Consensus Report on the Future of Animal-Free Systemic Toxicity Testing

Marcel Leist  
University of Konstanz

Nina Hasiwa  
University of Konstanz

Costanza Rovida  
University of Konstanz

Mardas Daneshian  
University of Konstanz

David Basketter  
DABMEB Consultancy Ltd

See next page for additional authors

Follow this and additional works at: https://www.wellbeingintlstudiesrepository.org/acwp_arte

Part of the Bioethics and Medical Ethics Commons, Laboratory and Basic Science Research Commons, and the Research Methods in Life Sciences Commons

Recommended Citation
Leist, M., Hasiwa, N., Rovida, C., Daneshian, M., Basketter, D., Kimber, I., ... & Rossi, A. M. (2014). Consensus report on the future of animal-free systemic toxicity testing. Alternatives to Animal Experimentation: ALTEX, 31(3), 341-356.

This material is brought to you for free and open access by WellBeing International. It has been accepted for inclusion by an authorized administrator of the WBI Studies Repository. For more information, please contact wbisr-info@wellbeingintl.org.
Authors
Marcel Leist, Nina Hasiwa, Costanza Rovida, Mardas Daneshian, David Basketter, Ian Kimber, Harvey Clewell, Tilman Gocht, Alan Goldberg, Francois Busquet, Anna-Maria Rossi, Michael Schwarz, Martin Stephens, Rob Taalman, Thomas B. Knudsen, James McKim, Georgina Harris, David Pamies, and Thomas Hartung

This report is available at WBI Studies Repository: https://www.wellbeingintlstudiesrepository.org/acwp_arte/107
Summary
Since March 2013, animal use for cosmetics testing for the European market has been banned. This requires a renewed view on risk assessment in this field. However, in other fields as well, traditional animal experimentation does not always satisfy requirements in safety testing, as the need for human-relevant information is ever increasing. A general strategy for animal-free test approaches was outlined by the US National Research Council’s vision document for Toxicity Testing in the 21st Century in 2007. It is now possible to provide a more defined roadmap on how to implement this vision for the four principal areas of systemic toxicity evaluation: repeat dose organ toxicity, carcinogenicity, reproductive toxicity and allergy induction (skin sensitization), as well as for the evaluation of toxicant metabolism (toxicokinetics) (Fig. 1). CAAT-Europe assembled experts from Europe, America and Asia to design a scientific roadmap for future risk assessment approaches and the outcome was then further discussed and refined in two consensus meetings with over 200 stakeholders. The key recommendations include: focusing on improving existing methods rather than favoring de novo design; combining hazard testing with toxicokinetics predictions; developing integrated test strategies; incorporating new high content endpoints to classical assays; evolving...
1 Introduction and background

The discussion leading to this summary report started with the 7th Amendment to the Cosmetics Directive 76/768/EEC, which called for a complete ban of testing on vertebrate animals for the toxicological characterization of cosmetics ingredients in 2013. The European Commission asked experts to evaluate the availability of alternative non-animal methods. Their conclusion that alternative methods would not be available during the next 10 years (Adler et al., 2011) met with some criticism (Taylor et al., 2011), but was mostly endorsed by an independent expert group invited by CAAT also including specialists from Japan and the USA (Hartung et al., 2011). They also noted that significant advances had been made in the time between the publication of the Adler Report and the evaluation by the international group of experts. The next step, i.e., assembling experts to create a perspective for the future, was initiated by CAAT-Europe in a series of commissioned white papers on sensitization, repeated dose organ toxicity, toxicokinetics, carcinogenicity and reproductive toxicity. Importantly, this work addressed a broad range of chemical testing, including also the fields of drugs, pesticides and industrial chemicals in addition to cosmetics ingredients. A workshop with 35 experts discussed these white papers. This activity resulted in the extensive report A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing (Basketter et al., 2012). To involve all potential stakeholders, this

---

**Fig. 2 Timeline of events leading to this public expert consultation report**

The 7th amendment of the European Cosmetics Directive required the phasing out of animal testing to be completed by 2013. The European Commission evaluated the availability of non-animal methods and the outcome was published (Adler et al., 2011). The conclusions of that report were confirmed by independent experts (Hartung et al., 2011). In order to outline a roadmap for further development of non-animal methods for addressing systemic toxicity, an expert consortium was convened in a series of CAAT-Europe workshops to elaborate on the issue. As a result, an extensive report on the roadmap for non-animal methods for systemic toxicity testing was published (Basketter et al., 2012) and presented for a public expert consultation in 2012 in Brussels and in 2013 in Washington at an FDA-hosted event. The present report summarizes the recommendations resulting from the public expert consultation in which over 200 experts from academia, industry and regulatory authorities were involved. Grey boxes refer to actions taken by CAAT/CAAT-Europe.

---

1 http://ec.europa.eu/consumers/sectors/cosmetics/files/doc/antest/(2)_executive_summary_en.pdf
The roadmap was presented in March 2012 in Brussels\(^2\) by several experts in front of about 200 stakeholders from governmental organizations, academia, industry and NGOs from all over the world (Fig. 2). A second workshop, *Scientific roadmap for the future of animal-free systemic toxicity testing*, similar in size and scope, was organized in Washington at the FDA in May 2013 to give updates on the Basketter report and scientific advances in the fields\(^3\).

Each of the lectures on the five major fields still requiring better non-animal safety testing methods (Fig. 1) was followed by one hour of intensive discussion to consolidate or improve the suggested strategies. Here we report the final outcome. This roadmap is expected to pave the way for a new toxicology that can better predict the effect of chemicals on humans, using fewer or even no vertebrate animals.

2 Animal models

A successful strategy to replace animal testing must take its starting point from the analysis of the current technology. The weaknesses of animal testing could then be avoided by the new approach (Fig. 3). It must be noted that rational comparisons are made difficult by the fact that almost all of the currently used animal models have never been formally validated. The rationale of their use is therefore not based on scientific data (Hartung and Leist, 2008).

Some of the problems related to animal models derive from the high doses that are tested and the multiplicity of endpoints that are measured. Experts claim, that “Half of all chemicals, whether natural or synthetic, are positive in high-dose rodent cancer tests. These results are unlikely to be relevant at the low doses of human exposure.” (Ames and Gold, 2000). Many of the attendees expressed the opinion that widespread knowledge on the limited value of certain animal studies frequently contributes to the decision by authorities to waive testing.

The lack of predictivity of animal models is particularly apparent from the field of drug development (Leist and Hartung, 2013). Only 8% of drugs entering clinical phase I (first human dose) gain approval by authorities and half of them fail in phase III\(^4\). US Health and Human Services Secretary Mike Leavitt commented that “currently, nine out of ten experimental drugs fail in clinical studies because we cannot accurately predict how they will behave in people based on laboratory and animal studies.”\(^5\)

The conclusions are easily drawn: We need human-predictive, rapid and economical methods to evaluate whether or not a compound, no matter if chemical, drug or cosmetic ingredient, is safe for intended human use.

3 Future safety science and pathways of toxicity

A key step in the paradigm shift in toxicology, as far as regulatory authorities are concerned, was the 2007 US National Research Council report *Toxicity Testing in the 21st Century – a Vision and a Strategy* (NRC, 2007; Leist et al., 2008a). It promoted the idea that the number of ways that a chemical or drug could disturb a cell is finite and can therefore be identified by appropriate screening methods. Quantitative information on the concentration-dependence of such disturbances can be used to predict the overall network of cellular regulatory reactions.

| Problems with animal experiments |          |
|---------------------------------|----------|
| High false negatives rate | → compromised safety |
| High false positives rate | → economically not acceptable |
| Lack of validation | → unclear scientific basis |
| Low predictivity and high costs | → waiving praxis |
| Frequent waiving | → compromised safety |
| High dose – low dose extrapolation | → uncertainty → compromised safety |
| Metabolism different from man | → uncertainty → compromised safety |
| Biology different from man | → uncertainty → compromised safety |

\(^2\) The forum was co-organized by Center for Alternatives to Animal testing (CAAT), Cosmetics Europe, Doerenkamp-Zbinden Foundation, European Chemical Industry Council (CEFIC), European Consensus Platform for Alternatives (ECOPA), European Society for Alternatives to Animal Testing (EUSAAAT), European Society for Toxicology in vitro (ESTIV), US EPA ToxCast\(^\text{TM}\), Humane Society International (HSI), Institute for In-vitro Sciences (IVIS) and In-vitro Testing Industrial Platform (IVTIP).

\(^3\) The forum was co-organized by Agilent Technologies, American Cleaning Institute (ACI), Animal Research & Development Foundation (ARDF), American Society for Cellular and Computational Toxicology (ASCCCT), Center for Alternatives to Animal testing (CAAT), CropLife Canada, CropLife America, FDA Center for Food Safety and Advanced Nutrition (CFSAN), Grocery Manufacturer Association (GMA), Human Toxicology Project Consortium, Humane Society of the United States (HSUS), Institute for In-vitro Sciences (IVIS), National Institute for Environmental Health Sciences (NIEHS), People for the ethical Treatment of Animals (PeTA) and The Hammer Institute.

\(^4\) http://www.the-scientist.com/?articles.view/articleNo/15910/title/More-Compounds-Failing-Phase-I/

\(^5\) http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2006/ucm108576.htm

Fig. 3: Problems with animal experiments

To assess the hazard posed by substances humans are exposed to, all available approaches need to be evaluated for their usefulness. The present system of animal testing needs critical evaluation of its predictive power for human safety. The limitations of animal testing, which is often considered the “gold” standard, may compromise human safety and pose an economic threat. Under such conditions, its ethical acceptability is also doubtful.
4 General forward strategies

While discussing the five specific toxicological domains, the experts identified issues relevant to all areas of in vitro methods. Joint knowledge management and sharing of expertise between different sectors, stakeholders and application domains were identified as important drivers for accelerated progress, in addition to accelerated test establishment and validation, and better use of computational toxicology methods (Fig. 6).

Sharing of data would hopefully lead to the creation of a human safety database, bigger and more complete than existing ones (e.g., OPENTOX, EPA, TOXBANC, IMI activities). Most importantly, it would be more accessible and should be tailored for in vitro-in vivo comparisons as well as data mining by non-specialists in bioinformatics. The focus would not be on data collection as such, but on the accessory information linked to the primary outcome data as far as mechanisms of toxicity are concerned. As there are major hurdles (e.g., intellectual property rights and industrial competitiveness issues) to be overcome, it is clear that substantial incentives must be granted to encourage industry to share their proprietary data (Fig. 7).
None of the future challenges in the field of in vitro toxicology can be addressed by individual test systems. The solution will rather lie in the construction of batteries of tests to be combined in integrated testing strategies (ITS). This will need to be considered right from the conception of a test, throughout its development and especially during the evaluation of its performance. The latter evaluation needs to consider the test alone, but also in the context of the added value it brings to a test battery. More research and experimentation is required on how to build ITS. One example of a flexible, yet fully quantitative approach, is the Bayesian network (McDowell and Jaworska, 2009).

**Fig. 5: Vision of a smooth transition from current to future toxicology in safety science**

It is envisaged that the types of test systems employed will change over the time course of the establishment of a new safety science. At present, complex test systems that are specific for organ functions and developmental stages are preferentially used. Only few programs use simple assays of elementary biochemical and cellular function (e.g., ToxCast™ Program). Over time, more and more critical biomarkers of toxicity may be identified by the application of HCS (high-content screening) and omics technologies to the complex systems, and simple test systems may suffice to measure key processes (Rossini and Hartung, 2012). Case studies, e.g., from PBPK (physiologically based pharmacokinetic modeling) and skin sensitization fields, could be used as learning models for the transition. The principles of evidence-based toxicology and the resulting quality control will lead to an accelerated method development and validation. Over time, the goal is to shift from the present deterministic risk classification to a probabilistic risk assessment.

**Fig. 6: General points to consider when moving forward towards new approaches for systemic toxicity testing**

- **Collaboration**
  - Join forces: academia, industrial sectors, regulators
  - Render valuable proprietary toxicological data accessible
  - Generation of curated and searchable data base

- **Test systems**
  - Develop better in vitro methods
  - Optimize existing test systems
  - Develop strategies to establish, validate and use integrated testing strategy
  - Abolition of useless in vivo tests

- **Computational toxicology**
  - Use of systems biology approaches
  - Define pathways-of-toxicity and bring to practical application
  - Reduce complexity by focusing on key events
  - Use information-rich endpoints as biomarkers-of-toxicity (BoT)
  - Broad use of quantitative in vitro → in vivo extrapolation (qVIVE)

**Fig. 7: Creation of a high quality database for relating in vivo and in vitro information**

The key players (pharmaceutical, chemical and cosmetics industry, basic research and regulators) share common goals that are of high value to them. The benefit of working together should outweigh disadvantages (opening of proprietary databases). This would allow the generation of a large, high quality database of in vitro toxicity data. It should be publicly available, include rich data that informs on the mode of action of compounds and allows for in vitro – in vivo correlations. It should also be quality controlled and suitable for case studies. This can be achieved by joint projects and the common use of legacy data from hitherto proprietary in-house databases. The collection of human data by micro-dosing, from clinical trials and from epidemiological studies plays a major role.
Alternative in vitro methods have been developed for all toxicological questions, including even the most complex fields, ranging from developmental neurotoxicity (DNT) to xenobiotic metabolism (Adler et al., 2011; Basketter et al., 2012; Leist et al., 2012a, 2008b; van Thriel et al., 2012; Smirnova et al., 2014; Taylor et al., 2011). Many of these methods should be formally validated for immediate use, or they could form the basis for accelerated further development. Optimization of existing systems is an important part of the strategy to accelerate the implementation of a mostly animal-free safety science, in addition to the more time demanding development of entirely new methods (Fig. 9). One specific way to improve available tests is the incorporation of highly information-rich endpoints provided by omics technologies.

**Fig. 8: Example for the use of Bayesian networks in the establishment of integrated testing strategies**

LLNA (local lymph node assay) potency prediction is used here as an example from the area of skin sensitization. Information from different assays (circles) is fed into the network. The dimension of the circles represents mutual information values; the length of the arrows has no mathematical correlation. For instance, information can be obtained on how the in vivo outcome (LLNA: local lymph node assay) is predicted by physicochemical compound properties (such as molecular weight or lipophilicity (K<sub>OW</sub>)), biological assays (e.g., GARD assay or dendritic cell assay) and peptide reactivity measures (DPRA: direct peptide reactivity assay). The advantage of the approach is that it can be coupled to other networks or other assays, as they are desired and become available. The original paper (Jaworska and Hoffmann, 2010) contains all the details on the background.

**Fig. 9: Strategies to improve in vitro test systems**

Test systems that already have been developed can still improve in quality and robustness to arrive relatively quickly at predictive test systems fit for regulatory use. A list of features to be considered has been compiled here.

**Fig. 10: Overview of different omics technologies that can inform on chemicals’ adverse outcome pathways and underlying modes of action**

Omics technologies provide data-rich endpoints. The biological information flow in a cell leads from gene sequences (the code) via RNA (the messages) to enzymes and other functional proteins (the tools). Within this infrastructure small molecule metabolites may be regarded as the goods that are produced and traded. They comprise energy substrates, building blocks and signaling messengers. As there are feedback loops between all levels, the different omics technologies address these four organization levels. The disturbance of a cell by chemicals may be measured by any single technique. Combinations of more than one approach lead to a better prediction of the true human situation.
eral mechanisms of sensitization are well-defined. There are three well-developed animal models (Buehler Guinea Pig Test, Guinea Pig Maximization Test (GPMt) and Mouse Local Lymph Node Assay (LLNA)) currently used to identify chemicals with toxic potential. The LLNA, which is already a step towards refinement and reduction of the use of animals, is the preferred method for safety assessment as it provides a quantitative value (the concentration of the chemical which causes a threshold positive response (EC3)) that can determine the potency of the sensitizer.

Already over a dozen different in vitro tests to identify sensitisers have been submitted to the European Union Reference Laboratory for Alternatives to Animal testing (EURL ECVAM). Currently, two of these are validated for risk assessment. These are the direct peptide reactivity assay (DPRA), based on the chemical understanding and correlation with sensitization, and the human cell line activation test (h-CLA t), based on the activation of dendritic-like cells (Bauch et al., 2011; Sakaguchi et al., 2006; Ashikaga et al., 2010). The KeratinoSense™ luciferase-reporter gene model (based on the anti-oxidant response element in the HaCat keratinocyte cell line) (Natsch, 2010; Andreas et al., 2011) has already been validated by Givaudan and is accepted by OECD. Integrated testing strategies (ITS) will be

6 Specific approaches for the five toxicological endpoints still lacking validated replacement methods

A detailed strategy has been elaborated for each endpoint and described in detail (Basketter et al., 2012). The consensus meeting of the roadmap initiative highlighted specific points for immediate attention and action:

6.1 Skin Sensitization

Although the sensitization process is a disease-free state, subsequent exposures can lead to allergic contact dermatitis, the most common adverse effect of chemicals on human health. One in 5 adults suffers contact allergy to one chemical or another (Peiser et al., 2012). This area differs from others as a large amount of human data is available. Moreover, the general mechanisms of sensitization are well-defined. There are three well-developed animal models (Buehler Guinea Pig Test, Guinea Pig Maximization Test (GPMt) and Mouse Local Lymph Node Assay (LLNA)) currently used to identify chemicals with toxic potential. The LLNA, which is already a step towards refinement and reduction of the use of animals, is the preferred method for safety assessment as it provides a quantitative value (the concentration of the chemical which causes a threshold positive response (EC3)) that can determine the potency of the sensitizer. 

Already over a dozen different in vitro tests to identify sensitisers have been submitted to the European Union Reference Laboratory for Alternatives to Animal testing (EURL ECVAM). Currently, two of these are validated for risk assessment. These are the direct peptide reactivity assay (DPRA), based on the chemical understanding and correlation with sensitization, and the human cell line activation test (h-CLA t), based on the activation of dendritic-like cells (Bauch et al., 2011; Sakaguchi et al., 2006; Ashikaga et al., 2010). The KeratinoSense™ luciferase-reporter gene model (based on the anti-oxidant response element in the HaCat keratinocyte cell line) (Natsch, 2010; Andreas et al., 2011) has already been validated by Givaudan and is accepted by OECD. Integrated testing strategies (ITS) will be

Combination of such in vitro tests in an integrated strategy (ITS) would allow comprehensive predictions for unknown xenobiotics.
the way forward, as each assay on its own has 80% accuracy, but if combined in an ITS, 90% accuracy can be reached (Bauch et al., 2012). This level of predictivity would perform better than the validated LLNA. Thus, an ITS would fully replace the existing animal models.

Skin sensitization is a field in which several formally validated methods and ITS are expected to emerge in the near future. The reasons for this are the following: first, for skin sensitization the validation process has clear anchors: this is the only toxicological domain that is based on a formally validated animal test model (LLNA). Moreover, a large set of human data on positive control compounds is available, e.g., from diagnostic patch testing in dermatology clinics. Second, the mechanisms of skin sensitization are well understood, and the individual steps are amenable to modeling. Third, several in vitro models that seek to mimic each single step in the pathway are already available, and they now need only to be combined in an ITS. CAAT organized a workshop on ITS using the example of skin sensitization in June 2013 in Ranco, Italy; the respective report is currently being completed.

The application of the OECD-promoted concept of “adverse outcome pathways” (AOPs) to skin sensitization is relatively straightforward. Virtually all key events of the AOP already are covered by in vitro assays (Fig. 11)6. Despite this favorable situation, validation of a complete ITS for skin sensitization will require further work. It is, for instance, not yet clear how the individual tests that cover the steps of the AOP will be combined, including how much weight is given to the results of each assay and how the decision points of tiered testing would be structured. The final prediction model must be built as a whole on the assembly of tests and on the ITS rules linking them. The process of building and optimizing this overall test strategy is made difficult by the fact that the LLNA, even though it is one of the most advanced in vivo methods, can yield false-negative and false-positive results. Despite these weaknesses, and although human data are available as an alternative reference point, the LLNA is the only accepted reference for the determination of potency and for providing background data for ITS validation.

The conclusions on the status and roadmap for skin sensitization testing are as follows:

- Many non-animal methods for skin sensitization testing have been proposed and some of them have been/will be validated for the purpose of hazard identification. The development of non-animal methods for the evaluation of the relative skin sensitizing potency of contact allergens will require more work.
- Better measurements and tests for exposure are needed, and little is known about how to assess mixtures yet.
- Complications may arise when there is a need to test hydrophobic compounds or formulations as the proposed models may not be adequate. These problems must be tackled sooner rather than later and the applicability of each model should be assessed accordingly. This will provide opportunities for the development of other assays with other applicability domains.

- Many other areas of toxicology can follow skin sensitization as a good example where a detailed understanding of mechanisms can lead to the development of specific assays needed to identify compound toxicity.
- Computational models based on quantitative structure-activity relationship models (QSARs) provide promising tools to identify sensitizers, as the toxicity of the chemical is implicit in its structure. There have been major advances in QSAR models, although studies use data from the LLNA rather than human data and have difficulties in obtaining accuracy in models for “moderate” sensitizers (Li et al., 2007).

6 http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

62 Repeated dose toxicity

Repeated dose testing (RDT) consists in the evaluation of a chemical’s potential to cause chronic toxicity and organ-specific toxicities. Classically, tests for RDT are based on 4 (sub-acute toxicity), 13 (sub-chronic toxicity) and 26-102 (chronic toxicity) week rodent and non-rodent studies. Toxicity occurs after a chemical is absorbed into the general circulation. There is great concern about the relevance of these studies performed in animals for predicting human toxicity (Basketter et al., 2012; Chen et al., 2014; Hengstler et al., 1999; Olson et al., 2000). Different organizations (FDA, EPA, EMEA) and initiatives (REACH, TSCA and the EU Cosmetic Directive) are pushing in vitro methods in the chemical toxicity evaluation process.

RDT includes chronic adverse effects on major organs. On the one hand, the assessment of RDT requires lengthy in vivo experiments, which are difficult to model in vitro. On the other hand, inter-species differences can limit the usefulness of animal data for the prediction of human hazard in this area (Leist and Hartung, 2013). In vitro methods based on human cell lines may provide more human-relevant information (Pfaller et al., 2001). Biological models for different organs, e.g., liver, kidney, lung or brain, have been established, and new culture techniques, especially in form of 3D organoids, are expected to solve present issues about long-term culturing, absence of relevant inflammatory and immune cells (Hengstler et al., 2012) and availability of fully mature cell phenotypes. Stem cells, especially pluripotent stem cells, will be a major source of tissues and cells not available otherwise. Therefore, research on the generation of 2D cultures and 3D tissues from stem cells is of high importance. One of the approaches in this direction is the European SEURAT-1 project (following the long-term strategic target: “Safety Evaluation Ultimately Replacing Animal Testing”, http://www.seurat-1.eu). It started in 2011 with 50 million € joint funding from the European Commission and Cosmetics Europe and is focusing on the development of non-animal test systems in the field of repeat dose systemic toxicity following a case study approach based on the AOP concept. The Tox21 consortium and the US EPA’s ToxCast™ activity in the USA (Dix et al., 2007; Judson et al., 2010, 2014) as well as other activities in Europe and worldwide take similar or complementary approaches (NRC, 2007; Adler et al., 2011; Basketter et al., 2012; Judson et al., 2012; Leist et al., 2012b). Key to all these activities is the concept that most late (longer term)
effects of chemicals will be predicted from the early changes they cause in cellular signaling and regulation (Kleensang et al., 2014; Blaauboer et al., 2012). Therefore, signaling pathway identification and analysis is a crucial research necessity in toxicology, and very detailed quantitative information needs to be derived (Fig. 12) to use such data for systems biology modeling (Jennings et al., 2013; Krug et al., 2014; Hartung et al., 2012). Toxicogenomics technologies (Ramirez et al., 2013) are important tools that cover a multitude of cellular events. However, it is important to apply them to the right biological models. For instance, monocultures can hardly model the inflammatory responses frequently seen after long-term exposure to hazardous chemicals.

For repeated dose toxicity, two different approaches are taken in the development of alternative methods: (a) substitution of animals by a battery of relatively complex surrogate models that reflect important features of target tissues and organs. They often use ‘apical’ phenotypic endpoints (e.g., cell death markers) as readouts; (b) an integrated and tiered systems biology approach based on mechanistic endpoints and using the vast knowledge on biological regulation and homeostasis. Pathways-of-toxicity (PoT), emerging from such approaches, will guide hazard evaluation and risk assessment when combined with toxicokinetics modelling (Hartung and McBride, 2011; Boekelheide and Andersen, 2010). The two types of approaches may also be combined.

In summary, repeated dose toxicity will probably be the last method to be replaced, the use of PoT and new culture systems combined with new technologies and sharing of data on pharmaceutical case studies could be the opportunity to reduce the need for such expensive and long-term studies.

6.3 Toxicokinetics and quantitative in vitro – in vivo extrapolation (qIVIVE)

To relate data from non-animal test systems to the human situation, the in vitro concentration levels need to be correlated with the real exposure in vivo. Procedures for such extrapolations (qIVIVE) have been established (Fig. 13). The starting point is a determination of the “real” toxicant concentration that a cell is exposed to. This may be different from the nominal concentration, because of evaporation, metabolism, binding to plastic or uneven distribution in cells. Next, a physiologically-based pharmacokinetic (PBPK) model would be constructed for absorption and distribution in the whole organism, followed by metabolism and excretion. In vitro test systems to predict drug effects of chemicals will be predicted from the early changes they cause in cellular signaling and regulation (Kleensang et al., 2014; Blaauboer et al., 2012). Therefore, signaling pathway identification and analysis is a crucial research necessity in toxicology, and very detailed quantitative information needs to be derived (Fig. 12) to use such data for systems biology modeling (Jennings et al., 2013; Krug et al., 2014; Hartung et al., 2012). Toxicogenomics technologies (Ramirez et al., 2013) are important tools that cover a multitude of cellular events. However, it is important to apply them to the right biological models. For instance, monocultures can hardly model the inflammatory responses frequently seen after long-term exposure to hazardous chemicals.

For repeated dose toxicity, two different approaches are taken in the development of alternative methods: (a) substitution of animals by a battery of relatively complex surrogate models that reflect important features of target tissues and organs. They often use ‘apical’ phenotypic endpoints (e.g., cell death markers) as readouts; (b) an integrated and tiered systems biology approach based on mechanistic endpoints and using the vast knowledge on biological regulation and homeostasis. Pathways-of-toxicity (PoT), emerging from such approaches, will guide hazard evaluation and risk assessment when combined with toxicokinetics modelling (Hartung and McBride, 2011; Boekelheide and Andersen, 2010). The two types of approaches may also be combined.

In summary, repeated dose toxicity will probably be the last method to be replaced, the use of PoT and new culture systems combined with new technologies and sharing of data on pharmaceutical case studies could be the opportunity to reduce the need for such expensive and long-term studies.

6.3 Toxicokinetics and quantitative in vitro – in vivo extrapolation (qIVIVE)

To relate data from non-animal test systems to the human situation, the in vitro concentration levels need to be correlated with the real exposure in vivo. Procedures for such extrapolations (qIVIVE) have been established (Fig. 13). The starting point is a determination of the “real” toxicant concentration that a cell is exposed to. This may be different from the nominal concentration, because of evaporation, metabolism, binding to plastic or uneven distribution in cells. Next, a physiologically-based pharmacokinetic (PBPK) model would be constructed for absorption and distribution in the whole organism, followed by metabolism and excretion. In vitro test systems to predict drug effects of chemicals will be predicted from the early changes they cause in cellular signaling and regulation (Kleensang et al., 2014; Blaauboer et al., 2012). Therefore, signaling pathway identification and analysis is a crucial research necessity in toxicology, and very detailed quantitative information needs to be derived (Fig. 12) to use such data for systems biology modeling (Jennings et al., 2013; Krug et al., 2014; Hartung et al., 2012). Toxicogenomics technologies (Ramirez et al., 2013) are important tools that cover a multitude of cellular events. However, it is important to apply them to the right biological models. For instance, monocultures can hardly model the inflammatory responses frequently seen after long-term exposure to hazardous chemicals.

For repeated dose toxicity, two different approaches are taken in the development of alternative methods: (a) substitution of animals by a battery of relatively complex surrogate models that reflect important features of target tissues and organs. They often use ‘apical’ phenotypic endpoints (e.g., cell death markers) as readouts; (b) an integrated and tiered systems biology approach based on mechanistic endpoints and using the vast knowledge on biological regulation and homeostasis. Pathways-of-toxicity (PoT), emerging from such approaches, will guide hazard evaluation and risk assessment when combined with toxicokinetics modelling (Hartung and McBride, 2011; Boekelheide and Andersen, 2010). The two types of approaches may also be combined.

In summary, repeated dose toxicity will probably be the last method to be replaced, the use of PoT and new culture systems combined with new technologies and sharing of data on pharmaceutical case studies could be the opportunity to reduce the need for such expensive and long-term studies.

6.3 Toxicokinetics and quantitative in vitro – in vivo extrapolation (qIVIVE)

To relate data from non-animal test systems to the human situation, the in vitro concentration levels need to be correlated with the real exposure in vivo. Procedures for such extrapolations (qIVIVE) have been established (Fig. 13). The starting point is a determination of the “real” toxicant concentration that a cell is exposed to. This may be different from the nominal concentration, because of evaporation, metabolism, binding to plastic or uneven distribution in cells. Next, a physiologically-based pharmacokinetic (PBPK) model would be constructed for absorption and distribution in the whole organism, followed by metabolism and excretion. In vitro test systems to predict drug effects of chemicals will be predicted from the early changes they cause in cellular signaling and regulation (Kleensang et al., 2014; Blaauboer et al., 2012). Therefore, signaling pathway identification and analysis is a crucial research necessity in toxicology, and very detailed quantitative information needs to be derived (Fig. 12) to use such data for systems biology modeling (Jennings et al., 2013; Krug et al., 2014; Hartung et al., 2012). Toxicogenomics technologies (Ramirez et al., 2013) are important tools that cover a multitude of cellular events. However, it is important to apply them to the right biological models. For instance, monocultures can hardly model the inflammatory responses frequently seen after long-term exposure to hazardous chemicals.

For repeated dose toxicity, two different approaches are taken in the development of alternative methods: (a) substitution of animals by a battery of relatively complex surrogate models that reflect important features of target tissues and organs. They often use ‘apical’ phenotypic endpoints (e.g., cell death markers) as readouts; (b) an integrated and tiered systems biology approach based on mechanistic endpoints and using the vast knowledge on biological regulation and homeostasis. Pathways-of-toxicity (PoT), emerging from such approaches, will guide hazard evaluation and risk assessment when combined with toxicokinetics modelling (Hartung and McBride, 2011; Boekelheide and Andersen, 2010). The two types of approaches may also be combined.
Inductive toxicity testing has shown high background variability (even among untreated control animals) and is characterized by low species concordance (ToxCast™, for example, showed 60% concordance between rat and rabbit studies; and 56% concordance between zebrafish and rat). In some cases, to overcome low sensitivity, studies in a second species may be requested by regulators. However, the two-species approach increases both the cost of the studies and the false-positive rates dramatically (Hartung, 2009). For this reason, in 2009, a revision of the REACH legislation reduced the use of a second species.

Further progress in this area would be accelerated by regulatory steps that preclude the use of in vivo data unless they come from a formally validated model and therefore have a known predictivity (Carney et al., 2011). Uses of the zebrafish assay (Selderslaghs et al., 2011; Padilla et al., 2012; Truong et al., 2014), the embryonic stem cell test (eSt) (van Dartel and Piersma, 2011; Seiler and Spielmann, 2011) and further developments on the basis of ReProtect test systems (Piersma, 2010) could immediately fill the gap until assays based on human cells become available. The field is developing very dynamically, and, especially in the area of developmental neurotoxicity, many new test systems are emerging (Zimmer et al., 2014; Smirnova et al., 2014; Leist et al., 2013; Bal-Price et al., 2012). The new assays will need to be assembled into an advanced test battery using concepts of ITS design (Fig. 15).

The ReProTect project assembled 35 European partners from academia, SMEs (Small-Medium Enterprises) and governmental institutes in order to develop in vitro reproductive toxicity approaches (http://www.reprotect.eu/). The scientific problem of identifying non-animal test methods in this field was addressed (Hareng et al., 2005). The project was based on a battery of in vitro methods that covered different steps of the reproductive cycle (Fig. 16). In a so-called “feasibility study” conducted at the end of the project, 10 blinded chemicals were tested by the consortium. Effects on 3 endpoints, namely male fertility, female fertility and embryotoxicity were predicted. The results of the feasibility study demonstrated that the vast majority of the predictions made were correct (Schenk et al., 2010).

### 6.4 Reproductive toxicity

Reproductive toxicology, including developmental toxicology, is a particularly difficult field as far as animal-to-human predictions are concerned (Knudsen et al., 2011; Makris et al., 2011). Reproductive toxicity aims to assess possible hazard to the reproductive cycle, with a high interest in the early stages of embryonic development (embryotoxicity). Tests like the two-generation study are among the most costly and require up to 3,200 animals per substance (Hartung, 2008; Rovida and Hartung, 2009). This makes it impossible to test the enormous amount of chemicals present in the market, leading to a lack of information on reproduction and development toxicity of tens of thousands chemicals. Moreover, animal-based tests offer little mechanistic insight into a chemical’s toxic mode-of-action (MoA) (Knudsen et al., 2013; Knudsen, 2013). Animal reproductive toxicity testing has shown high background variability (even among untreated control animals) and is characterized by low species concordance (ToxCast™, for example, showed 60% concordance between rat and rabbit studies; and 56% concordance between zebrafish and rat). In some cases, to overcome low sensitivity, studies in a second species may be requested by regulators. However, the two-species approach increases both the cost of the studies and the false-positive rates dramatically (Hartung, 2009). For this reason, in 2009, a revision of the REACH legislation reduced the use of a second species.

Further progress in this area would be accelerated by regulatory steps that preclude the use of _in vivo_ data unless they come from a formally validated model and therefore have a known predictivity (Carney et al., 2011). Uses of the zebrafish assay (Selderslaghs et al., 2011; Padilla et al., 2012; Truong et al., 2014), the embryonic stem cell test (eSt) (van Dartel and Piersma, 2011; Seiler and Spielmann, 2011) and further developments on the basis of ReProtect test systems (Piersma, 2010) could immediately fill the gap until assays based on human cells become available. The field is developing very dynamically, and, especially in the area of developmental neurotoxicity, many new test systems are emerging (Zimmer et al., 2014; Smirnova et al., 2014; Leist et al., 2013; Bal-Price et al., 2012). The new assays will need to be assembled into an advanced test battery using concepts of ITS design (Fig. 15).

The ReProTect project assembled 35 European partners from academia, SMEs (Small-Medium Enterprises) and governmental institutes in order to develop _in vitro_ reproductive toxicity approaches (http://www.reprotect.eu/). The scientific problem of identifying non-animal test methods in this field was addressed (Hareng et al., 2005). The project was based on a battery of _in vitro_ methods that covered different steps of the reproductive cycle (Fig. 16). In a so-called “feasibility study” conducted at the end of the project, 10 blinded chemicals were tested by the consortium. Effects on 3 endpoints, namely male fertility, female fertility and embryotoxicity were predicted. The results of the feasibility study demonstrated that the vast majority of the predictions made were correct (Schenk et al., 2010).

### Roadmap to toxicokinetic predictions

**Development of generic PBPK platforms**
- User friendly, open access
- Database for physiological parameters
- Different exposure (inhalation, dermal, oral)
- Multiple parallel metabolic pathways models

**Data collection to support modeling**
- Metabolite identification
- Protein binding in cell-based assays
- Tissue partitions (some classes of compounds)
- Metabolism rates / hepatic clearance
- Gut absorption / metabolism
- Transporter substrates / renal clearance

### qIVIVE case studies

**In vitro models**
- Concurrent intestinal absorption / metabolism
- Dermal absorption
- Biobarriers
- Hepatocyte clearance
- Pathway / metabolite ID / kinetics (organotypic)
- Renal clearance

### Characterization of free concentration in cell-based assays

---

**Roadmap to reproductive toxicity predictions**

- Generate relevant biological material (e.g., fetal tissue)
- More physiological culture conditions (e.g., 3D)
- New technologies (e.g., microstructuring, epigenetics)
- Quality assurance
- Evaluation of animal test performance → abolition of bad tests
- Validation of (human) embryonic stem cell tests (“ESTs”)
- Validation of zebrafish egg test (“DART”)
- Extension of ITS approaches “ReProTect II”

---

**Fig. 14: Roadmap to animal-free toxicokinetic predictions**

The experts in the area of toxicokinetics identified research areas requiring further work to obtain human-relevant toxicokinetic data on xenobiotics independent of animal experiments.

Metabolism or certain distribution parameters can provide data for such modeling (Vinci et al., 2012; Gebhardt et al., 2003) but better assays are still required for local specialized metabolism, distribution mediated by transporters, and for excretion processes (e.g., in the kidney). Altogether, this area is far advanced, e.g., for drug development, but its general application for chemicals requires further development (Bessem et al., 2014). Detailed case studies are required to explore the performance of currently available methods (Fig. 14).

---

**Fig. 15: The roadmap for animal-free reproductive toxicity predictions**

In the area of reproductive toxicity the experts suggested, in addition to the points summarized in Figures 4 and 6, to include several specific measures and research lines to be followed.
In the follow-up European project ChemScreen (http://www.chemscreen.eu) 12 chemicals were tested for embryotoxicity in a final performance test. The battery correctly detected 11 out of 12 compounds tested. The consortium concluded that “this study illustrates added value of combining assays that contain complementary biological processes and mechanisms, increasing predictive value of the battery over individual assays” (Piersma et al., 2013).

In silico models (e.g., the US EPA’s Virtual Embryo project) have also shown potential application in the reproductive toxicology field. It is expected that ultimately a computer model that simulates cellular function in the growing embryo can be used to determine the effects of teratogens. Some promising first results come from an in silico modeling platform: A novel multi-cellular agent-based model (ABMs) of vasculogenesis using the CompuCell3D (http://www.compucell3d.org/) modeling environment supplemented with semi-automatic knowledgebase creation has been developed by EPA. Dynamic cell ABMs have been shown to simulate complex developing systems and, consequently, display a potential to simulate adverse effects (Kleinstreuer et al., 2013; Hester et al., 2011; Shirinifard et al., 2013) and aberrant tissue fusion (Ray and Niswander, 2012).

### 6.5 Carcinogenicity

At present, the carcinogenicity hazard of chemicals is determined by a costly and lengthy animal test, the “cancer bioassay”, although its relevance for human health is seriously doubted (Allden et al., 1996; Knight, 2007; Gottmann et al., 2001). Results of more than 3,500 cancer bioassays, which cost about € 800,000 per substance and species, are publically available: 53% of all substances tested were positive, suggesting an enormous false-positive rate, but still some accepted human carcinogens were not positive (Baskettet et al., 2012). The correlation of findings in rats and mice is less than 60%, even less if the site of cancer in the organism is considered. The experts suggested a thorough evaluation of the test, taking into account the principles of evidence-based toxicology (Hoffmann and Hartung, 2006b). This might lead to an abolition of the in vivo assay (Fig. 17). Chemical carcinogenicity may be based either on genotoxic or non-genotoxic (epigenetic) mechanisms (Oliveira et al., 2007). Alternative methods for the determination of genotoxicity have been in use for over 40 years. An ItS has been suggested to combine such available methods (Pfuhler et al., 2010; Aldenberg and Jaworska, 2010). Testing for non-genotoxic carcinogens has proven more difficult, but good results have been obtained recently by various cell transformation assays (Vanparys et al., 2011). A combination of mutagenesis assays, tests for DNA damage, cell transformation assays and targeted tests for frequent epigenetic changes has been suggested to improve the prediction of carcinogenesis. In silico models (e.g., the US EPA’s Virtual Embryo project) have also shown potential application in the reproductive toxicology field. It is expected that ultimately a computer model that simulates cellular function in the growing embryo can be used to determine the effects of teratogens. Some promising first results come from an in silico modeling platform: A novel multi-cellular agent-based model (ABMs) of vasculogenesis using the CompuCell3D modeling environment supplemented with semi-automatic knowledgebase creation has been developed by EPA. Dynamic cell ABMs have been shown to simulate complex developing systems and, consequently, display a potential to simulate adverse effects (Kleinstreuer et al., 2013; Hester et al., 2011; Shirinifard et al., 2013) and aberrant tissue fusion (Ray and Niswander, 2012).

**Roadmap to carcinogenicity predictions**

| Generate appropriate biological material (e.g., human cells) |
| Controlled culture conditions (defined genetic backgrounds) |
| New technologies (operator-independent, automation) |
| Quality assurance |
| Validation of state-of-the-art → abolition of current test |
| Optimize battery for genotoxic carcinogens → new assays of repair, recombination ... |
| ITS including “non-genotoxic” modes of action, epigenetics, PoT basis |
| Further evaluation of "CTAs" (cell transformation assays) |

**Fig. 17: Roadmap for animal-free carcinogenicity predictions**

In the area of carcinogenicity the experts suggested, in addition to the points summarized in Figures 4 and 6, to include several specific measures and research lines to be followed.
mechanisms (e.g., nuclear receptor activation) will most likely form the basis for a future ITS. Most elements are available in some form, but they will require further development and optimization for satisfactory predictivity (Fig. 17).

7 Evaluation of test system performance

Evaluation of test system performance has classically considered three aspects (Fig. 18): (1) the technical reliability of the test; (2) the scientific background, rationale and scope; (3) the correlation of test data with a gold standard (e.g., animal data). The latter point has also been called predictivity (Hartung et al., 2004; Hoffmann and Hartung, 2006a; Moore et al., 2009). The validation procedure has until now followed very strict and rigid rules in the field of chemical testing. This has led to high costs and long delays before new assays were introduced. Moreover, the definition of assay predictivity on the basis of animal data has proven to be problematic because of the shortcomings of the in vivo experiments. Therefore, new validation concepts have to be considered. For instance, high throughput screening assays need to be treated differently from other tests as, e.g., ring trials cannot be performed when certain robotics equipment is available only in one place (Judson et al., 2013). In cases where predictivity cannot be determined from correlation studies, en-

Fig. 18: New validation approach for novel toxicity tests
At present, test validation relies on three pillars: reliability, scientific basis and predictivity. Predictivity has in practical terms been determined by the correlation of in vitro test results with animal data. This approach is not possible for many of the toxicity domains discussed here and many of the assays that are developed are part of a test battery. Future validation must therefore rely on two pillars: Even more focus is required on test quality (reliability). Moreover, the scientific basis of a test needs to be broadened to provide a rationale for the predictive capacity of the test, not based on statistical correlation but based on scientific (mechanistic) explanations.

Fig. 19: Vision for the future of toxicity testing
The current approach is first to test unknown chemicals in animal tests. This limits overall throughput and leads to a high rate of false positives and false negatives. Mixtures are hardly ever tested because of the limitation of resources. High costs in combination with a low predictivity lead to many cases of “no testing”. Mechanistic studies are only carried out in few cases of particular interest to identify the factors causing the toxicity. The new approach, suggested here, is based on 21st century in silico and in vitro methods identifying PoT. This will in most cases lead to an amount of data that is sufficient to decide whether a substance is toxic (positive) or non-toxic (negative) for the intended scenario. Only in few cases, when not enough information can be obtained, will animal tests be performed as an additional source of information. Good information can be provided on all chemicals and, due to the high throughput of the approach, also on mixtures.
tirely different approaches need to be considered. Mechanistic validation, i.e., focus on the scientific background and consistency of an assay (Hartung et al., 2013a; Leist et al., 2012a), is the most promising general option. For specialized assays, e.g., within a test battery or as the basis of a screening approach, the test performance will also need to be judged according to its fit for the specific purpose (Fig. 18). The concept of flexible evaluations of test system performance has been pioneered in the field of drug discovery. Fit-for-purpose evaluation and rating of test predictivity based on the scientific rationale of a test are commonplace in this field, and this experience could help to establish faster and more efficient assay evaluation procedures for chemicals, cosmetics and pesticides. In the discussion on test system predictivity it is important to keep in mind that test reliability will always be a necessary, absolutely mandatory condition, whatever evaluation process is chosen. The progress and acceptance of non-animal based testing will depend on this criterion and its strict implementation in the field (Leist et al., 2010).

8 The future of toxicity testing

The participants of the roadmap consensus symposium envisaged that two key features will distinguish the future from the present toxicity testing (Fig. 19). First, the present animal-based testing (sometimes followed by in vitro tests to supply mechanistic information) will be substituted by ITS using in vitro and in silico approaches (sometimes followed by animal tests, where further data are needed). Second, according to the vision for a new toxicology, data will be generated for every chemical and possibly also for important mixtures. This contrasts with the present situation, in which hardly any data is available on many chemicals (Crofton et al., 2012) and many tests are being waived. Moreover, a large part of the available data lacks mechanistic background and consistency controls. It therefore cannot be used to supply information concerning adverse outcome pathways. Despite the many shortcomings and, in particular, a lack of formal validation, animal data are still being used as gold standard. This contributes to an underestimation of the success of non-animal methods. The expectations regarding scientific validity and predictivity are usually higher for alternative methods than for the respective in vivo models. This is a key issue for the roadmap initiative, as the primary goal is to provide methods that are as good as animal models (not better). This goal might already have been reached in some areas. The future goals would then be to further improve the quality of safety testing beyond that of animal experiments.

As detailed already in the NRC report on the vision for toxicity testing in the 21st century (NRC, 2007; Leist et al., 2008a), the radically new approach to risk assessment has large economic and scientific advantages over the present animal-based system. The roadmap outlined in this overview highlights important steps towards this goal. All experts agreed that considerable work is still needed, but there was also strong consensus that already an impressive advance has been achieved and that the goal is well worth the required efforts.

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. http://dx.doi.org/10.1007/s00204-011-0693-2

Alden, C., Smith, P. F., Piper, C. E. et al. (1996). A critical appraisal of the value of the mouse cancer bioassay in safety assessment. Toxicol Pathol 24, 722-725. http://dx.doi.org/10.1177/01926239602400610

Aldenberg, T. and Jaworska, J. (2010). Multiple Test In Silico Weight-Of-Evidence for Toxicological Endpoints. In M. Cronin and J. Madden (eds.), In Silico Toxicology: Principles and Applications. Cambridge, UK: Royal Society of Chemistry. http://dx.doi.org/10.1039/9781849732093-00558

Ames, B. N. and Gold, L. S. (2000). Paracelsus to parascience: the environmental cancer distraction. Mutat Res 447, 3-13. http://dx.doi.org/10.1016/S0165-1218(00)00194-3

Andersen, M. E., Clewell, H. J., Carmichael, P. L. et al. (2011). Can case study approaches speed implementation of the NRC report: “Toxicity testing in the 21st century: a vision and a strategy?”. ALTEX 28, 175-182. http://dx.doi.org/10.14573/altex.2011.3.175

Andreas, N., Caroline, B., Leslie, F. et al. (2011). The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers in vitro: Results of a ring-study in five laboratories. Toxicol In Vitro 25, 733-744. http://dx.doi.org/10.1016/j.tiv.2010.12.014

Ashikaga, T., Sakaguchi, H., Sono, S. et al. (2010). A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). Altern Lab Anim 38, 275-284.

Bal-Price, A. K., Coecke, S., Costa, L. et al. (2012). Advancing the science of developmental neurotoxicity (DNT): Testing for better safety evaluation. ALTEX 29, 202-215. http://dx.doi.org/10.14573/altex.2012.2.202

Basketter, D. A., Clewell, H., Kimber, I. et al. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing – the 4th report*. ALTEX 29, 3-91. http://dx.doi.org/10.14573/altex.2012.1.003

Bauch, C., Kolle, S. N., Fabian, E. et al. (2011). Intralaboratory validation of four in vitro assays for the prediction of the skin sensitizing potential of chemicals. Toxicol In Vitro 25, 1162-1168. http://dx.doi.org/10.1016/j.tiv.2011.05.030

Bauch, C., Kolle, S. N., Ramirez, T., et al. (2012). Putting the parts together: Combining in vitro methods to test for skin sensitizing. Regul Toxicol Pharmacol 63, 489-504. http://dx.doi.org/10.1016/j.yrtph.2012.05.013 potentials

Bessems, J. G., loizou, G., Krishnan, K. et al. (2014). PBtK modelling platforms and parameter estimation tools to enable animal-free risk assessment: Recommendations from a joint EPAA--EURL ECVAM ADME workshop. Regul Toxicol Pharmacol 68, 119-139. http://dx.doi.org/10.1016/j.yrtph.2013.11.008

Blaauboer, B. J., Boekelheide, K., Clewell, H. J. et al. (2012). The use of biomarkers of toxicity for integrating in vitro hazard estimates into risk assessment for humans. ALTEX 29, 411-425. http://dx.doi.org/10.14573/altex.2012.4.411
Boekelheide, K. and Andersen, M. E. (2010). A mechanistic re-definition of adverse effects – a key step in the toxicity testing paradigm shift. ALTEX 27, 243-52. http://www.altex.ch/All-issues/Issue_50.html?id=121&aid=2

Bouhifd, M., Hogberg, H. T., Kleensang, A. et al. (2014). Mapping the human toxome by systems toxicology. Basic Clin Pharmacol Toxicol 115, 24-31. http://dx.doi.org/10.1111/bcpt.12198

Carney, E. W., Ellis, A. L., Tyl, R. W. et al. (2011). Critical evaluation of current developmental toxicity testing strategies: A case of babies and their bathwater. Birth Defects Res B Dev Reprod Toxicol 92, 395-403. http://dx.doi.org/10.1002/bdrb.20318

Chen, M., Bisgin, H., Tong, L., et al. (2014). Toward predictive models for drug-induced liver injury in humans: are we there yet? Biomark Med 8, 201-213. http://dx.doi.org/10.2217/bmm.13.146

Crofton, K. M., Mundy, W. R. and Shafer, T. J. (2012). Developmental neurotoxicity testing: a path forward. Congenit Anom (Kyoto) 52, 140-146. http://dx.doi.org/10.1111/j.1741-4520.2012.00377.x

Dix, D. J., Houck, K. A., Martin, M. T. et al. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95, 5-12. http://dx.doi.org/10.1093/toxsci/kfl103

Gebhardt, R., Hengstler, J. G., Muller, D. et al. (2003). New hepatocyte in vitro systems for drug metabolism: Metabolic capacity and recommendations for application in basic research and drug development, standard operation procedures. Drug Metab Rev 35, 145-213. http://dx.doi.org/10.1081/DMR-120023684

Gottmann, E., Kramer, S., Pfahringer, B. et al. (2001). Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. Environ Health Perspect 109, 509-514. http://dx.doi.org/10.1289/ehp.10110509

Haren, L., Pellizzer, C., Bremer, S. et al. (2005). The integrated project ReProTox: A novel approach in reproductive toxicity hazard assessment. Reprod Toxicol 20, 441-452. http://dx.doi.org/10.1016/j.reprotox.2005.04.003

Hartung, T., Bremer, S., Casati, S. et al. (2004). A modular approach to the ECVAM principles on test validity. Altern Lab Anim 32, 467-472.

Hartung, T. (2008). Food for thought ... on alternative methods for cosmetics safety testing. ALTEX 25, 147-162. http://www.altex.ch/All-issues/Issue_50.html?id=101&aid=1

Hartung, T. and Leist, M. (2008). Food for thought ... on the evolution of toxicology and the phasing out of animal testing. ALTEX 25, 91-102. http://www.altex.ch/All-issues/Issue_50.html?id=99&aid=1

Hartung, T. (2009). Toxicology for the twenty-first century. Nature 460, 208-212. http://dx.doi.org/10.1038/460208a

Hartung, T. and McBride, M. (2011). Food for thought ... on mapping the human toxome. ALTEX 28, 83-93. http://dx.doi.org/10.14573/altex.2011.2.083

Hartung, T., Blaauboer, B. J., Bosgra, S. et al. (2011). An expert consortium review of the EC-commissioned report “Alternative (non-animal) methods for cosmetics testing: current status and future prospects – 2010”. ALTEX 28, 183-209. http://dx.doi.org/10.14573/altex.2011.3.183

Hartung, T., van Vliet, E., Jaworska, J. et al. (2012). Food for thought ... systems toxicology. ALTEX 29, 119-128. http://dx.doi.org/10.14573/altex.2012.2.119

Hartung, T., Hoffmann, S. and Stephens, M. (2013a). Mechanistic validation. ALTEX 30, 119-130. http://dx.doi.org/10.14573/altex.2013.2.119

Hartung, T., Luechtfeld, T., Maertens, A. et al. (2013b). Integrated testing strategies for safety assessments. ALTEX 30, 3-18. http://dx.doi.org/10.14573/altex.2013.1.003

Hengstler, J. G., van der Burg, B., Steinberg, P., et al. (1999). Interspecies differences in cancer susceptibility and toxicity. Drug Metab Rev 31, 917-970. http://dx.doi.org/10.1081/DMR-100101946

Hengstler, J. G., Marchan, R., and Leist, M. (2012). Highlight report: towards the replacement of in vivo repeated dose systemic toxicity testing. Arch Toxicol 86, 13-15. http://dx.doi.org/10.1007/s00204-011-0798-7

Hester, S. D., Belmonte, J. M., Gens, J. S. et al. (2011). A multicell, multi-scale model of vertebrate segmentation and somite formation. PLoS Comput Biol 7, e1002155. http://dx.doi.org/10.1371/journal.pcbi.1002155

Hoffmann, S. and Hartung, T. (2006a). Toward an evidence-based toxicology. Hum Exp Toxicol 25, 497-513. http://dx.doi.org/10.1191/0960327106hett648oa

Hoffmann, S. and Hartung, T. (2006b). Designing validation studies more efficiently according to the modular approach: retrospective analysis of the EPISKIN test for skin corrosion. Altern Lab Anim 34, 177-191.

Jaworska, J. and Hoffmann, S. (2010). Integrated Testing Strategy (ITS) – Opportunities to better use existing data and guide future testing in toxicology. ALTEX 27, 231-242. http://www.altex.ch/All-issues/Issue_50.html?id=121&aid=1

Jaworska, J., Gabbert, S. and Aldenberg, T. (2010). Towards optimization of chemical testing under REACH: A Bayesian network approach to Integrated Testing Strategies. Regul Toxicol Pharmacol 57, 157-167. http://dx.doi.org/10.1016/j.yrtph.2010.02.003

Jaworska, J., Harol, A., Kern, P. S. et al. (2011). Integrating non-animal test information into an adaptive testing strategy – skin sensitization proof of concept case. Arch Toxicol 85, 119-128. http://dx.doi.org/10.1007/s00204-010-0919-y

Judson, R. S., Houck, K. A., Kavlock, R. J. et al. (2010). In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118, 485-492. http://dx.doi.org/10.1289/ehp.0901392

Judson, R. S., Mortensen, H. M., Shah, I. et al. (2012). Using pathway modules as targets for assay development in xenobiotic screening. Mol Biosyst 8, 531-542. http://dx.doi.org/10.1039/c1mb05303e
Judson, R., Kavlock, R., Martin, M. et al. (2013). Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX 30*, 51-56. http://dx.doi.org/10.14573/alteX.2013.1.051

Judson, R., Houck, K., Martin, M. et al. (2014). In vitro and modelling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast Programme. *Basic Clin Pharmacol Toxicol* 115, 69-76. http://dx.doi.org/10.1111/bcpt.12239

Kavlock, R., Chandler, K., Houck, K. et al. (2012). Update on EPA’s ToxCast program: providing high throughput decision support tools for chemical risk management. *Chem Res Toxicol* 25, 1287-1302. http://dx.doi.org/10.1021/tr3009394

Kleinstreuer, N. C., Yang, J., Berg, E. L., et al. (2014). Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat Biotechnol* 32, 583-591. http://dx.doi.org/10.1038/nbt.2914

Knudsen, T. B., Kavlock, R. J., Daston, G. P. et al. (2011). Developmental toxicity testing for safety assessment: new approaches and technologies. *Birth Defects Res B Dev Reprod Toxicol* 92, 413-420. http://dx.doi.org/10.1002/bdrb.20315

Knudsen, T., Martin, M., Chandler, K. et al. (2013). Predictive models and computational toxicology. *Methods Mol Biol* 947, 343-374. http://dx.doi.org/10.1007/978-1-62703-131-8_26

Knudsen, L. E. (2013). Animal-free toxicology: The use of human tissue to replace the use of animals – examples from human biomonitoring and human placental transport studies. *Altern Lab Anim* 41, 443-447.

Knudsen, T. B., Kavlock, R. J., Daston, G. P. et al. (2011). Developmental toxicity testing for safety assessment: new approaches and technologies. *Birth Defects Res B Dev Reprod Toxicol* 92, 413-420. http://dx.doi.org/10.1002/bdrb.20315

Leist, M. and Hartung, T. (2013). Inflammatory findings on species extrapolations: humans are definitely no 70-kg mice. *ALTEX 30*, 227-230. http://dx.doi.org/10.14573/alteX.2013.2.227

Leist, M., Ringwald, A., Kolde, R. et al. (2013). Test systems of developmental toxicity: state-of-the art and future perspectives. *Arch Toxicol* 87, 2037-2042. http://dx.doi.org/10.1007/s00204-013-1154-x

Li, Y., Pan, D., Liu, J. et al. (2007). Categorical QSAR Models for skin sensitization based upon local lymph node assay classification measures part 2: 4D-fingerprint three-state and two-state logistic regression models. *Toxicol Sci* 99, 532-544. http://dx.doi.org/10.1093/toxsci/kfm185

Makris, S. L., Kim, J. H., Ellis, A. et al. (2011). Current and future needs for developmental toxicity testing. *Birth Defects Res B Dev Reprod Toxicol* 92, 384-394. http://dx.doi.org/10.1002/bdbr.20335

McDowell, R. M. and Jaworska, J. S. (2002). Bayesian analysis and inference from QSAR predictive model results. *SAR QSAR Environ Res* 13, 111-125. http://dx.doi.org/10.1080/10629360290002280

Moore, N., Bremer, S., Carmichael, N. et al. (2009). A modular approach to the extended one-generation reproduction toxicity study. The outcome of an ECETOC Task Force and International ECETOC/ECVAM Workshop. *Altern Lab Anim* 37, 219-225.

Natsch, A. (2010). The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers--functional relevance and a hypothesis on innate reactions to skin sensitizers. *Toxicol Sci* 113, 284-292. http://dx.doi.org/10.1093/toxsci/kfp228

NRC (2007). *Toxicity Testing in the Twenty-first Century: A Vision and a Strategy*. N. A. Press.

Oliveira, P. A., Colaco, A., Chavez, R. et al. (2007). Chemical carcinogenesis. *An Acad Bras Cien* 79, 593-616. http://dx.doi.org/10.1590/S0001-37652007000400004

Olson, H., Betton, G., Robinson, D., et al. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32, 56-67. http://dx.doi.org/10.1006/rthrop.2000.1399

Padilla, S., Corum, D., Padnos, B. et al. (2012). Zebrafish developmental screening of the ToxCast Phase I chemical library. *Reprod Toxicol* 33, 174-187. http://dx.doi.org/10.1016/j.reprotox.2011.08.018

Paparella, M., Daneshian, M., Hornek-Gausterer, R. et al. (2013). Uncertainty of testing methods – what do we (want to) know? *ALTEX 30*, 131-144. http://dx.doi.org/10.14573/alteX.2013.2.131

Peiser, M., Tralau, T., Heidler, J. et al. (2012). Allergic contact dermatitis: epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. Current knowledge assembled at an international workshop at BfR, Germany. *Cell Mol
Pfuhler, S., Kirst, A., Aardema, M. et al. (2010). A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: genotoxicity. A COLIPA analysis. Regul Toxicol Pharmacol 57, 315-324. http://dx.doi.org/10.1016/j.yrtph.2010.03.012

Pfuhler, S., Kirst, A., Aardema, M. et al. (2010). A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: genotoxicity. A COLIPA analysis. Regul Toxicol Pharmacol 57, 315-324. http://dx.doi.org/10.1016/j.yrtph.2010.03.012

Piersma, A. H. (2010). Reproductive toxicology. editorial. Reprod Toxicol 30, 1. http://dx.doi.org/10.1016/j.reprotox.2010.05.008

Piersma, A. H., Bosgra, S., van Duursen, M. B. et al. (2013). Evaluation of an alternative in vitro test battery for detecting reproductive toxicants. Reprod Toxicol 38, 53-64. http://dx.doi.org/10.1016/j.reprotox.2013.03.002

Ramirez, T., Daneshian, M., Kamp, H. et al. (2013). Metabolomics in toxicology and preclinical research. ALTEX 30, 209-225. http://dx.doi.org/10.14573/altex.2013.2.209

Ray, H. J. and Niswander, L. (2012). Mechanisms of tissue fusion during development. Development 139, 1701-1711. http://dx.doi.org/10.1242/dev.068338

Rossini, G. P. and Hartung, T. (2012). Towards tailored assays for cell-based approaches to toxicity testing. ALTEX 29, 359-372. http://dx.doi.org/10.14573/altex.2012.4.359

Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals – a report by the transatlantic think tank for toxicology (t³). ALTEX 26, 187-208. http://www.altex.ch/All-issues/Issue.50.html?id=107&aid=4

Sakaguchi, H., Ashikaga, T., Miyazawa, M. et al. (2006). Development of an in vitro skin sensitization test using human cell lines; human Cell Line Activation Test (h-ClAT). II. An inter-laboratory study of the h-CLAT. Toxicol In Vitro 20, 774-784. http://dx.doi.org/10.1016/j.tiv.2005.10.014

Schenk, B., Weimer, M., Bremer, S. et al. (2010). The ReProTect Feasibility Study, a novel comprehensive in vitro approach to detect reproductive toxicants. Reprod Toxicol 30, 200-218. http://dx.doi.org/10.1016/j.reprotox.2010.05.012

Seiler, A. e. and Spielmann, H. (2011). The validated embryonic stem cell test to predict embryotoxicity in vitro. Nat Protoc 6, 961-978. http://dx.doi.org/10.1038/nprot.2011.348

Sellderslags, I. W., Blust, R. and Witters, H. E. (2011). Feasibility study of the zebrafish assay as an alternative method to screen for developmental toxicity and embryotoxicity using a training set of 27 compounds. Reprod Toxicol 33, 142-54. http://dx.doi.org/10.1016/j.reprotox.2011.08.003

Shirinifard, A., McCollum, C. W., Bolin, M. B. et al. (2013). 3D quantitative analyses of angiogenic sprout growth dynamics. Dev Dyn 242, 518-526. http://dx.doi.org/10.1002/dvdy.23946

Smirnova, L., Hogberg, H. T., Leist, M. et al. (2014). Developmental neurotoxicity – challenges in the 21st century and in vitro opportunities. ALTEX 31, 129-156. http://dx.doi.org/10.14573/altex.1403271

Stephens, M. L., Andersen, M., Becker, R. A. et al. (2013). Evidence-based toxicology for the 21st century: Opportunities and challenges. ALTEX 30, 74-103. http://dx.doi.org/10.14573/altex.2013.1.074

Sturla, S. J., Boobis, A. R., FitzGerald, R. E. et al. (2014). Systems toxicology: from basic research to risk assessment. Chem Res Toxicol 27, 314-329. http://dx.doi.org/10.1021/tr400410s

Taylor, K., Casalegno, C. and Stengel, W. (2011). A critique of the EC’s expert (draft) reports on the status of alternatives for cosmetics testing to meet the 2013 deadline. ALTEX 28, 131-148. http://dx.doi.org/10.14573/altex.2011.2.131

Truong, L., Reif, D. M., St Mary, L. et al. (2014). Multidimensional in vivo hazard assessment using zebrafish. Toxicol Sci 137, 212-233. http://dx.doi.org/10.1093/toxsci/kft235

van Dartel, D. A. and Piersma, A. H. (2011). The embryonic stem cell test combined with toxicogenomics as an alternative testing model for the assessment of developmental toxicity. Reprod Toxicol 32, 235-244. http://dx.doi.org/10.1016/j.reprotox.2011.04.008

van Thriel, C., Westerink, R. H., Beste, C. et al. (2012). Translating neurobehavioural endpoints of developmental neurotoxicity tests into in vitro assays and readouts. Neurotoxicology 33, 911-924. http://dx.doi.org/10.1016/j.neuro.2011.10.002

Vanparys, P., Corvi, R., Aardema, M. et al. (2011). ECM validation of three cell transformation assays. ALTEX 28, 56-59. http://dx.doi.org/10.14573/altex.2011.1.056

Vinci, B., Murphy, E., Iori, E. et al. (2012). An in vitro model of glucose and lipid metabolism in a multicompartmental bioreactor. Biotechnol J 7, 117-126. http://dx.doi.org/10.1002/biot.201100177

Zimmer, B., Pallocca, G., Dreser, N. et al. (2014). Profiling of drugs and environmental chemicals for functional impairment of neural crest migration in a novel stem cell-based test battery. Arch Toxicol 88, 1109-1126. http://dx.doi.org/10.1007/s00204-014-1231-9

Acknowledgements

This work was supported by the Doerenkamp-Zbinden Foundation, the State of Baden-Württemberg, Germany, the German BMBF and the Centers for Alternatives to Animal Testing in the US and Europe.

Correspondence to
Marcel Leist, PhD
University of Konstanz
Box M657
78457 Konstanz
Germany
email: marcel.leist@uni.konstanz.de