTC is accepted as a basis for first-into-man studies. But as stated, it can be very useful in establishing* in vitro-in vivo *scaling factors for various enzyme and transporter activities and in validating* in vitro/in silico *and PBTK approaches."*

4.3 Are there promising methods and approaches that were omitted?

Quite a few aspects are suggested for additional consideration:
– *"As a great supporter of physiologically based kinetic and dynamic models, this reviewer was a bit dismayed that more emphasis was not placed on the essentiality of modeling the dynamics as well as the kinetics of chemical exposure. The chapter did discuss this aspect in the latter portion of the discussion, however. Given the fact that the chapter is called kinetics and that the executive summary is silent on dynamics, this reviewer would strongly support adding this concept to the title and throughout the chapter in a more prominent manner. It is the opinion of this reviewer that these models will be as essential as the kinetic models for extrapolation and interpretation of* in vitro *observations for humans."*

– *"The mentioning of the importance that kinetic and dynamic modeling will go hand in hand and that such models can be integrated is also very well covered; however, some key literature on the application of these methods is missing. There are more examples of this approach, especially in the field of pharma development that could very well be used in the area of cosmetics too. This field of research, however, needs to be stimulated further. The possibilities of implementing this in risk assessment are increasing, and we need to take advantage of these developments more than is now possible. A good example, also mentioned in the report, is the Predict-iv program in FP7."*

– *"This chapter, as well as other sections of the report, refer to exposure and emphasize the importance of calculating potential for bioavailability through the use of kinetics for specific exposure routes. This data would then be used to assess potential for toxicity from the* in vitro *alternative models using a "threshold of toxicological concern." However, it was remarkable that no discussion of variations in kinetics was included, especially given the application for cosmetics. Somewhere the report should acknowledge that skin absorption is not the only concern here; determining testing for potential for skin abrasion, use on broken skin (i.e. diaper rash or sunburn) or occlusion (use of cosmetics under gloves or masks,*

etc) should all be a significant part of the kinetic modeling as well. This could be done by using Bayesian based models for probability of skin condition and linking this with more standard kinetic models for absorption across normal intact skin. Adding a section acknowledging this additional need for modeling would be important. Human information could be used to supplement these assessments."

– *"There will be a need to use PBPK-type models to bridge from exposure estimates to concentrations at the level of the target tissue and even at the level of cells within tissues. This latter capability is what is really needed, and that will depend on the development of virtual tissues. While development of some virtual tissues is underway, having these available for routine use will require years of dedicated research."*

– *"The* in silico *approaches to estimate tissue:blood partitioning give reasonable results. An updated algorithm has been published that unifies the earlier algorithms of Poulin and Theil and Schmitt mentioned here (Peyret et al., 2010)."*

– *"Plasma protein binding can be measured* in vitro *by several methods. Some* in silico *approaches are mentioned, but these seem obsolete, with the accurate and easy to standardize* in vitro *approaches now available."*

– *"Various* in vitro *systems are used for estimation of oral bioaccessible fraction, ranging from simple dissolution to full physiology-mimicking systems. Subsequent intestinal absorption/metabolism is usually measured in monolayer cultures such as Caco-2 or* ex vivo *material. The combination of advanced* in vitro *bioaccessibility models and absorption assays gives promising results: TNO's intestinal model (TIM) has successfully replaced dog studies in early R&D at large pharmaceutical companies."*

– *"Development of* in vitro *placental models is not mentioned here. This still seems challenging. Useful alternatives could be* ex vivo *placental material, and transporter assays in combination with transporter abundance measurements."*

– *"I miss discussion of data available from volunteer human studies as e.g., microdosing in drug testing for PBTK."*

– *"I miss the internal TTC as a concept of big potential in these recommendations."*

– *"There is a large effort underway at the NTP to increase our knowledge about these complex chemicals and their related mixtures. The report could refer to this developing database."*

4.4 Do you agree with the conclusions for the area? If not, where do you disagree?

In general, the document was endorsed:
– *"I agree with the main conclusion of the working group."*
– *"The framework of* in vitro/in silico *predictions in combination with PBTK modeling that is proposed seems the most promising way forward."*
– *"The kinetic section specifically indicates needs for determining kinetics for plant-derived products that can be used in cosmetics and indicates that international dialogue is underway for this need."*

– *"I agree to the pivotal role of TK and PBTK modeling in the 1R approach for cosmetics…The crosstalk between toxicologists involved in TK and TD is indeed crucial, also to address the matter of concentration-time mentioned earlier."*

4.5 Further advice to the area

A number of recommendations add to the report:
– *"Challenges lie mainly in filling gaps or improving model input, rather than in modeling the relevant processes. Apart from meeting these challenges, which may indeed need some more years, sensitivity analyses of model parameters to guide these efforts are important, as is a discussion on acceptable model uncertainties. This may contribute to a different way of looking at validation of the alternative methods. As said previously, algorithms to describe processes for which alternative (*in vitro *or* in silico*) predictive methods are lacking or need improvement should not take (much) more time than the development of the alternatives themselves."*
– *"Specifically, a publicly available generic model framework would be beneficial."*
– *"Another prerequisite for a proper use of* in vitro *data is the emphasis on the use of good considerations of the actual exposure of* in vitro *biological systems (i.e. the cells), in other words the* in vitro *biokinetics."*
– *"This final comment is related to the idea of metrics for deciding when an alternative approach is sufficiently mature to be used. The toxicokinetics section as a whole seems to be based on the idea that* in vitro *and* in silico *methods will be developed (eventually) to the point where they are capable of providing datasets equivalent to those that can be obtained using* in vivo *methods. No consideration seems to have been given to whether or not this is really the best way to proceed. The US NAS document 'Toxicity Testing in the 21st Century (TT21),' for example, proposes an alternative approach in which perturbations of toxicity pathways* in vitro *are used in place of* in vivo *testing. This document has generated a lot of discussion, and it's not entirely clear exactly what its routine application would look like. Nevertheless, it seems that the challenge facing the EU given the impending deadline on cosmetics testing presents an opportunity for thinking about truly alternative approaches, possibly but not necessarily something along the lines of TT21, and not just a refined means of obtaining the same kinds of data we are used to obtaining."*
– *"I underline the statement that sensitivity/uncertainty analyses should guide the determination of which aspects of the model are critical."*
– *"A new way of approaching validation indeed seems needed."*
– *"Pharmaceutical industry shows an increasing willingness to share non-competitive data and to address non-competitive issues such as PK in consortia (e.g. IMI, Simcyp). Cosmetic industry should be able to benefit, or contribute."*
– *"International attention on nano engineered materials (NEMs) and the need to assess these materials using a new paradigm offers both opportunities and challenges, especially*

for the linkage of QSAR and kinetics with alternative models for assessment. This reviewer would suggest some cross-reference to the OECD efforts to coordinate testing and review of these materials. This might be a very fertile application of the kinetics and QSAR approaches outlined in this section. It is also of relevance due to the current use of NEMs in cosmetics and sun creams."

5 Sensitization

5.1 Overall assessment of the adequacy of this status report

The overall assessment is positive, e.g.:
– *"This status report gives an exhaustive overview of the alternative methods available."*
– *"The section on skin sensitization was a good overview that touched upon the major assays that are recognized as having potential for an* in vitro *test to replace or reduce the use of animals, or assays that are currently undergoing the process of validation. There was an emphasis on assays that are in an advanced enough state of development to be used by industry in a timeframe to meet a tentative deadline of 2017. The strengths of the report are that they recognized the major assays that have potential and categorized them into the context of the major pathways of the sensitization cascade."*
– *"An exhaustive review of the state of the art on alternative methods for hazard identification of allergens is presented."*
– *"The report gives a systematic overview of tests for hazard identification (sensitization) with a brief description of the underlying mechanisms addressed by each test. As the report is set up, it indicates that the available tests cover, to our current knowledge, the main steps in the mechanism of skin sensitization induction."*

More critical remarks:
– *"The individual approaches are nicely summarized and graded. However, it is not always clear what exactly the gaps are and what is needed to fill these gaps to reach the goal within the given time frame."*
– *"Not all works published have been mentioned."*
– *"While the report lists most of the more advanced approaches, describes correctly what the test is about and what it measures (read-out), and mentions the (pre-) validation status where relevant, little information is provided about the quality of performance of each of these tests, where available (e.g., chemicals tested, accuracy, limitations, applicability, reproducibility). In my opinion, this information is highly relevant for assessing whether or not a test is well established."*

5.2 Is the judgment on promise and timelines adequate, overly optimistic, overly pessimistic?

Assessments were somewhat mixed with regard to suggested timelines:
– *"I agree with the timelines."*

– *"Adequate timelines (see Table 1, page 398) have been estimated."*
– *"Overall, the segment on skin sensitization was optimistic in that it estimated the use of predictive testing without animal testing to be in place before 2017-19, based on the state of development of the current lead assays. This concept is based on the assumption that the currently developed assays will move forward. The emphasis of the article was related to advanced assays, their validation, and availability to meet the 2017-19 timeframe."*
– *"Currently there are four methods under validation at ECVAM. A hope would be that alternative methods able to discriminate between sensitizers and non-sensitizers might become available in the next couple of years. Therefore, hazard identification may be possible by 2013."*
– *"Overall, the tools listed in the report (supplemented with the tests missing…) should provide a good basis for hazard identification."*
– *"I do not agree with the idea that the existing toolbox of* in vitro *assays is complete enough to identify hazards without whole-animal testing by 2017-19. I do agree with their idea that the estimation of allergen potency is elusive at this time."*
– *"It is estimated, that it will take at least another 7 to 9 years for replacement of the current* in vivo *animal tests. Unfortunately, it is not clear how this is estimated and on which assumptions it is based."*

Specific comments:
– *"Reference is made to the importance of skin bioavailability, which could be a critical point in a risk-based decision making approach."*
– *"Table 1 (Toolbox for skin sensitization risk assessment: estimated timelines) is very helpful and instructive. It could serve as an example for the other areas, even though it is well understood that a clear solution may not be available for all methods and gaps."*
– *"It is correct that none of these tests has been formally prevalidated or validated, but most of them have been demonstrated to be reproducible (within as well as between laboratories) and transferable. It is generally accepted that these tests are physiologically relevant. In combination with the increasing numbers of chemicals being tested, this must build further on the scientific validation of these tools. The point I want to make is that, while indeed formal validation has not been performed on any of these tests, scientific validation should provide confidence in the capacity of the respective tests and an understanding of the limitations, which in turn should drive the different tests to incorporation in OECD guidelines."*
– *"To date, it is still not properly understood what, for example, the CYP profile of normal skin looks like* in vivo. *Consequently, the* in vitro *models may not be physiologically relevant. Nevertheless, keratinocyte (KC) as well as dendritic cell (DC) based tests can properly identify sensitizers, including pre- and pro-haptens. Furthermore, human reconstituted skin models were found capable of classifying chemicals, identified by other means as sensitizers (including pre- and pro-hap-*

tens), in the proper potency classes (92% concordance with the LLNA, see www.sens-it-iv.eu). The mechanism(s) behind this good performance remain(s) to be fully elucidated."
– *"Cell-based assays focusing on a single marker or pathway may misclassify (typically false-negatives) chemicals."*
– *"Potency information is lost in the currently used direct peptide reactivity assays."*
– *"I partially disagree on the potency issue… It is indisputable that the LLNA provides several advantages with regards to both scientific progress and animal welfare. It studies the induction phase of skin sensitization and provides data suitable for dose response assessment. It is important, however, to point out that LLNA EC3 values might be useful to protect individuals from induction of sensitization, but they have no value to protect already sensitized individuals. Furthermore, an evaluation conducted by ICCVAM on the possibility of using the LLNA as a stand-alone test method to determine potency categorization of allergens concluded that using EC3 ≤2%, a criterion recently adopted by the United Nations Globally Harmonized System of Classification and Labeling of Chemicals, correctly categorized 52% of the strong human sensitizers. Nearly half (48%) of the strong human sensitizers had an EC3 >2% or were negative in the LLNA. The LLNA can be used to categorize substances as strong sensitizers when EC3 ≤2%, but it cannot be used as a stand-alone assay to predict sensitization potency categories. It is important to consider that the GHS only foresees two categories for potency classification: strong sensitizers and sensitizers. Regulatory requirements and industry necessities are not always on the same track. We should therefore ask what we want from* in vitro *methods. Nevertheless, even if limited numbers of chemicals have been tested, excellent correlations between* in vivo *LLNA EC3 values and* in vitro *data have been found by several authors using different* in vitro *methods. Possibly, a deep and careful re-evaluation of the methods currently proposed for hazard identification for their ability also to classify chemicals according to potency may result in a full replacement of animal testing for sensitization within the same timelines foreseen for hazard identification."*
– *"Just to give a few examples: in the peptide binding assay, non-allergens and weak allergens generally demonstrated minimal to low peptide reactivity, whereas moderate to extremely potent allergens displayed moderate to high peptide reactivity. Classifying minimal reactivity as non-sensitizers and low, moderate, and high reactivity as sensitizers, it was determined that a model based on cysteine and lysine gave a prediction accuracy of 89% (Gerberick et al., 2007). Using one of the dendritic cell tests (h-CLAT) currently under validation at ECVAM, a first indication of potency may come, for example, from the concentration required to induce a threshold of positive response (CD86 ≥150) in the h-CLAT system. A good correlation (R = 0.839, p < 0.01) was indeed found between the h-CLAT thresholds and LLNA EC3 values (Sakaguchi et al., 2009). In our work using THP-1 and IL-8 release to identify sensitizers (Mitjans et al., 2010), the calculation of concentration of allergen that induced a release of IL-8 of 100 pg/ml by linear regression analysis of data*

showed a significant correlation with the EC3 LLNA data (R = 0.924, p= 0.0248). There is a general trend for IL-8 release at lower concentrations for strong sensitizers, whereas higher concentrations are needed in the case of weak sensitizers. Therefore, in vitro *methods already can provide useful information regarding potency. It is true, however, as the expert group concludes, that sensitizer potency currently is not possible with sufficient confidence to allow risk assessment decisions using only* in vitro *methods, but I am not convinced that animal models are there either. Also, with the LLNA an integrated decision strategy is necessary."*

– *"As reported, immunocompetent reconstituted human skin models have been developed that are useful for acquiring mechanistic understanding of the cellular processes and the impact of a chemical on these processes. These models are not yet mature enough to identify sensitizers accurately."*

– *P. 402, 3rd paragraph: "There has been some discussion on the role of systemic metabolism (e.g., liver) following initial exposure to an allergen. Absorption through skin activation in the liver is followed by haptenization and subsequent events. For prohaptens this could introduce the issue of sensitization reactions not related to skin alone. This could be especially important for compounds that are respiratory sensitizers, e.g., menadione. There also may be instances where systemic conversion of a prohapten results in dermal lesions."*

– *P. 402, 4th paragraph: "This information should be combined with information on rates of metabolism in the skin, preferably human skin. It is unlikely that the mouse ear has the same metabolic capacity as human skin. Auto induction should also be considered. The ability of a prohapten to induce its own metabolism could significantly increase the rate of electrophile formation. This in turn would change the potency category."*

– *P. 404, 6th paragraph: "HaCaT keratinocyte cell line (Natsch and Emter, 2008; Emter et al., 2010) – "The original model was a transfected liver cell line? Did this change?"*

– *P. 404, 7th paragraph: "to SenCeeTox add 'combines concentration response, cell viability, reactivity with GSH, and ARE controlled gene expression.' The human 3D models used are SkinEthic and MatTek. Are these proprietary? I guess the culture process may be, but it sounds as though the cells belong to SenCeeTox, and of course they do not. …This last sentence is very important. This is the only method that attempts to place chemicals into potency categories. We have had success placing chemicals into three categories (Extreme/strong, Moderate, Weak-non). The accuracy of this is about 70%."*

– *P. 404, 8th paragraph: "Stratatech skin cell model from the University of Wisconsin, Madison is a full thickness human skin model with both epidermis and dermis layers (see SOT poster 2011, M. Hoffmann[2]). They have communicated to me that dendritic cells are present."*

– *P. 405, 5th paragraph: "McKim et al. presented a comparison study at SOT 2011 in which the SenCeeTox method was evaluated in HaCaT, and RHE cells from both SkinEthic and MatTek. A manuscript has been written and submitted. … In* addition, we have discovered that considerable care must be exercised when using these models as they can vary greatly in their stress status when received by laboratories. For example, we have found that cells shipped from the EU to the USA undergo repeated high dose exposure to X-rays, which greatly increase the background oxidative stress markers and ROS. This data is also provided in the manuscript above. A standardization of the test systems should be implemented based on basic biochemistry to ensure that the test system used has the required dynamic range and correct sensitivity to reproducibly identify chemical sensitizers."*

– *P. 406, 3rd paragraph: "The test system is a co-culture, yet the endpoint that is used to base results on is CD86 expression. How is this different from the hCLAT? Of what value are the KCs in this method? Possibly more explanation?"*

– *"It should be recognized that there are certain limitations that may necessitate the use of TG 406 (guinea pig tests), e.g., test substance classes or substances containing functional groups shown to act as potential confounders. Leaving open the question of potency, as guinea pig methods have not been designed for potency classification. I totally agree with experts when they say nothing, including the LLNA, is perfect."*

5.3 Are there promising methods and approaches that were omitted?

– *"The list of methods is quite complete, even though some statements indicate that they may not be sufficient to allow for a complete replacement. More details on what is missing and what could be done to fill this gap might be helpful, for example: 'Several expert systems that claim to predict skin sensitization are commercially available, to mention some: … It must be kept in mind that none of the above mentioned approaches represents a complete replacement for current* in vivo *methods, nor has any undergone formal validation.'"*

– *"Transgenic mice, mouse ear swelling assays were not described. The emerging cell types being recognized in ACD (such as mast cells, NK, and NKT-cells, and innate molecules such as anti-microbial peptides, toll-like receptors) were ignored. Molecular signaling pathways were not described, and cellular and macroscopic assays (tissue-skin equivalent) were emphasized instead."*

– *"There was also an inventory of the currently available animal assays. Particular emphasis was placed on the local lymph node assay because of hazard identification, as well as the possibility of identifying potency. Although the mouse ear swelling assay is not utilized for hazard identification, it deserves mention and some description, because the mouse ear swelling assay is very closely related to the local lymph node assay. The mouse ear swelling assay is a widely utilized whole-animal test used in developing the immunology of allergic contact dermatitis. Additionally, numerous mice are used in the skin allergy research arena, and this deserves to be mentioned as another potential target for animal reduction. Along the lines of use of* in vivo *mouse models, the*

---

[2] http://www.stratatechcorp.com/download/Products/Allen-Hoffmann%20SOT%202011%20Poster.pdf (last accessed August 17, 2011).

potential for transgenic mice, which have transformed immunology research, should be considered either as a potential for refinement of the local lymph node assay, or to define other important pathways of allergic contact dermatitis that can be moved from an in vivo *pathway to a mechanistic* in vitro *pathway."*

– *"The article recommended against the use of* in vivo *predictive testing in humans (the human repeated insult patch test). While I agree with this statement, I am of the opinion that the whitepaper is missing an opportunity to recommend the use of* ex vivo *studies from humans with ACD (either skin biopsies from skin challenge sites, or peripheral blood) to define pathways of ACD. Thus, this manuscript misses opportunities to recommend* ex vivo *studies of patients with existing skin reactions to either common or rare contact allergens to define human pathways that could be utilized to develop predictive* in vitro *assays of human allergy. It is understandable that access to allergic patients may be limiting to industry segments, but it should be recognized that such studies represent a valuable opportunity. This deserves mention because this type of material is available from academic and contract research organizations."*

– *"This segment lacked the insight to address opportunities in the immunology arena related to the field of experimental contact dermatitis. For example, the role of innate immunity and cells of the innate immune system play a critical role in ACD, and assays based on cell recognition or signaling molecules should have been mentioned as emerging opportunities. Although dendritic cells (DC) were described in great detail because of the DC-based assays that utilize this cell type, other innate cells such as NKT-cells, NK cells, and mast cells deserve mention as potential targets of assays. The report also did little to emphasize the need for the development of molecular signaling pathways that could be developed for virtually all of the steps of the pathways involved in ACD (keratinocyte inflammation, DC maturation, and T-cell responses). The article emphasized the cellular assays (for example, exposure of keratinocytes to a chemical allergen, and subsequent inflammatory cytokine release). This cellular approach has the potential to be transformed into a molecular method that defines the pathways leading to inflammatory cytokine release."*

– *"The direct peptide reactivity assays classify chemicals correctly in 85% of the cases. Since peptide/protein reactivity may not be specific for sensitizers, subsequent testing is required. Physiologically relevant tests are available for assessing the impact of the chemical on keratinocytes (KeratinoSens, SenCeeTox and the NCTC2544/IL-18 test) as well as on DCs (GARD (DC activation), DotScan, MUSST and h-CLAT (DC maturation and DC migration)."*

– *"KeratinoSens (87% accuracy) and SenCeeTox (84% accuracy) assess the impact of chemicals on the oxidative stress response pathway (more specific ARE promoter) of human keratinocytes (HaCat cell line). A preliminary assessment suggests that about two out of three chemicals actually affect the KEAP-1/Nerf-2 pathway."*

– *"The NCTC2544 cell line-based test (98% accuracy on 25 compounds) measures the increase in intracellular IL-18 expression upon exposure to chemicals. The higher accuracy (as compared to KeratinoSens and SenCeeTox) might reflect the fact that this test measures a phenomenon that occurs downstream from KEAP-1/Nerf-2 activation, and therefore may reflect also the activation of elements other than KAEP-1/Nerf-2."*

– *"In vivo KC activation seems to go hand in hand with DC activation. Sens-it-iv has shown that KC and DC activation is required in order to get the inflammation that is required for sensitization to occur. Two rather advanced assays assessing DC activation can be put forward. The VITOSENS test uses 13 genes (of which two are essential) as predictive biomarkers following treatment of human CD34$^+$-progenitor cell derived DC (95% accuracy on 21 compounds). The obvious drawback of this test is the use of primary human cells. The Genomic Allergen Rapid Detection (GARD) test provides a high throughput test system discriminating skin sensitizers from respiratory sensitizers and irritants using a specific marker profile consisting of 11 genes. This test uses the MUTZ-3 cell line. The accuracy of this test is as high as 98% for skin sensitizers (80 compounds tested)."*

– *"There are at least three tests that can be used to assess the next step in sensitization: DC maturation. The most advanced tests are the MUSST (measuring CD86) and h-CLAT (measuring CD86 and CD54). The overall accuracy of these tests is around 85% (about 100 chemicals tested). A more flexible and high throughput approach for assessing DC maturation is the DotScan, which uses a chip coated with monoclonal antibodies directed against all the known cell surface markers. The accuracy of this approach for skin sensitizers is >90% (25 compounds tested)."*

– *"In vivo the next step is DC migration. The in vitro DC migration developed within Sens-it-iv is a functional test using, again, the MUTZ-3 cell line. So far this test has not produced misclassification of either skin or respiratory sensitizers (only 12 compounds tested)."*

– *"Finally, T cell responses are induced in vivo. During recent years our understanding of T cell stimulation and the regulation of T cell responses has expanded significantly. This has allowed the establishment of an in vitro T cell priming assay. This approach is promising but faces the drawbacks of complexity and donor-to-donor variation."*

– *"Thus far, the described tests contain only one cell type and (with the exception of the epidermal equivalent models for potency assessment) all represent the submerged cell format. This has the obvious consequence that the tests are limited to chemicals with sufficient solubility in an aqueous environment. In addition, the submerged single cell format does not allow for assessing the impact of chemicals on cell-cell interactions, which might be relevant in the context of sensitization induction."*

– *"In my opinion, two important tests are missing: i) The Genomic Allergen Rapid Detection (GARD) test for assessing xenobiotic sensing, and ii) the DotScan for assessing the im-*

pact of chemicals on DC maturation (both described above). Both tests have so far revealed a high accuracy and have a flexible high-throughput format.”

– *“There is evidence emerging from the FP6 project Sens-it-iv indicating that sensitizers of different potency stimulate different intracellular pathways. While the mechanism behind this differentiation is not understood, proteomic analysis seems to indicate that different proteins are modified. What are the potential consequences?”*

– *“I am surprised that no mention of photosensitization is included at all (with the exception of page 373). The transposition of the protocols developed to identify in vitro sensitizers to the identification of photoallergens is not as straightforward as one may think.”*

5.4 Do you agree with the conclusions for the area? If not, where do you disagree?

– *“The conclusion is clear and easy to understand: ‘in isolation, non-animal tests for hazard identification will not be sufficient to replace fully the need for animal testing for this endpoint, although they might reduce the overall need.’ Again, this area would benefit from more details (or a summary) on what exactly is missing, how these gaps can be bridged, what a tiered safety assessment strategy (decision tree) – beyond the one-to-one replacement of animal studies – might look like, and how this can be achieved by 2019.”*

– *“I agree with the idea that potency prediction is lacking without animal testing. I am not convinced that the existing panel of predictive tests will provide a robust toolbox for hazard identification to distinguish between sensitizers and non-sensitizers with in vitro tests. I also agree with the idea of the need for a toolbox approach, with multiple endpoints that address numerous steps in the sensitization cascade. It is optimistic, perhaps overly optimistic, to estimate that the elimination of animal testing and hazard identification can be completed in the absence of animal data by 2017-2019. It is my opinion that the existing toolbox needs to be expanded and refined to the molecular level. Much more work related to the molecular pathways of ACD is necessary to develop novel tools. These new assays will need time for development.”*

– *“I also advise the authors to recognize that complete prediction of safety will never be possible for human use products. The existing predictions do not recognize the misuse/abuse of commercially available prescription and non-prescription products, as well as the use of products correctly but in the setting of co-morbid conditions that produce skin inflammation (leg ulcers, skin infections, underlying skin diseases such as atopic dermatitis, psoriasis, etc). Such human conditions subvert the model systems that may rely on assumptions such as steady-state health (and associated lack of inflammation), which is not always the situation in the above-described conditions.”*

– *“Did the authors anticipate the possibility that tests, while not formally validated, can be incorporated in OECD guidelines if well established, scientifically accepted, and implemented?”*

– *“Overall, I support the view that the mechanism of skin sensitization is fairly well understood. However, I can see gaps that may affect the capability of the existing tests to properly assess sensitizing potential and potency.”*

– *“Even though no validated alternative in vitro tests to assess immunotoxicity exist, much progress has been made toward these assays in the last decade. Such models can, at least, be used for the pre-screening and hazard identification of unintended immunosuppression and contact hypersensitivity of direct immunotoxicants.”*

5.5 Further advice to the area

– *“This should give space to start developing integrated testing strategies for safety assessment of ingredients (single compounds).”*

– *“The main challenge remains: mixtures.”*

– *“It might be worthwhile for this area to look on the other side of the fence, specifically at the lessons gained from the Carcinogenicity area. Many of the same mechanisms are involved in both endpoints, e.g., inflammation, ROS-related pathways, apoptosis, and intracellular communication.”*

– *“Focus should be on scientific validation of the existing tests, development of integrated testing strategies, and assessment of mixtures.”*

– *“In vitro exposure of KC or DC to low concentrations of a chemical allows for the identification of genomic/proteomic marker profiles that discriminate sensitizers and irritants. The apparent importance of (low) concentration for prediction of sensitization and non-irritation seems to urge for a good understanding of the fate of a compound in the cellular test system (toxicokinetics). Currently, research within the FP7 project Predict-IV is aimed at developing tools and strategies for assessing the ‘real’ in vitro exposure of cells to a chemical compound. This is not mentioned in the report.”*

– *“Most of the reported tests have a certain degree of maturity, which points towards scientific validity. Several of these tests are also implemented by industry for hazard identification, early decision-making, and in a weight-of-evidence approach, e.g., in the context of REACH. What can we learn from these experiences? This information, in my opinion, is highly relevant for estimating the time line for a test or a testing strategy including this test to be accepted. Indeed, assays can be incorporated in OECD guidelines without formal validation (considered by the report as the bottleneck for acceptance), provided they are well established, scientifically accepted, and implemented.”*

– *“I understand the need for robust assays, but these numbers seem a little pessimistic. Scientific validation and blinded studies done by potential end users should be enough to allow companies to use alternative methods. Would it be useful to add a section that discusses regulatory validation versus scientific validation? There are numerous examples of methods being ‘validated’ and a protocol generated that no one can deviate from even when science dictates it. At the very least, it seems to me that it would be useful to define what is meant when the term ‘validation’ is used. There must*

be room for the use of what is available and scientifically validated now, and there must be flexibility in the process to allow evolution of each method's technology without the burden of revalidation. One suggestion is to allow labs to run the validated protocol, but when science dictates, allow the labs to run additional endpoints, or use alternative approaches, as long as correct controls and an explanation for the addition are included."

– "*My advice is to pay attention to the immunology and methods, and to recognize opportunities to transform the field with more forward thinking about areas of potential and unmet needs. The article focused on existing technologies and the timelines/deadlines. From this perspective, the review article was pragmatic but not forward-thinking. Certainly it makes sense to describe the best-developed assays, because these assays are the only available tools that can be studied enough to meet the proposed deadlines. It was a true review of the state of the art about the development of a toolbox. It needed some visionary thinking about potential new directions and the real need to revise the deadline for the implementation of* in vitro *testing and replacement of whole animal testing. The existing toolbox needs more emphasis on subcellular, molecular pathways, as well as the development of new mechanistic pathways. Considering the timeline pressure for the development of the assays, this may not be practical for meeting the deadlines in question. It is important to recognize that research to develop new assays will be necessary to advance the field, which may lead to the assays needed to successfully identify hazards as well as potency.*"

– "*A review, and possibly a workshop, on allergen potency and* in vitro *tests is highly advisable, together with a workshop on the state of the art of* in vitro *immunotoxicity.*"

6 Repeated dose toxicity

6.1 Overall assessment of the adequacy of this status report

An overall positive assessment with comments like:
– "*A reasonably well conceptualized overview of this topic and an executive summary that addresses the current situation appropriately, specifically with respect to alternative methods.*"
– "*A nice example of out-of-the-box thinking and a risk-based approach for a new way to generate a high quality safety assessment that could serve in the future to replace the current approach and strategy, above the replacement of individual animal studies.*"
– "*This chapter gives a realistic picture of the complexity of understanding repeated dose toxicity of compounds. The manifold interactions that may occur in an intact organism will have their impact on the effect of exposure to compounds. (This, however, is also true for acute effects.)*"
– "*Examples from other areas (TTC, QSAR, benefit-risk analysis), in particular pharma development, are used in a very intelligent way and adjusted for the purpose of assessing the safety of cosmetics. The specific issues and problems of this*

area are well covered, particularly the importance of interaction in the entire organism. Gaps are identified."
– "*The individual approaches described are quite complete and the advantages as well as the gaps are well identified. The paragraph on QSAR and* in silico *modeling is very instructive and provided ideas from other areas (e.g., pharma), which serve as examples for the cosmetic area.*"
– "*The notion that certain organs or tissues are more specifically of interest in repeated dose toxicity is well documented. It opens the possibility of studying effects on* in vitro *systems derived from these tissues or organs. Specially mentioned are the liver, the kidney, the nervous system, the cardiovascular system, the lung, the immune system, and the hematopoietic system.*"

Some criticism:
– "*Certain gaps, ambiguities, and misinterpretations or conflicting statements.*"
– "*Not focused on cosmetic ingredients.*"
– "*No clear conclusions from the past experience with regard to animal experiments are given and, unfortunately, no suggestions are made of how reduction and refinement with regard to RTT can be achieved.*"
– "*What is missing or insufficiently covered in the introduction is some kind of prioritization in testing the many target organs that can be affected by cosmetic ingredients.*"
– "*Not focused on cosmetic ingredients but rather frequently refers to and discusses aspects related to pharmaceuticals and/or chemicals. This is specifically valid for the passage on QSAR use in RTT.*"
– "*I would have expected some details on what the* in vitro *models developed in the past and currently used are lacking in order to be useful for RTT. How long should cells be exposed to test compounds, when do we need two-dimensional and when three-dimensional cultures? How important is the growth support? What strategies are applied in order to achieve optimum differentiation of the cell models utilized?*"
– "*It is obvious that the authors also have significant competence and expertise in pharma development, which is highly beneficial for this area. However, in some paragraphs, the aspects taken from pharma development seem not clearly enough distinguished from the actual scope of this report.*"

6.2 Is the judgment on promise and timelines adequate, overly optimistic, overly pessimistic?

Reviewers missed suggestions for timelines, e.g.:
– "*No clear timeline given in this area; the adequacy of timelines cannot be evaluated.*"
– "*No time line given except to say that alternatives to replace* in vivo *methods will not be available by 2013.*"
– "*The experts further state that full replacement for repeated dose toxicity is extremely challenging (an understatement in this reviewer's mind) and the time needed to achieve this goal will depend on a number of things: 1) progress at the research and development level, 2) adequate prioritization, 3) funding, and 4) coordination of efforts.*"

– *"The statement is correct, that almost all of the* in vitro *cell culture based models currently being pursued as models for RTT have not yet overcome the status of early basic research."*

– *"The timing for full replacements will continue to be difficult to predict."*

– *"I am concerned that these estimates are grossly overestimated. There is at least one method that could be evaluated now. CeeTox Panel. This method has already been used for nine years in pharma and was recently used to establish an* in vitro *approach to predicting acute oral LD₅₀ with excellent results. The LD₅₀ work was done in collaboration with L'Oréal and was presented at the WC7 and at SOT2011."*

– *"New prediction models based on a systems biology/toxicology-based approach need to be developed and applied. In combination with human-derived cellular* in vitro *models (including stem cell or iPSC derived systems) the currently used animal-based, long-term repeat toxicity models could be replaced by* in vitro *alternatives within the next 10 years, provided that funding agencies together with regulators are helping to continue the ongoing research in that direction."*

Pertinent comments organized by target organ:
Kidney

– *"Amazingly, it is stated that the kidney is not a major target in 'preclinical development of pharmaceuticals…,' which definitely is wrong. It's the second most affected target organ in xenobiotic, chemical, and drug-induced toxicity, as correctly stated in a later passage on nephrotoxicity (this, by the way, also is stated for cardiotoxicity). Most interestingly, in the new joint EU-COLIPA project within EU FP-7, the kidney is not considered a target organ in the research program."*

– *"The kidney is considered to be of high importance when assessing the safety of new drugs, and many companies have* in vitro *screening programs in place to identify this type of toxicity."*

– *"The statement that primary renal cultures of rodents are widely used is incorrect or at least misleading. I am not aware of any reasonable, well-standardized procedure that allows maintaining stable rodent renal primary cell phenotypic function for more than several hours. Human primary cell cultures are more important as they can even be sub-cultured and thus be maintained in culture displaying stable phenotypic functions over far longer periods.*

In contrast to other epithelia, non-transformed continuous renal cell lines are available from several species (cattle, pig, dog, opossum, and monkey; see Tab. II). Some of these cell lines even comprise sub-strains that have retained the property of gluconeogenesis, a biochemical pathway normally expressed only in proximal tubular cells and hepatocytes. These continuous lines possess a karyotype identical to that of their in vivo ancestors. This karyotype is stable over more than a hundred passages.

Human renal cell lines so far immortalized by viral oncogene transfection, e.g., HK-2 cells, suffer from passage number dependent changes in their karyotype."

– *"The recent introduction of telomerase transfected renal proximal tubular cells (RPTEC/TERT1) … (Wieser et al., 2008) is quoted in the report. However, the description is completely misleading. This cell line has characteristics that make it more comparable to the human correspondent than other proximal tubular cell lines, because the energy metabolism of the RPETEC/TERT1 cell line is less dependent on glycolysis, but rather on oxidative metabolism, as in the* in vivo *situation. This property can be even enhanced by allowing the cells to differentiate, which is achieved by keeping them under conditions of contact inhibition in monolayer cultures over a certain period of time."*

– *"The statement that no models for the renal medulla are available is incorrect. The dog MDCK cell line represents a perfect collecting duct model, as it can be easily adapted to high osmolality media (600 mosmol/l and above) (Nakanishi et al., 1989). In addition, several renal interstitial cell lines originating from the medullary region have been established. Cell lines representing the distal nephron segment are available only from the mouse (GRB-MAL1, Mdct, Univ. of Pittsburg), but these lines have hardly been used as an* in vitro *model in nephrotoxicity testing. Immortalized lines exist for glomerular mesangial and epithelial cells (Nguyen et al., 1996; Sarrab et al., 2011; Delarue et al., 1991). In addition, fairly well established culture models for primary human glomerular endothelial cells exist, which potentially could be used for repeat dose toxicity testing (Aydin et al., 2008)."*

– *"Gaps are identified. …for example, nephrotoxicity:* 'in vitro *models for other segments of the nephron are limited. No model is available for assessment of the potential for toxicity to the kidney medulla.' What is the risk of this gap (historically and statistically, how many compounds induce toxicity in the medulla), and what is proposed (methods or alternative strategies) to fill this gap?"*

Tab. II: Continuous renal epithelial cell lines (not transformed, normal karyotype)

Cell line	Animal derived from	Nephron segment of origin
LLC-PK₁	Yorkshire pig	proximal nephron
OK	North American Opossum	proximal nephron
JTC-12	Monkey	proximal nephron
MDCK	Dog (Cocker Spaniel)	(cortical?) collecting duct
MDCK	*Xenopus laevis*	distal tubule/collecting duct

Liver
– *"The section on hepatotoxicity gives a short but adequate outline of the current status regarding* in vitro *models, which potentially may be used in RTT."*
– *"90% of all biochemical functions that define the liver are carried out in the hepatocyte. While there are examples of drugs and chemicals that activate Kupffer cells, which then elicit toxicity via release of proinflammatory cytokines, the biochemical changes in the hepatocyte are key to determining organ toxicity. I don't believe that it is essential to know which organ resulted in animal toxicity. It may be enough to be able to relate plasma concentration or AUC to general toxicity, which may occur in any organ. If one knows that an* in vitro *value provides a plasma toxicity threshold, (e.g., a concentration above which toxicity would be expected in some organ, see McKim (2010). Using this approach we have screened more than 600 drugs with excellent results. It is the pharmacokinetic properties of the chemical that ultimately define target organ dose and hence organ specific toxicity. Clearly, the primary organs, in pharma at least, are liver, kidney, bone, and heart. Also, it is not the cell type used that defines target organ toxicity, but rather the specific endpoints monitored. Are they unique to liver for example? Once a panel of endpoints or biomarkers is identified, a good method for determining organ specificity is to expose another cell type. Liver versus kidney for example."*

Heart
– *"There is little to say about myocardial toxicity. The potential hazard by arrhythmogenic compounds is adequately discussed. Three-dimensional* in vitro *models already available could perhaps be discussed additionally. The relevance of stem cell derived cardiomyocytes with respect to their direct comparability to primary cultures could be discussed in more detail."*

Brain
– *"The* in vitro *neurotoxicity models listed cover those commonly used for testing and describe approaches currently being developed, which might become of relevance in repeat* in vitro *toxicity testing."*
– *"Neurotoxicity (CNS) is an important issue. We have found, however, that isolating key receptors in the brain and using these in an isolated format can provide information on critical brain-specific toxicity. For example, nicotine is relatively nontoxic in most* in vitro *models. Yet it has a low LD_{50} and is considered to be quite toxic. If the nicotinic receptor is used as an added piece, the toxicity profile changes dramatically and an accurate predication can be performed. Predicting brain dose and BBB penetration are also helpful, although good estimates can be done by LogP."*

Lung
– *"Although progress has been made over the past few years with regard to* in vitro *models of the various portions assembling the pulmonary system, these models suffer from several drawbacks, which are not sufficiently covered by the report.*

The models available for the airway epithelium are largely based on cells and cell lines derived from adenomas. Although they have retained numerous functions from their in vivo *correspondents (beating cilia, mucus secretion) they usually are not co-cultured with their immediate neighbor cells as smooth muscle cells and defined immune cells are. The cell models are not very well characterized with regard to their potency to produce certain mucus constituents and, specifically, surfactant proteins. A discussion of the lungs barrier cells relevant for gas exchange, the alveolar epithelium, is missing."*

Immunotoxicity
– *"Concerning systemic immunotoxicity and* in vitro *evaluation of immunosuppression, I am not particularly impressed by the work done (very short and superficial) and I don't like the order in which alternative tests are listed in Table 9. However, I have to agree with the timelines proposed (>10 y) as, at least in Europe, very limited efforts (and resources) are currently underway (or available) to address this aspect. ToxCast and Tox21 may offer better perspective in the near future, also regarding immunotoxicity."*
– *"There is an error in the reference cited on page 420; it should be Carfi' et al., 2007 and not Carfi' et al., 2010. Furthermore, several relevant reviews that specifically address* in vitro *immunotoxicity are not even mentioned, i.e. Galbiati et al., 2010; Corsini and Roggen, 2009; Lankveld et al., 2010."*

Further specific comments:
– *"It is interesting to read about an SCC(NF)P survey on cosmetic ingredients giving the percentages of 28 and 90 days subchronic toxicity studies carried out over the last 5 years or so. It would be of more interest, however, to know how many and which types of ingredients have been examined in these studies."*
– *"The benefit of one specific statement on botulinum neurotoxin and its impact on cosmetics is not clear. Consequently, it is suggested to delete this small paragraph."*
– *"In the description of nephrotoxicity models an overview reference is given, which is not easily accessible as it is a PhD thesis of the University of Brussels (VUB)."*
– *P. 409, line 1: "need to define long term."*
– *P. 409, second column, line 3 of Introduction: "I would recommend the sentence be modified to state 'Repeat-dose toxicity is non-target organ directed and comprises…'; the last sentence on page is very well said."*
– *P. 410, second column, second paragraph, line 6: "I recommend "tissue/systems" be the terminology, given the discussion immediately preceding this sentence."*
– *P. 410, second column, second paragraph, line 11: "I find this extremely difficult to understand. Does the importance of the liver vary according to whether a drug is in the preclinical, clinical or post market stage of development? I doubt that seriously. If so why?"*
– *P. 410, second column, second paragraph, line 13: "I think what the authors mean is that the appearance, not the importance, of a certain toxicity to an organ may only appear at certain stages of drug development, such as post-market*

exposure, and a drug that causes an idiosyncratic reaction is an example."

– *P. 410, last paragraph: "With regard to pharmaceutical compounds, the organs most associated with failure are the liver, kidney, and heart."*

– *P. 411, line 4: "This is an incorrect statement. The immune, respiratory, musculoskeletal, and gastrointestinal systems all can be identified as toxicity endpoints pre-clinically. Otherwise they would not be a part of the protocols for systemic toxicology studies."*

– *P. 411, second column, fifth full paragraph, line 1: "It should be noted that these are commercial models."*

– *P. 413, third full paragraph, line 11: "The wording 'compound that constitutes a baseline effect' is very ambiguous. If the report is intended for a varied scientific audience, terminology such as this, I would suggest, needs an example."*

– *P. 413, second column, line 3: "What about vitamins? Are they considered pharmaceutical or non-pharmaceutical?"*

– *P. 414, second column, last sentence in first paragraph: "an excellent point"*

– *P. 423, second column, second full paragraph, last line: "another excellent point"*

– *P. 426, first full paragraph, line 3: "what does 'pers' mean?"*

6.3 Are there promising methods and approaches that were omitted?

All together, the report was perceived as quite comprehensive, with lesser emphasis on modern approaches such as omics, ITS, organotypic cultures, and interlink with kinetics.

– *"Methods and approaches are quite complete."*

– *"The conclusion for this area and the gap analysis provided are complete and well justified. It could be of additional benefit to provide a table with the gaps currently encountered and milestones and rough timelines of what is needed to fill these gaps."*

– *"The section on hepatotoxicity... What is not adequately stressed is the use of cell culture technologies known to affect differentiation of hepatocytes, e.g., co-cultures (with endothelial cells, monocytes) and 3-dimensional cultures."*

– *"Toxicity testing for intestinal epithelia has not been stressed, although it might be of high importance for repeat toxicity of cosmetic ingredients!"*

– *"One model, organotypic cultures of brain slices could have been discussed as an additional methodology as it allows long term exposure of 'intact' brain tissue and the possibility of applying morphologic and electrophysiological monitoring techniques to gain information on mechanisms of neurotoxicity."*

– *"CeeTox (http://www.ceetox.com) for a variety of potential alternative approaches including dermal toxicity and percutaneous absorption,* in vitro *acute toxicity-LD$_{50}$, ocular corrosion and irritation, and organ specific toxicity panels for heart, liver, and kidney."*

– *"Hurel (http://www.hurelcorp.com/index2.php) for organ specific toxicity panels."*

– *"An excellent cardiomyocyte system derived from stem cells is currently available in the US. It has had multiple testing studies and should be included in this discussion."*

– *"A small section deals with the current activities aimed at implementing -omics technologies to* in vitro *long-term and RT testing. Several attempts have been undertaken and supported by the FP 6 and FP 7 EU research programs. The Predictomics FP-6 RTD-project, for example, tried for the first time to make use of toxicogenomics in human- and rat-derived liver and kidney cell culture models in order to be able to identify and predict potential target organ toxicity from gene activation profiles using whole genome DNA-arrays. The currently running FP 7 project, 'Predict-IV,' which was built upon the experiences of FP 6 'Predictomics,' tries with the use of an integrated approach utilizing metabolomics, proteomics, and transcriptomics, in combination with information from high content imaging, to identify gene activation patterns characteristic for the activation or changes in activity of certain intracellular signaling and/or metabolic pathways. In contrast to other comparable investigations based on* in vitro *alternatives, this integrated approach also tries for the first time to implement toxicokinetic and dynamic data, which can be implemented to existing or newly developed PBBK and PBPK models. The project can be regarded as a first step in the attempt to develop tools that in the future may be implemented into a systems toxicology approach, which in my view needs to be established, anyhow, in analogy to systems biology, and in the end should complement systems biology."*

– *"Before starting with* in vitro *tests, bioavailability should be considered. If the compound does not have appreciable bioavailability, immunotoxicity is unlikely to occur. As a general strategy,* in vitro *testing for direct immunotoxicity should be done in a tiered approach (Gennari et al., 2005), the first tier measuring myelotoxicity."*

– *"Compounds that are capable of damaging or destroying the bone marrow often will have a profound immunotoxic effect, since the effectors of the immune system itself will no longer be available. Therefore, if a compound is myelotoxic, the material will be a de facto immunotoxicant."*

– *"Compounds that are not overtly myelotoxic may still selectively damage or destroy lymphocytes, which are the primary effectors and regulators of acquired immunity. Compounds are therefore tested for lymphotoxicity (second tier). This toxicity may result from the destruction of rapidly dividing cells by necrosis or apoptosis; alternatively, chemicals may interfere with cell activation affecting signal transduction pathways. A variety of methods are available for assessing cell viability (e.g., colorimetric, flow cytometric assays). After myelotoxicity and overt cytotoxicity are excluded as endpoints, basic immune cell functionality should then be assessed by performing specific functional assays, i.e., proliferative responses, cytokine production, NK cell activity, etc. (third tier), using non-cytotoxic concentrations of the tested chemicals (viability > 80%)."*

– *"The* in vitro *system, named 'fluorescent cell chip,' is based on a number of cell lines derived from T-lymphocytes, mast cells, monocytes, each transfected with various cytokine reporter*

cell constructs for measuring cytokine expression (Ringerike et al., 2005). Although further refinement of this system is required, this assay holds promise for in vitro screening of chemicals for their immunotoxicity."

– "The human whole blood cell culture, introduced more than 20 years ago, may also be useful in studying the biological effects of potential immunomodulatory chemicals based on immune cell activation and cytokine secretion (Langezaal et al., 2001, 2002)."

– "Cytokine production, together with lymphocyte proliferation, is currently in a pre-validation phase (Carfi' et al., 2007). Based on results of two previous studies, the human T cell activation assay was selected as the most promising of the investigated in vitro immunotoxicity test. This assay is based on CD3/CD28-mediated T cell activation using proliferation and cytokine release (TNF-α and IFN-γ) as read-out parameters. Statistical analyses revealed that the human T cell activation test had a 'sensitivity' (correct prediction of immunosuppressive chemicals) of 76% and a 'specificity' (correct prediction of non-immunosuppressive chemicals) of 83% (manuscript in preparation). The human T cell activation assay may be a promising candidate for in vitro evaluation of immunosuppressive activity."

– "Immunotoxicogenomics represents a novel approach to investigate immunotoxicity. Hochstenbach et al. (2010) have recently reported the possibility of using a set of 48 genes to distinguish immunotoxic from non-immunotoxic compounds using human peripheral blood mononuclear cells. These genes might be considered as candidate biomarker genes for immunotoxicity screening. However, even if many of the annotated genes appear to be immunologically relevant, in vivo studies in the human population or in experimental models are necessary to demonstrate their effective relevance."

– "In animals, production of T-dependent antibodies is considered to be the 'gold standard.' However, there are currently no good systems for in vitro antibody production using human cells. Recently, Koeper and Vohr (2009) reported that using a modification of the Mishell-Dutton assay with female NMRI murine splenocytes, all six immunosuppressive compounds (with the exception of cyclophosphamide) and all four non-immunotoxic compounds tested were correctly identified. Further explorations of this model, therefore, are recommended."

6.4 Do you agree with the conclusions for the area? If not, where do you disagree?

In general, agreement is noted, e.g.:
– "The need for knowledge on the underlying mechanisms of toxicity, as well as the complete picture of the mode of action (i.e., the full process starting with the first chemico-biological interaction, the adaptive and pathological changes in the cells, tissues, and organs involved, up to the clinically observable toxic endpoints) is well emphasized in the report."

– "An overview is given of the different in silico (i.e. QSAR) and in vitro models for studying these organ-specific effects. The incorporation of rapidly developing technical possibili-

ties to have high-content information, viz. the omics, imaging techniques, and systems biology approaches opens excellent possibilities for interpreting systemic toxicity."

– "In conclusion, the report rightly elaborates on the above-mentioned aspects of repeated dose toxicity."

– "Methods such as the whole blood assay, lymphocyte proliferation, and cytokine production can be used for the hazard identification of the immunosuppressive potential of chemicals, whereas several in vitro methods are already available to identify allergens and, possibly, to discriminate contact from respiratory allergens and to classify sensitizers accordingly to potency."

6.5 Further advice to the area

The recommendation is clearly to move toward strategic planning for the future:
– "No concepts, ideas and suggestions on future strategies are given to further develop and position these methodologies."

– "The described pros and cons of the current in vivo testing strategy is very instructive and encourages the development of new and more unorthodox assessment strategies. The importance of pharmaco- and toxicokinetic information as the basis for an educated decision-making process and safety assessment is well emphasized. This aspect might encourage the preparation of an overall approach (beyond the individual areas evaluated in this report) to develop one strategy that can serve several elements of an overall safety assessment. It is obvious that toxicokinetics plays a key role for all other areas. So it might be beneficial to design one strategy (perhaps even with one or very few animal studies) that provides concise information crucial for all areas involved in the respective safety assessment."

– "The report also stresses the important role biokinetics play in these processes. It is concluded that the important dose-metric for systemic repeated-dose toxicity needs to be the internal exposure, i.e., the concentration at the site of toxic action. And since this implies knowledge of absorption, distribution, and elimination processes, this shows that biokinetic considerations are indispensable."

– "The authors could be encouraged to extend their excellent description of individual tiered testing strategies and provide an overall proposal for a tiered safety assessment strategy in repeated dose toxicology in the form of workflows and decision trees."

– "The industry strategies described are very instructive and could serve as the basis for proposing a new overall tiered safety assessment strategy for this area."

– "I support the general conclusion of the report that, although this field is rapidly evolving (if not revolving), and the introduction of new technologies and integrated testing strategies is very promising, we are not at a point that a full replacement of repeated-dose animal studies has been reached. We need to build on the existing examples that at least show the proof of principle that proper predictions of human risk can be made, so that we can widen the data bases for a better evaluation of these integrated approaches."

– *"It's true that the current* in vivo *methods provide the data listed above. A suggestion is to list what would be expected of an alternative method. For example, I'm not sure that target organ information is as useful as the NOAEL, toxicity in general, metabolic activation, margin of safety. For* in vitro *tests a realistic desired outcome should be decided upon."*

– *"Integration of data from different sources is the key to understanding systemic toxicity, if it will have to be based on* in vitro *toxicity data. It involves the evaluation of (organ- or tissue-specific) functional disturbances and the extrapolation of concentration-effect relationships to a dose-metric for the* in vivo *situation. It implies the choice of the relevant biomarkers to be measured* in vitro. *This might be done on the basis of information derived from structural properties of the chemical under study. It also implies that these biomarkers take into account the possible role of interactions between organ and tissue systems and the involvement of feed-back loops resulting from these interactions, i.e. adaptation or homeostasis. The major challenge here will then be: are we able to distinguish effects in the realm of adaptation vs. adverse (toxicity) effects?"*

– *"One aspect needs to be stressed here, though: the proper use of biomarkers in the manifold* in silico *and* in vitro *methods will need to take into account that we probably have to redefine the point-of-departure in a risk assessment procedure. The points now in use, viz. NOAELs or BMDLs, will almost by definition be cruder than the possibilities given by the new technologies* in vitro."*

– *"The endpoints relevant for regulatory decisions are, as correctly stated, derived from animal experiments and are based on NOAELs, which are derived from pathologic changes. When using* in vitro *models, changes in a cell type's phenotypic function will determine NOAEL. In order to assess such a dysfunction, the border conditions under which the respective* in vitro *cell or tissue culture model is kept need to be known. Various tissues and cells express different functions, depending on intra- and extracellular signals. Under appropriately standardized conditions changes can be identified in every cell type of interest with high sensitivity, depending on the markers selected and the analytical tools applied. We should never forget that we have not yet identified one marker for any one of the many mechanisms involved in toxicity. Attempts to apply the same procedures as in animal experimentation to predict and quantify risk brought about by potentially toxic cosmetic ingredients, pharmaceuticals, or chemicals will perhaps prevent the utilization of the predictive potential of tailored* in vitro *cell and tissue culture models."*

– *"Due to the complexity of the immune system, it is generally assumed that it would be very difficult to reproduce it* in vitro. *To a large extent,* in vitro *systems do not take into account the interactions of the different cellular and soluble components involved in the immune response, or the potential for neuro-immuno-endocrine interactions. Therefore, an assessment of* in vitro *immunotoxicity will be valuable only in the cases of a direct immunotoxicant. To this purpose, several isolated processes can be studied* in vitro, *such as lymphocyte proliferation, cytokine production, phagocytosis, lytic functions, and even primary antibody production."*

– *"Despite all these efforts, however, there is still a clear need for continued investment in the development of methods and approaches that will allow the correct identification* in vitro *of potential immunotoxic compounds, including immunogenicity and autoimmunity. Intensive international and interlaboratory cooperation and coordination will be necessary to reach this goal."*

– *"Without encouragement from governments, the process will continue to move slowly."*

– *"It seems to this reviewer that greater co-operation between the US and the EU would help move the process forward."*

7 Carcinogenicity

7.1 Overall assessment of the adequacy of this status report

There was an altogether very positive assessment of the document:

– *"The report gives an extensive overview of the approaches for the detection of (primarily) genotoxic carcinogens currently used, improved, or developed. Inclusion of* in vivo *methods implementing refinement and reduction strategies is appreciated but considered irrelevant for the purpose of this review (cosmetic industry and the 7th Amendment of the Cosmetic Directive)."*

– *"The report is well structured and provides clear assessments of status, applicability, and limitations of each discussed approach."*

– *"This is a very large, complex topic area, and the review handles this complexity with appropriate tone and depth. Appropriate context is provided on the successes from past research in this area. The review on this area discusses the well established framework for evaluating chemicals for genotoxicity and how this is used to evaluate that mechanistic underpinning for carcinogens. The review also highlights what limitations the lack of* in vivo *confirmation would mean if animal use is discontinued. For most other mechanisms, less well established endpoints are in use or are in development with alternative assays."*

Some negative remarks:

– *"It would have been appreciated, however, if concrete accuracy values were given where possible (as e.g., for QSARs for potency). As discussed above, this information must be highly relevant to assess the timeline for a test to be implemented and accepted. Indeed, the higher the relevance of the test and the higher the quality of its performance, the more likely it is to move forward towards acceptance (shorter timeline)."*

7.2 Is the judgment on promise and timelines adequate, overly optimistic, overly pessimistic?

There was an overall agreement with the suggested timelines:

– *"The proposed timelines (Table 10) are, in my opinion, adequate for formally validated tests. However, tests can be in-*

corporated in OECD guidelines if they are well established, scientifically accepted, and implemented, even when not formally validated. Did the authors anticipate this possibility while estimating the timelines?"

– *"The experts were unable to suggest a timeline for full replacement of the animal tests currently needed to fully evaluate carcinogenic risks of chemicals, but the timeline is expected to extend past 2013 (overly optimistic in this reviewer's mind); full replacement is at least 20 years off."*

– *"This reviewer agrees with the conclusion of the authors that currently there are not sufficiently developed and tested methods to support full replacement of animal-based tests, especially for non-genotoxic mechanisms."*

– *"The statement of rare use of the 2-year cancer bioassay in rodents for cosmetics is very important, as a reduction thus is already in place. Information from whole organisms is, however, not fully replaced by non-animal studies related to toxicokinetics, genotoxicity, and repeat dose. I agree this will not be possible within a short-term time range."*

Specific comments:

– *"This reviewer noted that this section of the report has a very useful discussion of TTC within the section on 'status of validation and/or standardization,' some of which related to endpoints other than cancer. Authors should consider whether, for reader clarity, this section should move forward, ahead of the previous area discussion on TTC in the toxicokinetic section."*

– *"The carcinogenicity section has a detailed discussion of the QSAR models available for this area. This reviewer would suggest putting some of these details into a table that, similar to Table 10, compares and contrasts the different methods. A column that describes the database of chemicals used to establish the QSAR relationships also would provide clarity and allow the reviewer more clearly to determine utility for future replacement efforts."*

– *P. 432, line 16:* *"The word 'no' gives a degree of surety that our methods cannot provide; thus, I would say 'little' as opposed to 'no.'"*

– *P. 434, Non-Testing Methods:* *"I suggest the subtitle be changed to 'in-silico' methods. 'Non-testing methods' is very vague."*

– *P. 437, last paragraph:* *"Cite Hayashi et al. (2005)."*

– *P. 439, first full paragraph, first line:* *"What interesting perspectives? This tends to leave the reader asking: What do the authors mean?"*

– *P. 441, second column, fifth full paragraph, line 4:* *"This test can be viewed in one sense as a higher tier in that it is closer to the human situation, but in another sense it can be viewed as a lower tier in that the endpoint is not as definitive as a gene or chromosomal mutation – it is DNA damage. I think this point needs to be made clear."*

– *P. 441, first paragraph:* *"Also ICH draft guideline S2 (R1)."*

– *P. 442, second column, and first full paragraph, line 2:* *"Should this be 2011?"*

– *P. 443, 5th paragraph:* *"Proper reference for Bhas 96-well plate method should be cited."*

– *P. 444, 4th paragraph:* *"Comet validation should acknowledge JaCVAM initiative cooperated with ECVAM and ICCVAM/NICEATM."*

– *P. 444, 5th paragraph:* *"T25 should be defined; insert "For the reduction of animals in genotoxicity studies, the incorporation into other toxicological studies, or the combination of two or more genotoxicity studies, e.g., comet assay and micronucleus assay using transgenic animals for gene mutation detection (ICH; Pfuhler et al., 2009)."*

– *P. 446, 2nd paragraph:* *"Add Hayashi et al. (2005) in (e.g., Matthews et al., 2008)."*

7.3 Are there promising methods and approaches that were omitted?

– *"This chapter also includes Read across and Threshold of toxicological concern approaches, while the weight of evidence approach is briefly mentioned. The latter could have been given more attention, since it is already implemented by the chemical industry in the context of REACH."*

– *"To my knowledge, no advanced approaches have been omitted. The review is quite extensive. With respect to toxicogenomics, it is a pity that the progress made by the FP6 project carcinoGENOMICS was not taken up. This project is currently evaluating the capacity of selected kidney, liver, and lung cell based test systems to discriminate genotoxic and non-genotoxic carcinogens."*

– *"This reviewer agrees with the details given in the area of the report that describes how mechanistic information about the myriad potential and proven mechanisms of cancer could be assessed. This reviewer also shares the authors' encouraging comments on the potential for toxicogenomics. As only gene expression profiles for rat hepatocarcinogenicity are listed as an example, however, using this for replacement of in vivo animal assays is more theoretical, and this development is well behind what is needed for replacement."*

– *"I would recommend the authors read a recent article by Paules et al. (2011). There is some very interesting information that has bearing on the just addressed problem."*

– *"I think it is worthwhile to describe the categorical approach based not only on the structure of chemical but on the activity of chemical (AAR)."*

7.4 Do you agree with the conclusions for the area? If not, where do you disagree?

– *"In general, I do agree with the conclusions, as well as with the timeline indicated in Table 10 for each of the individual approaches."*

– *"This reviewer agrees with the conclusion that the timeline for replacement is unknown except that it would be beyond 2013."*

7.5 Further advice to the area

– *"There is a wide gap in the area of non-genotoxic carcinogen identification. Mechanistic understanding needs to be built upon as soon as possible."*

– *"Especially with respect to* in vitro *identification of non-genotoxic carcinogens (but not exclusively), it might be worthwhile for this area to look at the other side of the fence, specifically at the lessons learnt from the Sensitization area. Many of the same mechanisms are involved in both endpoints, e.g., inflammation, ROS related pathways, apoptosis, and intracellular communication."*

8 Reproductive Toxicity

8.1 Overall assessment of the adequacy of this status report

The overall evaluation was very positive, e.g.:

– *"This area is very interesting, very current, and on-target. Reproductive toxicity includes a highly complex series of inter-related physiological processes that are well represented in this review. Inclusion of Figure 10, which shows the overall process and stages in reproduction, was essential for the reader to fully appreciate what the 3Rs means within this context. Further, the inclusion of a detailed discussion of the current, promising, and 'in development' assays was essential and was well done. Relevance specific to the 7th amendment to the EU Cosmetics Directive was given through this area of the report."*

– *"The 2011 report covers the status of non-animal testing for various stages of the reproductive cycle: conception to birth, postnatal development to sexual maturation, and gamete production. The complexity of development and reproduction has been adequately portrayed in terms of the multiple complex steps required for successful propagation of the species, and a lack of understanding of the mechanisms for reproductive developmental toxicity. The report also covers the main historical reasons for slow progress in the implementation of alternative methods for safety evaluations."*

– *"The description given in the section inventory of animal test methods currently used… is very good and informative for the comprehensive tale developed, naming methods currently developed within reproductive toxicology."*

– *"The individual approaches are well described and complete. Timelines for the individual methods are reasonable. The link between the individual approaches, how they contribute to the overall safety assessment, and where the gaps are in this concept, could be defined in more detail."*

More negative remarks:

– *"Following some general introductory comments on the complexity of reproductive toxicology, the currently available OECD guidelines on reproductive toxicology are listed and explained. This is followed by an inventory of alternative methods. The exact link between these two sections and how exactly the listed alternatives could replace the animal studies listed under the OECD guidelines could be more specific and transparent. Table 11 is certainly helpful in this respect, but some more explanation and perspective would be beneficial to the assessment of this report."*

8.2 Is the judgment on promise and timelines adequate, overly optimistic, overly pessimistic?

– *"I agree with the estimate that it will take >10 years to develop a sufficient armamentarium of alternatives for reproductive and developmental toxicity."*

– *"Full replacement is not probable within a foreseeable time frame (>10 years; overly optimistic; again in this reviewer's mind, full replacement is at least 20 years off)."*

– *"This reviewer agrees that the overall judgment on the promise and timeliness of approaches is on target, and this discussion and review is adequate. As noted above, because of the complexity of the process of reproductive toxicity, the individual assay advances that are present and which are significant were occasionally understated. However, the overall assessment is not significantly changed, and full replacement will not occur within the foreseeable future (>10 years)."*

– *"For reproductive and developmental toxicity, alternatives that could replace whole animal testing for cosmetics regulatory testing simply are not available at this time."*

– *"The reproductive toxicity subgroup concluded that it will take 'more than 10 years' to develop a suite of alternative methods sufficient to replace animal testing for reproductive toxicity. While this certainly will be perceived as undue pessimism by some, the conclusions of the authors are firmly grounded in reality. Reproductive toxicity assessment is arguably at the extreme end of the complexity spectrum, as it covers the creation of entirely new individuals and their subsequent development through embryogenesis, fetogenesis, the perinatal period, adolescence, adulthood, and finally reproductive senescence. Reproductive toxicology must even deal with unique, albeit temporary organs (e.g., placenta, primitive kidneys), which are formed only during specific developmental periods. For each one of these life stages there are numerous potential modes of toxic action, each of which could have its own specific assay. Toxicokinetics, which is a major determinant of toxicity, also varies according to life stage."*

– *"The 2011 report concludes that an ITS will require more than 10 years to complete a transition to non-animal testing for health protection. This is a reasonable assessment, given the feasibility being demonstrated for projects in the US (e.g., ToxCast), and Europe (e.g., ReProTect). The timeframe of 3 to 5 years for model development, 3 to 5 years for assay validation, and 3 to 5 years for regulatory acceptance seems reasonable at this time (9 to 15 years total). Some of the resources, tools, and approaches to overcome the limitations of* in vitro *systems and the challenges of extrapolation to* in vivo *toxicity are currently being actively investigated in Tox21 and ToxCast."*

Specific comments:

– *"Fig. 10 in the 2011 report provides a framework to map non-animal alternatives to discrete serial stages in the reproductive cycle. This reduces the complexity of the reproductive cycle to conceptual building blocks. Validated non-animal alternatives have been implemented to varying degrees of*

success at each level. No single platform seems to cover the breadth of molecular pathways, cellular behaviors, and tissue interactions across the whole cycle. Some of the potential non-animal tests listed in Table 11 still rely on animals for in vitro *samples to be generated and cannot be regarded as animal-free. The report indicates that refinement and reduction of animal studies are a more attainable goal than replacement. For example, the extended 1-generation study has been estimated to reduce the animal use by 40% relative to a 2-generation reproductive study."*

– *"The 2011 report describes an inventory of alternative methods designed to detect developmental and reproductive toxicants based on* in vitro *and* in silico *methodologies. Embryo-fetal development is the most critical area for hazard identification. Tests for prenatal developmental toxicity include integrated testing on whole embryos and complex cell cultures. The application of mammalian (rodent) whole embryo culture has the benefit of studying the embryo with intact circulation, with the drawbacks of a narrow test period (organogenesis) and the need for pregnant animals as a source of embryos. The most promising non-animal alternative is the zebrafish embryo test and embryonic stem (ES) cell test. The zebrafish embryo test has the benefit of screening and mechanistic studies (genetics, imaging) covering stages from fertilization through hatching, with the drawback of solubility and transferability of anatomical differences (e.g., gills versus lungs, fins versus limbs). The ES cell test provides a rational system for studying multiple pathways of differentiation and may be extended to human assays to identify predictive biomarkers for efficacy and safety assessment. In both the zebrafish and ES platforms, genetic manipulation can provide distinct biomarkers for developmentally important pathways (e.g., estrogens, WNT signals) to evaluate biologically important perturbations at a pathway-based level."*

– *"Tests for placental perfusion and trophoblast models have limited applications for assessing placental toxicity and transport. Tests of gamete biology in males (computer-assisted sperm analysis, Leydig cell and Sertoli cell cultures) and females (follicle culture, oocyte maturation) and fertilization (bovine IVF) have been relatively well-established. A test for preimplantation biology during zygote transport (mouse preimplantation embryo test) is still under development. In vitro tests for assessing effects on the endocrine system (proliferation in Ishakawa cells, MCF-7 cells) and nuclear receptor activity (primarily, ER, AR, TR) are undergoing development and validation for regulatory acceptance. ER and AR pathways are critical at multiple stages of development, and the TR pathway is highly relevant for neurodevelopment. The H295R adrenal cell line has the advantage of an intact steroidogenesis pathway and is being considered a validated non-animal alternative for hazard identification with regards to endocrine disruption."*

– *"I agree with the emphasis on the extended one generation study as a major effort toward reduction."*

8.3 Are there promising methods and approaches that were omitted?

A remarkably high number of missing, not up-to-date, or under-emphasized elements were suggested:

– *"The Adler et al. report was extremely comprehensive in its review of the available methods, and there were no noticeable omissions of promising methods or approaches. They covered long-standing developmental toxicity alternatives, such as whole embryo culture and the micromass assay, as well as newer methods such as the Zebrafish Embryo and Embryonic Stem Cell tests. In the reproductive area they reviewed an extensive number of* in vitro *assays for specific reproductive processes. As there are few reproductive alternatives currently used in toxicology, the authors drew upon assays from the recent ReProTect project, basic reproductive biology, and even clinical* in vitro *fertilization fields to ensure comprehensiveness. The scope was expanded even further by including* in vitro *endocrine assays, and also by discussing QSAR, SAR, read-across, and related non-testing approaches. All in all, the report was extremely comprehensive."*

– *"The report provides a complete overview on the currently available alternative methods and their use in the individual stages of the mammalian reproductive cycle. The general approach to compare current OECD guidelines on reproductive safety assessment with an inventory of alternative methods organized into the individual stages of the reproductive cycle seems to be quite specific and not always intuitive."*

– *"Several missing elements would have been great additions to the detailed report. For example, Table 11 presented the extensive list of alternative methods. It is striking, however, that the table does not have a column that relates the test to the stages in the mammalian reproductive cycle. Such a cross reference would show, for example, the importance of the availability of the extended one-generation reproduction toxicity study guidance (draft OECD guidance) and how many of the stages this one test covers. The discussion of this assay includes the statement that 'this procedure will possibly reduce the number of animals by 40%'. This is extraordinarily important, especially as the final conclusion of this area is that the full replacement of animal studies for reproductive toxicity hazard assessment is 'not probable within the foreseeable time frame (>10 years).' Further discussion on what progress can be made in the next 10 years: Discussion of the framework at the end of the section as the 'overall strategy' for working towards a conceptual OECD testing framework is presented, and this 'within the next 10 years' context needs some more details. Additional comments relevant for this assay also could include reference to current research at NIEHS NTP to optimize additional details of the one-generation assay (NTP report by Foster 2011)[3]. Also, in the discussions on pharmacokinetics, additional comments on how improvements in these pre-screening activities would make the use of the modified one-gen test even more effective could be added to illustrate progress anticipated within 10 years."*

---

[3] http://ntp.niehs.nih.gov/ntp/About_NTP/BSC/2011/April/MOGDesign.pdf (last accessed, 17 August 2011)

– *"The report could include several other examples. In the discussion of the currently available assays, the area report mentions OECD 426, developmental neurotoxicity. This reviewer would suggest that some cross-references to the 'repeated doses' testing area report should occur at this point in the reproductive toxicity area. In particular, this reviewer would suggest adding a few notes on the availability of alternative methods for evaluating neuro endpoints, such as neurite outgrowth during different periods of neurodevelopment in cell culture, as relevant for the reproductive toxicity area assessment. The section on zebra fish could also include some reference to behavioral assays being done as part of the tests on reproduction and development with this model organism."*

– *"Since some of the assays discussed in this section are quite early in their development, this reviewer would also suggest adding additional comments on daphnia, sea urchin and C. elegans. In particular, several labs work with the latter organism, including J. Freedman at NIEHS."*

– *"Several other 3-D germ cell assays are available, and these hold promise for reducing the number of animals. They are not regarded as full animal-free alternatives, however. These have recently been published by Yu et al. (2009, 2010, 2011)."*

– *"Missing important papers related to … work developed by the ReProTect program related to the human placental transport studies, also discussing these studies as supplementary to other studies and beneficial by providing human data (Mathiesen et al., 2010; Myllynen et al., 2010)."*

– *"Additional references on the concordance of WEC and other assays using transcriptomics can be identified in several other papers by Robinson et al. (2011)."*

– *"In silico methods are important for 'predictive toxicology,' although the 2011 report states, 'it is widely acknowledged that reproductive effects are among the most difficult endpoints to predict in silico.' A number of resources have appeared in the last few years that may help overcome at least some of this difficulty. These include toxicological databases for chemical structures and properties, high-throughput screening (HTS) data from in vitro assays, and legacy data from in vivo animal tests."*

– *"What the paper did not cover were the more integrated 'Toxicity Testing in the 21st Century (TT21C)' approaches as envisioned by a landmark NAS report issued in 2007. Instead, the Adler paper clearly took an endpoint by endpoint approach. Perhaps the choice not to discuss TT21C was because it was a new and somewhat controversial topic, and certainly much more research needs to be done before such approaches are ready for cosmetics testing. In terms of the charge question and the EU Cosmetics Directive, the lack of TT21C coverage is not an issue. However, … I believe this is where the field of alternatives needs to move."*

– *"The 2011 report provides relevant web links and cites papers into 2009-2010 and thus does not have the latest perspective. For example, EPA's ToxRefDB database, which puts $ 2-billion of legacy animal testing data into a computable format using standardized terminologies and ontologies, was described in the report, but only discussed from the initial paper (Martin et al., 2009a) focusing on chronic/cancer studies. Since the*

2011 report was written, ToxRefDB studies were published on multigenerational reproductive toxicity in rats (Martin et al., 2009b) and prenatal developmental toxicity studies in rats and rabbits (Knudsen et al., 2009). These ToxrefDB studies digitize thousands of traditional animal studies on hundreds of chemicals."

– *"The 2011 report states that the 'current status of in silico methods for reproductive toxicity reflects the problems with modeling this endpoint' and that these ambitious approaches 'may not account for the subtlety in the mechanisms, such as time-dependence and receptor binding effects.' The report has focused on more traditional in silico methods based on information from chemical structure-activity relationships (SAR). Recent dossiers from REACH legislation (2000-2009) provide information to regulators based on read-across methods; however, these approaches do not perform well for predictive toxicology of developmental and reproductive endpoints due to few well-defined structural alerts."*

– *"Newer studies from EPA's ToxCast research program in computational toxicology have demonstrated the feasibility of predictive modeling of fertility in rats (Martin et al., 2011), blood vessel development (Kleinstreuer et al., 2011) and prenatal developmental toxicity (Sipes et al., 2011). These predictive models have demonstrated the feasibility of predicting ToxRefDB animal toxicity solely from the results of in vitro data from >500 ToxCast HTS assays. This corroborates the statement in the 2011 report, 'the classical aim of one-to-one replacement of in vivo protocols by alternative tests is clearly not feasible for the complex reproductive and developmental toxicity animal study protocols.' That said, the aforementioned studies also corroborate what the report calls for in terms of an array of alternative assays that, in combination, reproduce sensitive endpoints. The aforementioned studies (currently in press or under review) clearly demonstrate the feasibility of predictive modeling of reproductive and developmental toxicity for a set of 309 chemicals."*

– *"Few non-animal testing methods are scientifically acceptable for regulatory use. As such, the current path forward is to consider building blocks in an 'integrated testing strategy' (ITS). In general, the ITS strategy represents an extension of methodologies for efficacy and safety assessment. These assays would typically tune the predictive capacity to high specificity in order to avoid false positives (e.g., development of a false lead) for preclinical screening; however, the regulatory toxicology application in safety assessment would more likely tune the predictive capacity of an ITS to high sensitivity to avoid false negatives (e.g., not to miss a hazard). The 2011 report describes the various methodologies but is not as clear on the line drawn between these applications."*

8.4 Do you agree with the conclusions for the area? If not, where do you disagree?

To a large extent, the reviewers agreed with the conclusions:
– *"In general, the conclusions are clear and are based on the facts provided in this status report. Some conclusions are quite*

general and could benefit from some more detailed explanations, milestones, and timelines. While the conclusion that it will take more than 10 years to complete the development of additional alternatives seems reasonable, more arguments and proposals could be provided, in addition to the timelines up to pre-validation as provided in Table 11, in terms of the overall strategy and changes in the current thinking process and milestones. More details on what specifically is needed to further define the role of alternative methods as part of integrated testing strategies for regulatory toxicity evaluations and what these milestones and timelines could look like would be beneficial to this report."

– "Again, this reviewer would reiterate agreement with the conclusion of this area report on reproductive toxicology. In addition to the comments given above, this area report also addressed the increased availability of comprehensive and searchable databases that inform reproductive toxicology. It should be noted that past efforts in validating alternative assays for reproductive toxicology have failed solely due to the lack of comparison databases for this complex series of endpoints. The report could emphasize the significance of these efforts to meeting its future goals."

– "Again, reproductive and developmental toxicology is likely to be one of the most challenging areas for incorporation of alternatives. Unfortunately, it is also the area that uses the greatest number of animals. To this point, I was glad to see that the new Extended One-Generation Reproductive Toxicity Study was discussed as an option for greatly reducing the number of animals used, while at the same time obtaining more information than ever before on postnatal functional development."

– "The section on reproductive toxicity closes by saying that testing strategy will be the future driving force. I couldn't agree with this more, as this is where the measureable progress is likely to come from in the future."

8.5 Further advice to the area

A number of suggestions were made:
– "The authors should be encouraged to provide an even more visionary evaluation of potential testing strategies in reproductive toxicology, develop a tiered safety assessment strategy, consider even more interaction with other areas (e.g., information on reproductive toxicity generated in repeated dose toxicity and how this can be best utilized, information generated vs. needed in toxicokinetics), and what specifically (including milestones, timelines) might be required. For this, the statement made in the conclusion of this report is very valuable (retrospective analyses to select critical endpoints and towards the definition of novel testing paradigms). It may be necessary to re-define what is needed for a high quality safety assessment and what could potentially be omitted as not toxicologically relevant based on the available experience with cosmetics, a re-designed inventory of the alternative methods currently available, and what their specific purpose is in this context, a gap analysis, and a redesigned testing strategy, considering high quality safety assessments for cosmetic

products, rather than the one-to-one replacement of OECD guidelines. This report could be the platform to initiate this important process."

– "REACH requires companies producing or selling chemicals in the EU to register toxicity data and tests to clarify biological activity, especially with respect to adverse effects on reproduction and development. For cosmetic products, a quantitative risk assessment is required for each single ingredient in a product. The 2011 report refers to the SCCS guidelines stating animal testing 'may become necessary' for reproductive toxicity depending on preconditions such as kinetics and acute toxicity, and that guidelines are less clear on what tests should be performed. Recent dossiers from 2000-2009 suggest that a prenatal rat study (OECD 414) has been the minimum information provided to regulators and considered sufficient. A very recent paper by Gilbert (2011) corroborates this and further states that "many dossiers rely heavily on old data and fail to suggest new tests, and that few include any mention of non-animal testing methods…""

– "The problem of alternatives for reproductive and developmental toxicity is not simply the lack of an adequate in vitro tool box. Because of the enormity and complexity of the reproductive cycle and all of the key events within, a tool box of in vitro methods that could cover all phases of the reproductive cycle would be far too large to carry around! As the authors mention, we need to shift our attention from one-to-one replacements of in vivo end points with in vitro assays, to a rethinking of testing strategies. This is where the big gains in alternatives are likely to be found within the next 5 to 10 years."

– "Although the long-range goal of replacing animal tests with non-animal alternatives for reproductive toxicity seems to more than 10 years away, Adler et al. briefly mentioned some opportunities for the use of alternatives in priority-setting and screening. While not under the strict purview of the regulatory process, this application of alternatives can be used in-house by industry or government labs as part of a larger integrated testing strategy. This point could have been given greater emphasis in the Adler et al. paper, as it is one of the few areas where reproductive and developmental alternatives can be used right now."

– "Improvements in testing paradigms also could be achieved by investing in improved human exposure models. The number of animals used for reproductive and developmental toxicity testing could be dramatically reduced if internal estimates of human exposure played more heavily into testing strategy and regulation. While not traditionally considered under alternatives, areas such as exposure research could and should play a large role in the advancement of non-animal alternatives."

– "One typographical item of note is that 'sperm' is plural without adding an 's'."

– "Again, the problem of alternatives for reproductive toxicity is not simply about the size of one's toolbox of in vitro assays. Even if enough in vitro assays were available to address all of the key processes in the reproductive cycle, integration of these processes would be lacking; not to mention the huge number of assays it would take to reconstitute the entire

reproductive cycle. Rather than trying to invent an in vitro *version of what currently exists, we should be thinking about redesigning toxicity testing strategies that are fit for 21st century purposes."*

– *"Several new assay platforms are amenable to a systems- or pathway-based approach, and these range from panels of human cells in culture to model organisms such as zebrafish embryo and the roundworm,* C. elegans. *As we move forward in developing non-animal alternatives for reproductive and developmental toxicity, I think it is these cross-cutting platforms that hold the greatest promise."*

9 Overall Conclusions

This independent review endorses to a large extent the conclusions of the Adler et al. (2011) report. The addition of non-European experts and further independent experts added a number of approaches that were missing or not sufficiently addressed. This does not affect the assessment of the status of the different areas with regard to full replacement of animal tests; it does, however, broaden the basis of methods and approaches to be considered for a roadmap for future research and development. The provenance of reviewers from different parts of the world did not lead to evident differences in view with two exceptions: Non-Europeans tended to recommend refinement and reduction methods, i.e. better animal tests, which do not represent an option in the EU cosmetics legislation. Stronger reference was made to the pathway-based approaches that dominate the US discussion.

Although there is general agreement on the time frames for complete replacement, there are also many methods that are just about ready for use and are being used to collect data and to evaluate the completeness of possible replacement (e.g., sensitization and toxicokinetics). Some of these methods may clear the current barriers in a time frame considerably shorter than 5 to 7 years. This does not include formal validation and acceptance, but it does represent a realistic view of the use of these for decision-making.

Throughout, it was noted that this important stocktaking must be furthered now to strategic planning and a roadmap. Both the original report and this review acknowledge the progress made in all areas discussed here. The lack of full replacements at this point is not a reflection of the efforts spent but rather of the sheer size of the challenge and the time science needs to develop new approaches and undergo transition. If we do not lessen these efforts, and follow the many suggestions to a more strategic combination of assays and further development of the methods that are almost there, the ultimate goal of animal-free testing of cosmetic ingredients can become a reality.

References

Adler, S., Basketter, D., Creton, S., et al. (2011). Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. *Arch. Toxicol. 85*, 367-485.

Aydin, S., Signorelli, S., Lechleitner, T., et al. (2008). Influence of microvascular endothelial cells on transcriptional regulation of proximal tubular epithelial cells. *Am. J. Physiol. Cell. Physiol. 294*, C543-54.

Carfi', M., Gennari, A., Malerba, I., et al. (2007). In vitro tests to evaluate immunotoxicity: a preliminary study. *Toxicology 229*, 11-22.

Corsini, E., and Roggen, E. L. (2009). Immunotoxicology: opportunities for non-animal test development. *ATLA 37*, 387-397.

Delarue, F., Virone, A., Hagege, J., et al. (1991). Stable cell line of T-SV40 immortalized human glomerular visceral epithelial cells. *Kidney Int. 40*, 906-912.

Emter, R., Ellis, G., and Natsch, A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. *Toxicol. Appl. Pharmacol. 245*, 281–290.

Eskes, C., and Zuang, V. (2005). Alternative (non-animal) methods for cosmetic testing: current status and future prospects. *ATLA 33, Suppl. 1*, 1-228.

European Commission (2003). Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products. *Official J. Eur. Union of 11.3.2003, L66*, 26-35.

European Commission (2009). Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. *Official J. Eur. Union of 30.11.2009, L342*, 59-209.

Gennari, A., Ban, M., Braun, A., et al. (2005). The use of in vitro systems for evaluating immunotoxicity: The report and recommendations of an ECVAM workshop. *J. Immunotoxicol. 2*, 61-83.

Galbiati, V., Mitjans, M., and Corsini, E. (2010). Present and future of in vitro immunotoxicology in drug development. *J. Immuntoxicol. 7*, 255-267.

Gerberick, G. F., Vassallo, J. D., Foertsch, L. M., et al. (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. *Toxicol. Sci. 97*, 417-427.

Gilbert, N. (2011). Data gaps threaten chemical safety law. *Nature 475*, 150-151.

Hartung, T., Bremer, S., Casati, S., et al. (2003). ECVAM's response to the changing political environment for alternatives: Consequences of the European Union chemicals and cosmetics policy. *ATLA 31*, 473-481.

Hartung, T. (2008). Food for thought … on alternative methods for cosmetics safety testing. *ALTEX 25*, 147-162.

Hayashi, M., Kamata, E., Hirose, A., et al. (2005) In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. *Mutat. Res. 588*, 129-135.

Hochstenbach, K., van Leeuwen, D. M., Gmuender, H., et al. (2010). Transcriptomic profile indicative of immunotoxic exposure: in vitro studies in peripheral blood mononuclear cells. *Toxicol. Sci. 118*, 19-30.

Kleinstreuer, N. C., Judson, R. S., Reif, D. M., et al. (2011). Environmental impact on vascular development predicted by high throughput screening. *Environ. Health. Perspect.,* Epub ahead of print.

Knudsen, T. B., Martin, M. T., Kavlock, R. J., et al. (2009). Profil-

ing the activity of environmental chemicals in prenatal developmental toxicity studies using the U.S. EPA's ToxRefDB. *Reprod. Toxicol. 28*, 209-219.

Koeper, L. M., and Vohr, H. W. (2009). Functional assays are mandatory for correct prediction of immunotoxic properties of compounds in vitro. *Food Chem. Toxicol. 47*, 110-118.

Langezaal, I., Coecke, S., and Hartung, T. (2001). Whole blood cytokine response as a measure of immunotoxicity. *Toxicol. In Vitro 15*, 313-318.

Langezaal, I., Hoffmann, S., Hartung, T., and Coecke, S. (2002). Evaluation and prevalidation of an immunotoxicity test based on human whole-blood cytokine release. *ATLA 30*, 581-595.

Lankveld, D. P. K. , Van Loveren, H., Baken, K. A., and Vandebriel, R. J. (2010). In vitro testing for direct immunotoxicity: state of the art. In Dietrich, D. (ed.) *Immunotoxicity Testing, Methods in Molecular Biology, 2010, Volume 598, Part 7*, (401-423) Humana Press, New York, USA.

Martin, M. T., Judson, R. S., Reif, D. M., et al. (2009a). Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. *Environ. Health Perspect. 117*, 392-399.

Martin, M. T., Mendez, E., Corum, D. G., et al. (2009b). Profiling the reproductive toxicity of chemicals from multigeneration studies in the toxicity reference database. *Toxicol. Sci. 110*, 181-190.

Martin, M. T., Knudsen, T. B., Reif, D. M., et al. (2011). Predictive model of rat reproductive toxicity from ToxCast high throughput screening. *Biol. Reprod. 85*, 327-339.

Mathiesen, L., Mose, T., Mørck, T. J., et al. (2010). Quality assessment of a placental perfusion protocol. *Reprod. Toxicol. 30*, 138-146.

McKim, J.M. (2010). Building a tiered approach to in vitro predictive toxicity screening: a focus on assays with in vivo relevance. *Comb. Chem. High Throughput Screen. 13*, 188–206.

Mitjans, M., Viviani, B., Lucchi, L., et al. (2008). Role of p38 MAPK in the selective release of IL-8 induced by chemical allergen in naive THP-1 cells. *Toxicol. In Vitro 22*, 386-395.

Myllynen, P., Mathiesen, L., Weimer, M., et al. (2010). Preliminary interlaboratory comparison of the ex vivo dual human placental perfusion system. *Reprod. Toxicol. 30*, 94-102.

Nakanishi, T., Turner, R. J., and Burg, M. B. (1989). Osmoregulatory changes in myo-inositol transport by renal cells. *Proc. Natl. Acad. Sci. 86*, 6002-6006.

Natsch, A., and Emter, R. (2008). Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitization potential of chemicals. *Toxicol. Sci. 102*, 110-119.

Nguyen, G., Delarue, F., Berrou, J., et al. (1996). Specific receptor binding of renin on human mesangial cells in culture increases plasminogen activator inhibitor-1 antigen. *Kidney Int. 50*, 1897-1903.

Paules, R. S., Aubrecht, J., Corvi, R. (2011). Moving forward in human cancer risk assessment. *Environ. Health Perspect. 119*, 739–743.

Peyret, T., Poulin, P., and Krishnan, K. (2010). A unified algorithm for predicting partition coefficients for PBPK modeling of drugs and environmental chemicals. *Toxicol. Appl. Pharmacol. 249*, 197-207.

Pfuhler, S., Kirkland, D., Kasper, P., et al. (2009). Reduction of use of animals in regulatory genotoxicity testing: Identification and implementation opportunities – Report from an ECVAM workshop. *Mut. Res. 680*, 31-42.

Ringerike, T., Ullerås, E., Völker, R., et al. (2005). Detection of immunotoxicity using T-cell-based cytokine reporter cell lines ("Cell Chip"). *Toxicology 206*, 257-272.

Robinson, J. F., Theunissen, P., van Dartel, D., et al. (2011). Comparison of MeHg-induced toxicogenomic responses across in vivo and in vitro models used in developmental toxicology. *Reprod. Toxicol.*, in press.

Sakaguchi, H., Ashikaga, T., Miyazawa, M., et al. (2009). The relationship between CD86/CD54 expression and THP-1 cell viability in an in vitro skin sensitization test--human cell line activation test (h-CLAT). *Cell. Biol. Toxicol. 25*, 109-126.

Sarrab, R. M., Lennon, R., Ni, L., et al. (2011). Establishment of conditionally immortalised human glomerular mesangial cells in culture, with unique migratory properties. *Am. J. Physiol. Renal Physiol.*, Epub ahead of print.

Sipes, N. S., Martin, M. T., Reif, D. M., et al. (2011). Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. *Toxicol. Sci.*, in press.

Verwei, M., van Burgsteden, J. A., Krul, C. A., et al. (2006). Prediction of in vivo embryotoxic effect levels with a combination of in vitro studies and PBPK modelling. *Toxicol. Lett. 165*, 79-87.

Wieser, M., Stadler, G., Jennings, P., et al. (2008). hTERT alone immortalizes epithelial cells of renal proximal tubules without changing their functional characteristics. Am. *J. Physiol. Renal. Physiol. 295*, F1365–F1375.

Yu, X., Hong, S., Moreira, E. G., and Faustman, E. M. (2009). Improving in vitro Sertoli cell/gonocyte co-culture model for assessing male reproductive toxicity: Lessons learned from comparisons of cytotoxicity versus genomic responses to phthalates. *Toxicol. Appl. Pharmacol. 15*, 325-336.

Yu, X., Robinson, J. F., Sidhu, J. S., et al. (2010). A system-based comparison of gene expression reveals alterations in oxidative stress, disruption of ubiquitin-proteasome system and altered cell cycle regulation after exposure to cadmium and methylmercury in Mouse Embryonic Fibroblast (MEF). *Toxicol. Sci. 114*, 356-377.

Yu, X., Sidhu, J. S., Hong, S., et al. (2011). Cadmium induced p53-dependent activation of stress signaling, accumulation of ubiquitinated proteins, and apoptosis in mouse embryonic fibroblast cells. *Toxicol. Sci. 120*, 403-412.

Zuang, V., and Hartung, T. (2005). Making validated alternatives available – the strategies and work of the European Centre for the Validation of Alternative Methods (ECVAM). *AATEX 11*, 15-26.

Correspondence to

Thomas Hartung, MD PhD
Johns Hopkins Bloomberg School of Public Health
Center for Alternatives to Animal Testing
615 N. Wolfe St.
Baltimore, MD, 21205, USA
e-mail: THartung@jhsph.edu