Toxicology Research

Cite this: *Toxicol. Res.*, 2012, **1**, 8

www.rsc.org/tx

REVIEW

Validation and quality control of replacement alternatives – current status and future challenges

Marcel Leist, Nina Hasiwa, Mardas Daneshian and Thomas Hartung

Received 5th February 2012, Accepted 9th May 2012
DOI: 10.1039/c2tx20011b

Alternatives to animal testing have been developed mainly in the fields of toxicology and vaccine testing. Typical examples are the evaluation of phototoxicity, eye irritation or skin corrosion/irritation of cosmetics and industrial chemicals. However, examples can also be found in other biomedical areas, such the control of the quality of drug preparations for pyrogens or for the control of the production process of biologics, such as botulinum neurotoxin. For regulatory purposes, the quality, transferability and predictivity of an alternative method needs to be evaluated. This procedure is called the "validation process" of a new method. It follows defined rules, and several governmental institutions have been established to perform, supervise or advise on this process. As this often results in a delay of method implementation, different alternatives for the evaluation of a method's suitability and quality are under discussion. We describe here the principles of model development and quality control. We also give an overview on methods that have undergone validation. Strengths and shortcomings of traditional approaches are discussed, and new developments and challenges are outlined.

1. Introduction

Validation is (or should be) a normal procedure in all fields of science, once a test is developed.[1] A test should be distinguished from a model as it includes a way to derive the test result (also known as data analysis procedure). The validation process is intended to provide confidence into the results, to define where the test may or may not be applied and to give account of test characteristics such as precision, limit of detection, accuracy, specificity, sensitivity, robustness and transferability (Fig. 1).

The establishment, validation and documentation of test methods in different areas of science have been extensively covered in the specialized literature and recently also in teaching programs.[2] This includes specific recommendations published by

Doerenkamp-Zbinden Chair for In-vitro *Toxicology and Biomedicine and Center for Alternatives to Animal Testing in Europe (CAAT-Europe), University of Konstanz, D-78467 Konstanz, Germany.*
E-mail: marcel.leist@uni-konstanz.de

Marcel Leist

Marcel Leist, PhD, is Director of the Center for Alternatives to Animal Testing in Europe (CAAT-Europe) and Professor at the University of Konstanz (D). There, he holds the endowed Doerenkamp-Zbinden chair for in vitro *toxicology and biomedicine. The research concentrates on* in vitro *models of developmental neurotoxicity, and of neurodegenerative diseases, based on the differentiation of stem cells. He holds an MSc in Toxicology from the University of Surrey (UK) and a PhD in Biochemical Pharmacology. Between 2000–2006 he worked in the pharmaceutical company H. Lundbeck A/S (DK). Before that, he was associate professor of Toxicology in Konstanz.*

Nina Hasiwa

Nina (Marina) Hasiwa, PhD, is Scientific Officer at the Center for Alternatives to Animal Testing (CAAT-Europe) and Managing Director and CEO of AtaX-Advice, 'Scientific Service Provider' around Alternative Methods. From 2003–2009 she was a Scientific Officer in the European Center for Validation of Alternatives Methods (ECVAM), where she also obtained her PhD in cooperation with the Chair of Biochemical Pharmacology at the University of Konstanz.

regulatory bodies. For instance, OECD GD 34 gives guidance on *Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment*. While the predictivity and scientific relevance are often difficult to quantify, the quality of an assay system may be assessed by strictly quantitative methods (Fig. 2).

2. Theoretical consideration on the setup of methods

What is essential for the setup of a method before validation? The setup and later validation of a toxicological test system

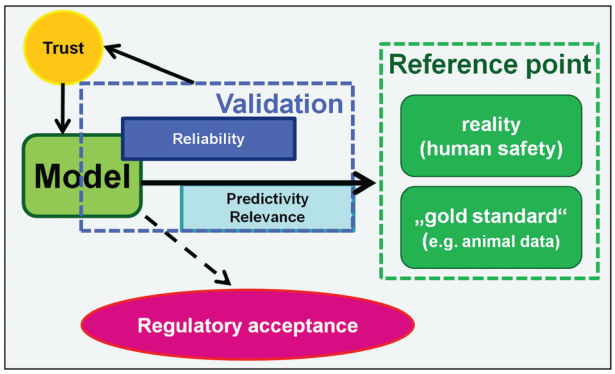

Fig. 1 The role of validation in the evaluation of a test model. A good model (*e.g. in vitro* test method) is expected to provide information on reality (*i.e.* domains outside the model). For instance, it may be constructed to predict human safety. Validation is a procedure that should provide objective information on the performance of the model. Successful validation increases the trust in the usefulness of the model, and therefore is considered a pre-requisite for regulatory acceptance. The validation process involves conceptually two different examinations. The first refers only to the test, and it evaluates its technical reliability (reproducibility, independence of place, time and operators, and so on). The second part of validation examines the relationship between model and reality, to assess its predictivity and scientific relevance. Most evaluations of predictivity examine the correlation of the model with a relevant reference point. This will ideally be human data. As these are often not available, some other gold standard needs to be chosen. In many cases in the past, this has been data from animal tests (*i.e.* from another model). Scientific relevance is defined as the agreement of the test principle with current scientific understanding.

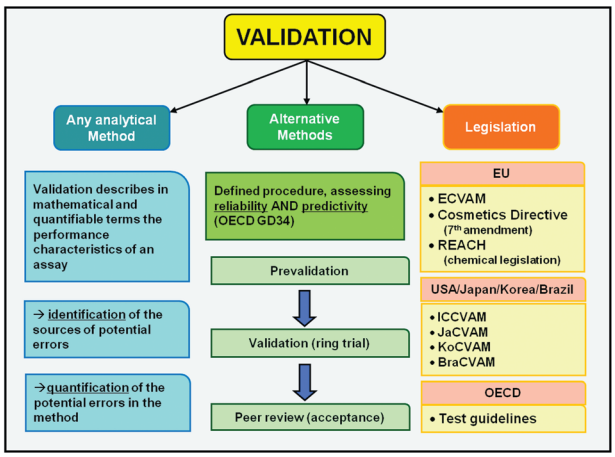

Fig. 2 Different aspects of the term "validation". The term validation is used differently in different contexts. Left: methods are technically validated in all fields of science and engineering. Here the process refers to the internal performance parameters of a method. Middle: in the area of 'Alternative Methods' predictivity also is validated, in addition to the internal performance characteristics (reliability). The OECD offers guidance on the process in its GD34 document. As this process is very resource-consuming, it is common to take a step-wise approach. Prevalidation examines *e.g.* reliability only. A validation performed as a ring trial would examine additional reliability parameters (*e.g.* inter-laboratory variability) and predictivity. An independent review and judgement of the data would then be a third step, before a method can be accepted for regulatory purposes. Right: EU legislation is using the terms "valid method" or "validated method" as legal terms, referring to specified validation procedures. Institutes specialising in the validation of alternative methods in toxicology have been established in many countries. In the field of chemical safety, the OECD guidelines give some good examples for validated and regulatory accepted alternative methods.

Mardas Daneshian

Mardas Daneshian, PhD, is the CEO of the Center for Alternatives to Animal Testing (CAAT-Europe) at the University of Konstanz. After his Postdoc in the 'International Research Training Group' (IRTG) 1331, he held a research position in the project 'Immunolab: European Aeronautic Defence and Space Company' supported by the 'Deutsches Zentrum für Luft- und Raumfahrt' (DLR) and European Space Agency (ESA). Furthermore he is a member of the 'European Society of Alternatives to Animal Testing' (EUSAAT) and a member of the ALTEX board.

Thomas Hartung

Thomas Hartung, MD PhD, is professor of toxicology, pharmacology, molecular microbiology and immunology at Johns Hopkins Bloomberg School of Public Health, Baltimore, USA, and University of Konstanz, Germany. He (co-) directs their Centers for Alternatives to Animal Testing (CAAT). From 2002–2008 he was head of the European Center for the Validation of Alternative Methods (ECVAM), European Commission. He has authored more than 350 publications.

requires some initial thoughts on model setup. The general rules for good scientific experiments may give some initial guidance.[3] The prerequisite of all validation efforts is standardization and documentation of the test. This includes also the application of quality assurance measures such as Good Cell Culture Practice (GCCP)[4,5] or Good Laboratory Practice (GLP).[6] Then, there are three basic requirements to be fulfilled:

(a) Reproducibility: The experiment needs to be independent of the observer/experimenter, the place and the time when it is performed. That means that it should be repeatable by anyone (skilled in the art) and anywhere. The data should be quantifiable. Otherwise it would be hard to establish reproducibility and comparability of the data.

(b) Scientific relevance: The reason and rationale for the experiment should be clear, and, most importantly, it should be embedded in a plausible biological context. This means in a wider sense, that it is hypothesis-driven.

(c) Hypothesis-generating: The results of the experiment must point beyond the experiment itself and make predictions for other conditions. This is for alternative methods formalized as a Prediction Model, *i.e.* an algorithm how to translate the test result into a prediction of the point of reference. The predictions made need to be testable and disprovable.

These can be transferred to the requirements for model development, where three major criteria need to be fulfilled:

(aa) Reliability: This refers to the robustness and reproducibility of the model. Validation of this aspect should be mandatory for each model used, independent of the legal context or other implications. It is an evaluation of the technical quality of a model.

(bb) Scientific relevance of the model: Often, judgement of this aspect will require time and experience to be gained from the use of the model. Also, deep knowledge of biological processes involved in the model and in reality is required.

(cc) Predictivity: This aspect deals with the capacity of the model to yield results that correlate well with reality. Specificity and sensitivity are amongst the parameters that describe this aspect. Notably, sensitivity and specificity are not only technical reliability parameters, and they change over time and with experience gained. Any given number is only valid in relation to the "gold standard" or the "reality" used as surrogate for reality. This is often neglected or not recognized.

The model (*i.e.* the toxicological test used *e.g.* as *in vitro* replacement method, or as animal model) itself is built from four elements. Each of them can be validated and adapted individually.

(aaa) Biological system: This may be for example a dendritic cell or a guinea pig, or a differentiating stem cell.

(bbb) Exposure scheme: The guinea pig may be dosed orally or dermally, once or repeatedly, with a certain vehicle, for a certain time. The stem cell may be exposed to a chemical with or without medium change, during a certain time window and in a specified solvent.

(ccc) Assay endpoint: Death of a cell or of the guinea pig, measured by a specified viability assay, or using a specified humane endpoint; or skin reddening or altered differentiation, determined by PCR or immunocytochemistry. The type of endpoint chosen can completely change the outcome of an assay. In this context it is of utmost importance to distinguish endpoints that describe the biological system from endpoints that describe the behaviour of the test in the presence of chemicals. These separate issues require independent optimisation and characterisation. For instance, a person's body weight can be measured well on scales (to give a good readout on general growth characteristics of a person = biological system), but this endpoint will hardly respond to acute poisoning of the person. Instead, blood pressure or vomiting activity may be good measures of human poisoning (toxicological test), but they in turn give little information on the growth activity over time.

(ddd) Data analysis procedure/Prediction Model: Translation of an endpoint outcome into toxicological information. For instance: is reversible light skin reddening interpreted as sign of sensitization? Or, is a change of gene expression of marker x interpreted as toxicological change? Is there a binary outcome (toxic—non toxic) or are there more than 2 classes (mild, moderate, severe irritants, and how are the boundaries defined); if there are two or more assay endpoints, how are they combined to a final toxicity statement? During validation, the prediction model also needs scrutiny and the questions asked are as follows: Is there a threshold (different from the statistical threshold) for when an effect can be considered biologically relevant? How is the outcome interpreted when more than one endpoint is measured (*e.g.* general cytotoxicity and functional impairment or effects on two different cell types)? Is an increase compared to normal good, when a decrease is bad? How should data be interpreted when a compound alters the baseline values for the endpoint (*e.g.* coloured compound in spectrophotometric assays, reducing agents in tetrazolium reduction assays)? The Prediction Model then translates the test result into a prediction of the point of reference, *e.g.* translating a level of cytotoxicity to a prediction whether the animal would have died.

Before validation of a method can be initiated, all this needs to be clearly defined. Recently, the validation process has been criticized for its slow progress and potentially faulty outcomes, and alternatives are being sought. The evidence based toxicology initiative has attempted to suggest alternative validation approaches.[7] However, it needs to be noted that these require even more stringent definitions of the above criteria and of assay quality. Technical assay quality is an indispensable step before any further validation steps addressing scientific relevance and predictivity.

3. Quality aspects of test systems

The description of a test system for regulatory purposes requires a standard operation procedure.[8] First, this has to define the purpose of the test and to the extent possible its applicability domain. This would also provide *e.g.* information on the source and characterisation of cells, a sufficient description of culture conditions for maintenance and experiment, and information on which parameters are critical and what affects them.[5] It also includes measurement methods, essential instrumentation, important manipulation steps, details on the determination of endpoints, positive and negative controls, assay acceptance criteria and a description of the data processing.

Validation of model relevance needs to answer for instance the following questions: What human problem is modelled? What

biological effect is it designed to measure? Which effects is the test designed to predict? Can it detect deviations from normal to both sides, or does the test work only for one side?

Important assay performance validation questions are: Does a compound that should change the endpoint do this – and by how much does it do this (effect strength = dynamics of the response, maximum possible deviation of endpoint); does a compound that is not expected to change the endpoint behave neutrally? It is frequently neglected although scientifically important that besides negative (NC) and positive (PC) controls (as above), many systems also require unspecific controls (UC) to ensure the quality of the test system. The response dynamics of a PC, and thus the performance of the test method, cannot be qualified without assessing the response to UC. It is important to re-challenge the test method with a new set of PC and NC (learning set, training set of chemicals) to assess its performance with respect to unknown compounds.

Desirably, test methods should assess specific adverse effects (SAE),[9] independent of general cytotoxicity.[10] For instance, inhibition of neurite outgrowth can only be measured meaningfully in a concentration range that does not kill the cells.[11,12] Many endpoints are available to specifically assess cell killing, also in complex settings.[13–15] The toxicity range of test compounds may be determined as follows: a general cytotoxicity/viability test is run over a wide range of concentrations, initially with 10-fold dilutions. After identification of the relevant range, re-testing is performed in a more narrow range (1.5-fold dilutions) to identify the highest non-cytotoxic concentration (HNCC) within the conditions of the assay (*e.g.* a given time frame). For most practical purposes this may be done by using the mathematically-defined IC_{10} value of the cytotoxicity concentration response curve, and moving to the left by a certain factor (*e.g.* HNCC = $EC_{10} \times 0.2$). Ideally, general cytotoxicity (GC) should be determined in parallel/simultaneously with specific adverse effects (SAE). Inability to detect GC does not mean that it does not occur (and could be detected by measuring more sensitive parameters, *e.g.* protein synthesis instead of cell disruption). This applies in particular to short term assays (few hours), as most GC endpoints require several hours to become manifest.

Each experimental setup requires controls of whether the experimental system reacts correctly, *i.e.* in the right direction, or in the right range. They give us an acceptance criterion for believing the other data obtained from unknown samples by the test method. The concept of acceptance criteria is highly important in all quantitative experimental sciences. Especially in *in vitro* toxicology, test systems are usually so complex that they require that known positive and negative controls are measured along with the unknown samples.[16] Only if these controls fulfil the acceptance criteria, the other experimental data can be taken into consideration. Data from an experiment that did not fulfil the acceptance criteria cannot be used.

4. Controls and considerations required for the validation of assay predictivity

The predictive capacity is usually validated by examination of the correlation of assay results with a gold standard (Fig. 3).

However, correlation does not mean causality, even if the correlation is very good.[17] On the one hand, the correlation may be

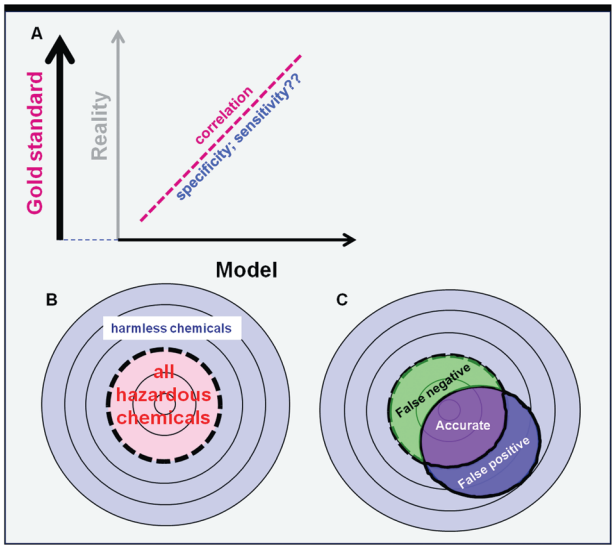

Fig. 3 The interrelation of test model and reality. The most difficult part of the validation of a toxicological test model is the assessment of its correlation with reality (human safety/hazard). (A) As "reality" is often unknown, the data are instead correlated with a "gold standard" assumed to reflect reality closely. Data on specificity, accuracy and sensitivity are obtained from a correlation matrix. (B/C) The meaning of these terms is illustrated graphically. (B) The outer circle comprises all tested chemicals. The outer grey rings (*e.g.* cosmetics, pesticides, drugs) contain all innocuous compounds. The inner (light red, delimited with bold dashed line) rings contain all hazardous compounds. (C) A typical real-life scenario of test performance is shown: the blue area indicates all compounds identified as hazardous by a given test. Some of the really hazardous compounds are correctly classified (= accurate). Some of the really hazardous compounds are not identified as hazardous. These are the false negatives of the test and their number defines test sensitivity. The test wrongly identifies several innocuous compounds as hazardous. These are the false positives and their number correlates with the test specificity. Toxicological tests need to be optimized, in order to reduce the number of false negatives. Usually this increase in sensitivity (*e.g.* expansion of the radius of the blue circle, by testing at very high doses) results in a decrease of specificity.

real but only exist within a small range or under specific conditions or for a limited class of compounds. On the other hand, the correlation may not really exist, but be suggested by the choice of compounds along the continuum of effects. This argument has an important practical implication for test compound selection. For instance, if the question is whether a simple, 24 h fibroblast cytotoxicity assay correlates with a complex endpoint, such as chronic toxicity or carcinogenicity, it can be possible to find a good correlation if the 20 test compounds comprise 10 compounds of very low cytotoxicity, and 10 compounds of high cytotoxicity. Some assays tend to agree when extremes are used, but the resulting (mathematically) good correlations may not hold true for test compounds in the intermediate range. Are such cases relevant and common? Yes, they are, in particular in studies using multiple endpoints. When dozens or hundreds of endpoints are used, such artificial correlations are likely to appear for at least some of them. Typical examples are -omics studies suggesting a correlation between some gene transcripts, metabolites or protein modifications with toxicity.[18] For such

studies, appropriate statistics use measures to counteract the effect of multiple testing on apparent significance of effects (false discovery rate corrections – FDR).[19]

The minimum information required on the response dynamics is the linear and dynamic range of the endpoint, and the detection limit. Moreover, information should be provided on how stable (robust) a readout is. For instance, when neurite growth is measured, data are required on the length under optimal conditions (S), and on the variation of length under these conditions (V); in addition, the minimum length (*N.B.* this is not necessarily zero. It may for example be 50% of the maximum length measured in the presence of the strongest known growth inhibitor) that can be observed under the given assay conditions needs to be determined (B). Also, its variation (N) is an essential piece of information. From these data, the signal-noise ratio (S/N-ratio or $(S - B)/N$)) can be calculated. These data can also be used to define the limit of detection (*e.g.*: $B + (5 \times N)$). Another quality parameter of the test system (independent of any test compound) is the z' factor, which should ideally be >0.5 and indicates the detection power of the system ($z' = 1 - (3 \times (V + N)/(S - B))$). The procedures used to determine z' or S/N ratio are also well suited to detect systematic errors in the assay setup.

Toxicity curves do not necessarily follow a simple mathematical model, and they do not need to reach zero (viability) within the tested range of concentrations. For instance, only a subpopulation of cells may be affected. This means that EC_{50} values cannot be extrapolated. A meaningful EC_{50} requires that real data points (ideally ≥2) exist on both sides of the EC_{50}. Alternatively, the onset of toxicity may be defined by a benchmark dose (BMD).

5 Validation criteria and the validation process

The validation process itself has evolved over time to allow higher throughput, flexibility and efficiency. For this, it is important to recall the main elements of an alternative method. Evidently, a test system is involved. This needs to be coupled with analysis endpoints and a data analysis procedure. Sometimes the third component is neglected: the prediction model relating the results of the method to predictions for human safety. A modular approach[8] has been useful to accelerate the validation procedure. First, the reliability of the test system needs to be validated. This includes testing of the descriptive assay parameters (accuracy, precision, detection limit, linear range, robustness, specificity, sensitivity, response dynamics), at increasing levels of complexity, *i.e.* within a laboratory (different operators) and between different laboratories (transferability). In parallel, the mechanistic and scientific relevance can be evaluated. In a third line of validation, the predictive capacity is evaluated. Up to now, this has been typically done by correlation of the test results with the results of animal experiments. This process may yield information on applicability domains (*e.g.* only certain types of chemicals, but not others).

6. Validation by comparison with animal data

In the field of alternative methods, there has been a lot of focus on one particular aspect of validation: the comparison to animal data. In this sense, validation and the wording "valid methods" and "validated methods" have been used in legal texts, such as the European chemical regulation REACH, the seventh amendment of the European cosmetics directive, and the new directive on the use and protection of experimental animals (2010/63EU).[20,21] Already in 1991 a 'European validation agency', the European Centre for Validation of Alternative Methods (ECVAM) in Ispra (Italy) was created to actively support the development, validation and acceptance of methods which could reduce, refine or replace the use of animal experiments. Comprehensive validation is a prerequisite for the adoption of a new method into a legal framework, such as the OECD test guidelines, or the European pharmacopoeia.

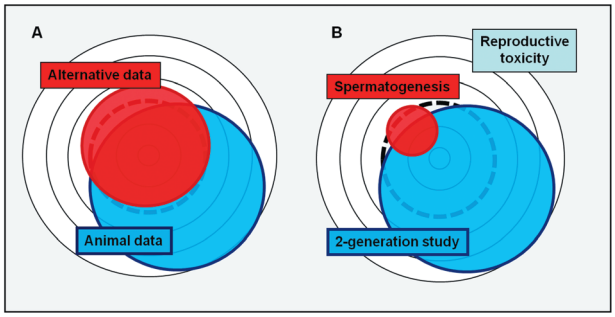

Fig. 4 Typical problems encountered at validation of alternative methods against animal data. Hazardous and innocuous compounds are displayed in circles as in Fig. 3. The dashed bold ring encloses all hazardous compounds. (A) The reference animal test is displayed by the blue area. An alternative test to be validated is represented by the red area. Both tests have good sensitivity, but they differ in the compounds they recognize as positive. The alternative test has a higher specificity than the animal test and should be considered valid, if not superior to the reference test. However, the validation of correlation is based on the ratio of the purple intersection area and the blue area. According to this procedure, the alternative assay would only have a predictivity/correlation of about 50%, and it would therefore fail. Thus, even very good alternative tests fail validation if a bad animal test is used as a gold standard. (B) The situation can be even more complex when an alternative test covers only one biological domain of a respective animal test. This is demonstrated for reproductive toxicity. The 2-generation study in rats covers a wide range of chemical effects. Effects on spermatogenesis cover only a small subgroup of chemicals. Even a perfect test for spermatogenesis would only detect a small subset of compounds detected in a 2-generation reproduction study. This test cannot be reasonably verified/validated against the respective animal experiment. It either requires detailed sub-analysis of the animal data (not always feasible), or the validation of a whole test battery of alternative tests against the animal experiment (with large technical and mathematical problems).

Validation comparing to animal data has been criticised a lot. One of the arguments is that animal experiments may not be suitable as a gold standard, as they do not correlate well enough with human data.[22–26] Another argument is that such a correlative process is not possible, when test batteries are used, that do not model a defined animal experiment[27,28] (Fig. 4).

Therefore, new ideas have been voiced to overcome this problem.[29] The most extreme approaches suggest neglecting the correlation aspects initially, and focusing instead much more on the first two domains of validation: high quality of the test system and high scientific relevance may provide by themselves a good predictivity for human safety. Such concepts are at present being tested and further developed with high speed.

The field of cosmetics is a good example for progress in establishment and validation of alternative methods: replacement methods for some toxicological domains have been validated. These include phototoxicity, skin corrosion, skin irritation, eye corrosion, and to some extent eye irritation. Refinement/reduction methods are also available for acute oral toxicity (tiered testing strategies for the LD_{50} test) and skin sensitization (local lymph node assay). Many of these tests have been accepted by the OECD, and some have substituted the corresponding animal experiments to a large extent.

According to current legislation animal testing for cosmetics was abandoned in 2009 and the marketing of cosmetic ingredients tested on animals for more complex endpoints is foreseen to be banned in 2013. These more complex domains include toxicokinetics, skin sensitization, repeated dose toxicity, carcinogenicity and reproductive toxicity. A recent report published by the European Commission stated that sufficiently validated methods are not available in these domains yet. This opinion was confirmed by a large expert panel assembled by the Centre for Alternatives to Animal Testing in Europe – CAAT-Europe.[30] Thus, test development and validation is on-going with high pressure in these domains.[31]

7. Toxicological and other methods that have been validated

There are more than 80 methods which have been validated or are in some more or less advanced state of validation; about 40 have received validity statements. These include more than 50 *in vitro* tests, 10 using isolated organs, several refined *in vivo* tests, and testing strategies, which combine *in vitro* and *in vivo* approaches. *In vitro* is defined as: 'no animals are involved', except as donors of cells or organs, and the test should be based on cell systems or isolated organs. Also in this area, replacing measures are under investigation.[32–35] Refined *in vivo* methods often involve the use of anaesthetics and analgesics and humane endpoints are applied.[25,36] Furthermore, the development of tiered testing or testing strategies reduces the number of animals involved.

Many alternative methods are anchored in OECD (Organization for Economic cooperation and development) Guidelines. Especially the guidelines for the testing of chemicals, as stated on their website:[37] 'are a collection of about 100 of the most relevant internationally agreed testing methods used by government, industry and independent laboratories to identify and characterize potential hazards of new and existing chemical substances, chemical preparations and chemical mixtures. They are a basic set of tools used primarily in regulatory safety testing and subsequent chemical and chemical product notification and chemical registration. In addition, they can also be used for the selection and ranking of candidate chemicals during the development of new chemicals and products and in toxicology research.'[38]

Another important source is the European Pharmacopeia and their mission is stated on their website:[39] 'The texts of the European Pharmacopoeia (Ph. Eur.) concern the qualitative and quantitative composition of medicines, the tests to be carried out on medicines, on the raw materials used in the production of medicines and on the intermediates of synthesis. It contains texts covering substances, excipients and preparations for pharmaceutical use of chemical, animal, human or herbal origin, homoeopathic preparations and homoeopathic stocks, antibiotics, as well as dosage forms and containers. The texts also cover biologicals, blood and plasma derivatives, vaccines and radiopharmaceutical preparations. They are legally binding'.

In the US the Office of Chemical Safety and Pollution Prevention (OCSPP)[40] under the umbrella of the Environmental Protection Agency[41] is taking care of the harmonization of test methods for chemicals and pesticides.

Due to the fact, that the area of test methods is under permanent development, it is rather challenging to keep track of the current situation. There are several sources available, which try to document the status of 3R methods, but none can claim to be complete.

Information databases, which may be consulted include:

AltTox.org: (update 27th September, 2011, used for this survey)

http://alttox.org/ttrc/validation-ra/validated-ra-methods.html

The Canadian Council of Animal Care in Science (CCAC/CCPA) (most information of April 2009, used for this survey)

http://threers.ccac.ca/en/alternatives/ATM-table-MRE/intro.html

The European Commission through the responsible Institute for Health and Consumer Protection (IHCP) on the website of the European Centre for the Validation of Alternative Methods (ECVAM) (last update 30.06.2011, used for this survey)

http://tsar.jrc.ec.europa.eu/documents/TSAR_public_ongoing_validation_studies_2011-06-30.pdf

An excellent compilation can also be found in form of a journal article.[42] Below a summary of the most prominent and widely accepted methods is provided. The implementation of such assays in regular testing differs largely between countries, institutions and exact data requirements.

Acute aquatic toxicity

One validated test is anchored in the OECD TG 203,[43] using an upper threshold concentration (UTC) step-down approach, which reduces the number of fish used by 65%.[44–46] Another test is under validation by ECVAM for an OECD Project to assess the transferability and reliability of the zebrafish embryo toxicity test for prediction of acute toxicity to fish. It is expected to be ready for implementation in 2012.[47,48]

Acute mammalian toxicity

Acute mammalian toxicity is divided into three subareas by their route of application.[49] For the oral route three tests have been validated, which were implemented in the OECD TG 420,[50] 423,[51] 425.[43] All three methods reduce the animals used from 25 to 5–9.[52–55] Two *in vitro* tests are recommended by the 'Interagency Coordinating Committee on the Validation of Alternative Methods' (ICCVAM) to be implemented into a tiered testing strategy, to reduce the number of animals further.[56] Another test is under validation by ECVAM and is considered to be a follow up validation study on the predictive capacity of 3T3/Neutral Red Uptake cytotoxicity test to identify non-toxic substances for

acute oral toxicity and its potential inclusion into an *in vitro* testing strategy for acute oral toxicity, which is expected to be finalized in 2012.[57] Results obtained in an industrial environment suggest that the 3T3NRU test alone has a limited reduction/refinement potential.[58]

With respect to inhalation exposure the original OECD TG 403[59] is under revision to implement two validated tests, which are suggesting humane endpoints, therefore they are considered to be refinement methods.

For acute dermal toxicity one test is available, also applying humane endpoints. This may lead to a new OECD document (Draft TG 434).

Non-vaccine biologics

A prominent example in the area of non-vaccine biologics is the Mouse LD_{50} Assay for Botox Potency Testing.[60,61] Eight alternative assays are available in different stages of regulatory acceptance. Some tests may have a large economic impact, as they are proprietary and implementation in a guideline would enforce their use by potential competitors. The Snap-25 test is listed as a method for replacement in the European Pharmacopeia for final batch testing,[62–64] while two other assays are recognized by ICCVAM, but further development is recommended. There are three non-lethal mouse models; two listed in the European Pharmacopeia and one is accepted only for BoNT type A. Furthermore, there are two organ models; one is listed in the European Pharmacopeia.[61] Another example for an accepted alternative in the field of biologics is the test for calcitonin bioactivity developed by Novartis.[65]

Vaccines

The testing of vaccines depends on their intended use, human or veterinary, and addresses their potency or safety separately.

For vaccine potency in veterinary use, the lethal challenge test was replaced by an enzyme-linked immuno-sorbent assay (ELISA), a biochemical analytical approach. It is implemented in the European Pharmacopeia, *e.g.*, for swine erysipelas vaccine.[66–68]

For testing the vaccine potency for human use, seven tests are implemented in the European Pharmacopeia. The lethal paralytic challenge test for batch potency of tetanus toxoid vaccines may be replaced by an ELISA[69] measurement and a toxin binding inhibition method,[70] the diphtheria vaccine may be tested *via* a cell-based assay and an ELISA. Hepatitis B and poliomyelitis vaccine are tested *via* serological antigen quantification and rabies potency testing is done by using one dilution only and humane endpoints are applied.[71]

The formerly used target animal vaccine safety test for veterinary use could be dropped because of a retrospective study conducted by ECVAM.

In the area of vaccine safety for human use the 4 following tests are available: (1) the abnormal toxicity test can be deleted from the testing scheme, when batch consistency can be demonstrated,[72] (2) the oral polio neurovirulence test conducted in monkeys may be replaced by an *in vitro* test called 'MAPREC', but only for type 3 oral polio virus vaccines,[73] (3) the use of transgenic mice instead of monkeys (TgPVR21) was validated by the World Health Organization (WHO) for type 1, 2, 3 oral polio vaccines[74] and (4) the residual toxicity of diphtheria may be replaced by the Vero Cell Test.[71]

Chronic toxicity

In the area of chronic toxicity for pesticides, the 1-year dog study was found to be unnecessary by a statement of the ECVAM Scientific Advisory Committee (ESAC) and the US Environmental Protection Agency.[41] It was found that the 1 year study does not provide more information than the 90 days study, but some countries still require these data.[75]

Eye corrosion and irritation

For eye corrosion and irritation studies ICCVAM will implement in 2012 the routine use of anesthetics, systemic analgesics and humane endpoints. Several validation studies were carried out, leading so far to the adoption of two organotypic assays anchored in OECD TG 437[76] and 438,[77] and two cytotoxicity and cell function-based assays. An ESAC validity statement was granted in 2009 leading to draft test guidelines[78,79] this process was mainly carried out by a retrospective weight-of-evidence validation[80] as a proof-of-principle.

Furthermore, there is one test under validation by ECVAM and the former European Cosmetics Association (COLIPA), now Cosmetics Europe (CosEU) to assess the transferability, reliability and predictive capacity of two *in vitro* test methods, based on reconstructed human tissue models, to be used as stand-alone test methods to identify chemicals not classified as eye irritant ('non-irritant').[81,82]

Food safety

In the area of food safety, two tests have been validated to replace the Mouse Bioassay for shellfish toxins (PSP). One screening method and a high performance liquid chromatography (HPLC) approach were accepted in the EU in 2010.[83–85]

Carcinogenicity

There are two tests under validation for carcinogenicity to assess protocol standardization, transferability and reproducibility (but not performance) of three protocols of cell transformation assays: the Syrian hamster embryo (SHE) pH 6.7, the Syrian hamster embryo (SHE) pH 7.0 and the BALB/c 3T3 assays.[86] Furthermore a validation study is ongoing to verify if the Bhas 42 cells based cell transformation assay might be an equivalent.[87–89] Results are expected in 2012.

Genotoxicity

The area of genotoxicity is covered by eight *in vitro* tests, validated to different extent, which are part of a tiered testing strategy to reduce the number of animals. They are reflected in several well-established OECD documents,[90–96] and the lately adopted OECD TG 487.[97] Furthermore, two *in vitro* comet assays are undergoing validation and several new approaches are under development.[98]

Hematotoxicity

One hematotoxicity test for acute neutropenia (CFU-GM) has been validated by ECVAM. The test can be applied instead of a second animal species. Therefore it is not considered a replacement, though a reduction of animals is achieved.[99]

Phototoxicity

To determine phototoxicity, the European Commission accepted the *in vitro* neutral red uptake[100] phototoxicity test as method B.41 in Annex V of the EU Council Directive 67/548/EEC and in 2004 it was also anchored in an OECD guideline: TG 432.[101] Animal methods to detect phototoxic effects of chemicals are prohibited in all member states.

Pyrogenicity

To replace the rabbit pyrogen test, five *in vitro* tests based on human cell models have been validated by ECVAM.[102] They can be used to detect Gram-negative mediated pyrogenicity. The official European Pharmacopeia listed test, the lymulus amoebocyte lysate assay[103] lacks the capability of detecting Gram-positive stimuli. The cell-based assays may also be useful for Gram-positive mediated pyrogenicity.[104–110] This might lead to a full replacement of the rabbit test in the near future.

Reproductive and developmental toxicity

Due to the complexity of the reproductive cycle and the importance of the developmental process, not many alternatives are available in these areas. Only recently the OECD accepted the extended one-generation study,[111] which replaces the two-generation study.[112] As stated there: 'For reproductive endpoints, it is envisaged that, as a first step and when available, information from repeat-dose studies (including screening reproductive toxicity studies, e.g. TG 422), or short term endocrine disrupter screening assays, (e.g. uterotrophic assay – TG 440; and Hershberger assay – TG 441) is used to detect effects on reproductive organs for males and females. This might include spermatogenesis (testicular histopathology) for males and oestrous cycles, follicle counts/oocyte maturation and ovarian integrity (histopathology) for females. The Extended One-Generation Reproductive Toxicity Study then serves as a test for reproductive endpoints that require the interaction of males with females, females with conceptus, and females with offspring and the F1 generation until after sexual maturity.'

There are also two ECVAM-validated *in vitro* tests using embryos from animals.[113,114] In addition, there is one stem cell-based test (EST) available,[115] which is recommended by ECVAM to be part of a tiered testing strategy, although reality might be different.[116]

Endocrine active substances

There are two OECD accepted methods, anchored in the OECD TG 455[117] and 456;[118] they may be used for screening purposes.

The US EPA accepted a Tier 1 Screening battery including, beside several *in vivo* assays, five *in vitro* tests, which are accepted by the Office of Prevention, Pesticides and Toxic Substances (OPPTS) and laid down as legally binding guidelines for the US as Series 890 OPPTS.[119]

There are two methods under validation by ICCVAM, ECVAM and the Japanese Centre for the Validation of Alternative Methods (JaCVAM) to assess the transferability and reliability of the assays to rank chemicals according to their potency for oestrogen receptor activation or suppression for use as a building block in future testing strategies to detect endocrine active compounds. Evaluation is expected to be finished in 2012.

Skin

There are many tests available for hazard assessment regarding the human skin. They are divided into 4 different areas: absorption, penetration, corrosion, and irritation.

For skin absorption and dermal penetration, a regulatory accepted dermal percutaneous test is available, which may replace the animal test, when human skin is used.[120] More information is available in the according OECD guidance document.[121]

For skin corrosion, three different methods are available, all integrated in OECD guidelines,[122–124] but their use is specified. For example, the test 'Corrositex' can be used to identify acids and bases and substances, which are identified as corrosives, will not proceed further to the animal test. The 'TER' test can distinguish between corrosives and non-corrosives, but non-corrosives will require further confirmation by an animal test. The human skin models (EPISKIN, EpiDerm, SkinEthic) are accepted in the EU as full replacement for corrosivity testing anchored in the Regulation 440/2008/EC. In the US these tests can be used to exclude corrosives, while negative results lead to an animal test.

Skin irritation can be detected *via* the above mentioned human skin models (EPISKIN, EpiDerm, SkinEthic), but using different protocols. Tests were adopted in Commission Regulation (EC) Nr. 761/2009: Method B.46 of the Annex to 440/2008/EC (EU Test Methods Regulation) included in July 2009 and OECD Test Guideline 439[125] published in July 2010.

Dermal sensitization

Sensitization is detected by local lymph node *in vivo* assays (LLNA) and is recommended by ICCVAM to be a stand-alone substitute for the guinea-pig sensitization test.[41,126–129] The reduced rLLNA is able to distinguish between sensitizers and non-sensitizers, and if a chemical is negative in the rLLNA it will not proceed to the full LLNA, which results in fewer animals,[130] which has been included into OECD Test Guideline 429 in July 2010.[131] There are 3 methods under validation by ECVAM, ICCVAM and JaCVAM to assess the assay's transferability and reliability in view of future incorporation into a testing strategy for full replacement of current regulatory animal tests. Cosmetics Europe (CosEU), the EU project Sens-it-iv and many others developed a whole battery of pure *in vitro* assays, which are considered to lead to a full animal replacement within few years.[132,133]

Toxicokinetics and metabolism

The area of toxicokinetics and metabolism is considered to be very complex and difficult to model. Nevertheless, two assays are under validation by ECVAM, ICCVAM and JaCVAM to assess the transferability and reliability of measuring liver enzyme (Cytochrome P450) induction using the human cryoHepaRG® cell line and cryopreserved human hepatocytes to provide a human-metabolically competent model for use in future testing.

Cosmetics

For the field of cosmetics testing, the status of replacement methods for sensitization, carcinogenicity, toxicokinetics, repeat dose toxicity and reproductive toxicity has been reviewed very extensively.[31,134–136]

Nanotoxicology, nanoparticles

In the emerging field of nanotoxicity, there is large potential for the application of alternative methods, but none of them has been validated yet.[137–140]

8. Old *versus* new approaches of validation

Validation approaches are closely linked to the concept initially used for model development. In this context, it is important to recall that there are two fundamentally different ways of constructing test systems, which we call here (a) 'correlative approach', and (b) 're-constructive approach'.

(a) Correlative approach

The correlative approach has been used most frequently for the establishment of alternative methods for animal experimentation, and therefore the whole theoretical concept of 'validation' has been adapted to this approach. In brief, this approach uses the test method as input–output system. Validation is in this case a lot concerned with an evaluation, how well input correlates with output. The model itself is often a kind of black box, with only limited information available on the relevant processes and reactions inside. This has the large advantage that reasonable correlations can be obtained without a need of knowledge on mechanisms of toxicity, or of regulatory mechanisms within the model. The disadvantage is obviously that the relevant processes are often not known. Examples illustrate this situation best. The first is the rat cancer bioassay. When it was established, it was a true black box model. The rationale was only that some correlation was expected between compounds known to be carcinogenic in man, and the ones triggering tumorigenesis in mice. Mechanisms of carcinogenesis were largely unknown and did not need to be known for this model. Very powerful carcinogens and clear non-carcinogens correlated nicely between this model and the situation in man. Problems became obvious when a lack of correlation was observed for several classes of compounds. For instance, the so-called peroxisome proliferators triggered hepatocarcinogenesis in the mouse/rat, but not in man. Another example from the same field of toxicology is the Ames bacterial mutagenesis assay that was introduced to detect mutagens, at that time believed to be also carcinogens. Backmutation of errors in the bacterial genes coding for histidine synthesis have obviously no resemblance or scientific relevance with respect to human carcinogenesis, but the model achieved a reasonably good overall correlation, and was therefore widely accepted. It became only later evident that about half of human carcinogens are non-mutagens, and can therefore not be detected in this assay. A correction or adaptation of the assay is not possible, as it does not reflect human biology. It was simply established as a correlative black-box model.

A third example also comes from the field of carcinogenesis. The human cell transformation assay predicts mutagenic and epigenetic carcinogens with an astonishingly high specificity. It is still unclear why the assay works, and what the underlying biological principle is. However, the validation of the assay is far advanced on the basis of correlations with human or animal data.

The strength of this correlative model setup is proven by the assays that have been developed on this basis and have been successfully validated and used. Model development was possible without the requirement of in depth biological knowledge. The weaknesses are demonstrated on the example of the embryonic stem cell test (EST). The EST uses murine embryonic stem cells (EST). They are differentiated with a very rough protocol to mixed cultures containing cardiomyocytes and pacemaker cells, which results in patches of cells that spontaneously start beating. Compounds are being tested for their potential to inhibit the development of these beating cell clumps. In initial validations, the assay was found to predict teratogens with high specificity and sensitivity, and it was recommended by ECVAM and the ECVAM Scientific Advisory Committee (ESAC) for regulatory use. A biological characterization of the assay had never taken place, and the mechanisms and regulations underlying this assay were never characterized. The use of a small number of validation compounds and the absence of biological knowledge harbours some dangers as demonstrated by the history of this assay. In a broader validation with compounds chosen within the context of the ReProTect study, the assay failed.[141]

A priori, it may not seem necessary to understand an assay as long as it delivers good (= predictive and reproducible) results. Toxicological testing has largely adopted this approach, not just *in vivo*, but also *in vitro*. However, there are strong reasons to move ahead to mechanism-based *in vitro* tests to attribute a scientific rationale to the correlations found in new test systems. Paradoxically, especially modern technologies settle for black box approaches and blind correlations. Such approaches bear the risk of measuring trivialities if they are not based on a mechanistic rationale. For example: new metabolomic or transcriptomic fingerprints to predict complex forms of toxicity (*e.g.* developmental toxicity) may indeed only be expensive and sensitive measures of classical cytotoxicity. Results only gain scientific validity when they are controlled by various approaches and when falsification attempts of their predictions have failed. The attempt to identify the underlying pathways of toxicity PoT) aims to give sense to such signatures of toxicity.[142–147]

(b) Re-constructive approach

The second type of modelling was termed above the "re-constructive approach" or mechanistic validation.[1,148] This name was chosen because such models try to reconstruct reality, using biological information and mathematical relationships between

model parameters as building bricks. This approach requires the understanding of the biological process to be modelled, not only in qualitative, but also in quantitative terms. A biological process needs to be dissected into all its components. Each component needs to be understood. Moreover, the relationships between the components need to be understood, and mathematical approaches need to be developed to describe the relationship between all components and parameters. Finally, these elements can be used for "re-constructing" reality as close as possible. An example is physiology-based pharmacokinetic modelling. The corresponding black box model is the injection of compounds into animals and the evaluation of their pharmacokinetic behaviour (time course of plasma concentrations, urinary excretion…) and the correlation of this information with the expected behaviour in man. PBPK modelling would use information on hepatic metabolism, solubility, lipophilicity and renal excretion to model the behaviour of the drug in a human body, using a set of differential equations. The validation of such models would not only refer to the input–output correlation, but also to the construct of the model. This is a difficult task and firm guidelines for this have not yet been established.[149]

It is noteworthy that in reality, the two extremes of black box modelling and pure reconstruction do hardly exist. Often, the approaches are combined to some extent. For instance PBPK models would use information obtained from rodent models. Then information would be used from human and rodent liver metabolism, and this information would then be used to translate rodent information better to human information in an optimised PBPK model using the so-called parallelogram approach. Other examples below illustrate the incorporation of biological and mechanistic information into correlative models. For instance, in the case of skin irritation, originally the damage to skin was measured by classical viability assays. Attempts are on-going to account for inflammatory processes and active reactions of cells in the skin by measurement of chemokine release. Also in the field of sensitization, biological information is incorporated into available models. One approach for instance tested the effect of keratinocyte addition in a dendritic cell activation model, to reflect their normal biological presence and role in co-stimulation and haptenization.

New dimensions of challenges (Fig. 5) are provided by the validation of integrated testing strategies.

This will somehow require the validation of individual components, but also of the relationships established between them and used for the overall modelling. Thus, this form of validation combines issues from the two types of model validation, correlative and re-constructive, discussed above. The challenges of such an approach may be illustrated by the example of dermal sensitization. An integrated test battery may involve a haptenization assay, measuring the covalent binding of the chemical to a peptide. It may also involve some physicochemical characterization to be used to predict skin penetration. A dendritic cell activation assay, in the absence or presence of keratinocytes would be added. Eventually, also T cell stimulation may be probed. The test strategy parts will then have to be linked and weighted. One simple rule may be: if a compound is positive in one of the assays, it is considered a sensitizer. More complex sets of rules would use a hierarchical decision setup. For instance, compounds unlikely to penetrate the skin, or without chemical reactivity may be considered of low hazard, even in presence of positive dendritic cell activation. Such integrated testing strategies (ITS)[133,150] may again be validated by a correlative approach of the overall ITS *vs.* reality.[151] This will most likely be the first and most immediate solution in the near future. An additional approach may be the validation of different new models or ITS (components) against one another in an iterative optimisation process. Pure re-constructive validations of such approaches will require huge amounts of data and experience, but they will become necessary in cases, in which little "reality" information is available, such as the area of developmental neurotoxicity.

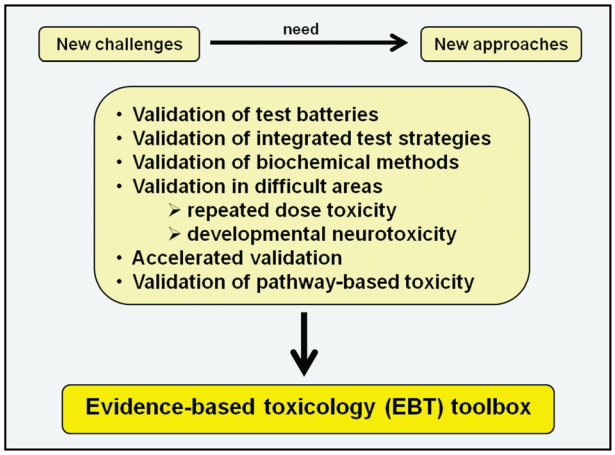

Fig. 5 Future challenges of test validation. More complex test systems (*e.g.* integrated test batteries) or difficult areas (*e.g.* developmental neurotoxicity) are calling for new approaches, which might be developed by the use of an evidence-based toxicology toolbox. Particular challenges come from all approaches that do not substitute animal models on a 1 : 1 basis. For example the approaches of the Tox21c strategy (pathways of toxicity) will require entirely new validation concepts.

9. Outlook on validation in the field of developmental neurotoxicity (DNT)

Developmental neurotoxicity is an area that requires such new validation concepts, as not enough animal data are available.[152] Therefore, this field is chosen as example. Today one out of six children is diagnosed with a developmental disorder[153–155] and in many cases this involves the central nervous system (CNS). Common neurodevelopmental disorders include learning disabilities, neurodevelopmental delays, autism, and attention deficit and hyperactivity disorder (ADHD). Although the assessment and reporting of these disorders have improved over the last few years, scientific evidence suggests that the incidence of such disorders is actually increasing. A major concern is that exposures to drugs and industrial chemicals contribute to this increase.[156–158] A recent review[159] revealed that just over 100 compounds have been tested in studies using the OECD 426 draft guideline on developmental neurotoxicity. Most of these compounds were pesticides (66%) and only 8 industrial chemicals were included. Another review[160] identified about 174 compounds for which neurobehavioral risk assessment had been performed, in many cases also on the offspring of the exposed

animals (F1 generation). Only 1% of these compounds were industrial chemicals. Thus, the available data regarding the developmental neurotoxicity of industrial chemicals is thus rather limited. For some compounds developmental neurotoxicity is the most sensitive of all toxicity endpoints evaluated in a broad safety evaluation battery. Thus, although developmental neurotoxicity appears to be an important domain of safety evaluation, test capacity is limited and test costs are extremely high, amounting to $1.4 million for a single substance according to the DNT guidelines (OECD TG 426 and US EPA 712-C-98-239). But there are also scientific concerns regarding the relevance of these studies for human health effects. The interpretation of the behavioural, morphological and histological data generated *in vivo* can be difficult. This makes it hard to predict human health effects. Consequently, testing according to current guidelines does often not provide rich information for regulatory decision-making, and the test throughput is severely limited. This puts pressure on the development of faster and cheaper *in vitro* systems that can predict developmental neurotoxicity, give information comparable to behavioural readouts, and facilitate screening or at least prioritization of relevant drugs and chemicals for further testing.

In response to this problem, experts in the field, guided by our Centres for Alternatives to Animal Testing (CAAT) at Johns Hopkins University and the University of Konstanz, have organized a series of International workshops ("DNT meetings") to discuss the current status and problems of developmental neurotoxicity assessment, to identify promising alternative approaches and to provide recommendations for the future.[148,152,161] One of the major recommendations after the workshop in 2008 was the development of an *in vitro* testing strategy based on a number of relevant *in vitro* approaches to be used for the initial prioritization of high-risk chemicals. Chemicals identified by such an *in vitro* testing strategy[162] should get the highest priority for regulatory DNT testing and risk assessment to ensure the protection of public health.

We envision that future *in vitro* test systems for developmental neurotoxicity will combine the above validation approaches with exposure information, and we suggest a strategy for test system development and cell-based risk assessment.

This approach is pioneering a different way to alternative test strategies: By expert consensus the relevant modes of action and reference substances are identified. Test models, for instance based on stem cells[163–165] are designed to reflect these modes of action and the known effects of the reference compounds. We have termed this "mechanistic validation"[1,166] to complement reproducibility assessments and substitute for the comparisons with the traditional animal model, for which no suitable data are available[7] It might well be that instead of global validity statements of a method, only local validity can be demonstrated, *i.e.*, that by testing of structurally similar reference substances, which give correct results, the extrapolation to the new substance without reference results is supported "test-across".[1]

We propose that the emerging knowledge from molecular and cellular neuroscience and mechanistic neurotoxicology can be exploited to design *in vitro* tests that read out cellular and molecular endpoints that are predictive of behavioural signs of neurotoxic exposures in humans.[167] A tremendous opportunity lies in a molecular definition of modes of action as pathways of toxicity (PoT)[31,143] as pioneered for DNT in our FDA-sponsored DNTox-21c project: It shall identify critical PoT based on already identified reference chemicals, based on transcriptomics and metabolomics, validate the newly identified PoT using specific pathway inhibitors and siRNA and create PoT-specific cell models based on reporter genes and produce a larger biological data-set with further reference chemicals and additional substances.

For acute or more chronic neurotoxic effects[168] the onset of these effects is temporally associated with the onset of the chemical exposure and usually follows a dose–response relationship. However, the discipline of developmental/neurodevelopmental toxicology faces an additional problem. It is difficult to provide evidence for cause–effect relationships for processes with a long time lag between exposure (*e.g.* gestational) and effect (*e.g.* adult life). Suitable test systems for delayed effects need to be identified. Research in this particular area is also motivated by increasing incidence of neurodevelopmental disorders such as autism, ADHD and schizophrenia and the growing awareness that environmental factors influence susceptibility and/or severity of these diseases.

10. Conclusions

Validation of alternative methods remains work in progress two decades after its formal implementation creating validation bodies like ECVAM. This holds true not only with regard to the pipeline of methods to be validated but also as to the process itself.[169,170] Its value lies in its scientific rigor and in its feasibility with regard to the efforts spent. Pragmatic compromises need to be made as to the latter, now that the proof of principle of *in vitro* replacement has been made. The challenges of new technologies (omics, HTS, high-content methods, ITS, PoT-based approaches) urge to adapt and expand the concepts of validation. Other areas of quality assurance of methods offer role models, *e.g.* the field of chemical analytical methods and evidence-based medicine, recently translated to evidence-based toxicology.[171–173] The concepts of validation, especially with regard to the establishment of scientific relevance of a test model have been pioneered in the field of alternative methods for regulatory purposes,[174,175] but they can serve as a role model for all life sciences, where too often these basic requirements for good science are neglected.

References

1 T. Hartung, Food for thought ... on validation, *ALTEX*, 2007, **24**(2), 67–80.
2 M. Daneshian, *et al.,* A framework program for the teaching of alternative methods (replacement, reduction, refinement) to animal experimentation, *ALTEX*, 2011, **28**(4), 341–352.
3 C. Burgess, *Valid Analytical Methods and Procedures*, Royal Society of Chemistry, 1st edn, 2001.
4 T. Hartung, *et al.,* Good cell culture practice. ECVAM good cell culture practice task force report 1, *Altern. Lab. Anim.*, 2002, **30**(4), 407–414.
5 S. Coecke, *et al.,* Guidance on good cell culture practice. A report of the second ECVAM task force on good cell culture practice, *Altern. Lab. Anim.*, 2005, **33**(3), 261–287.
6 L. A. Knight and I. A. Cree, Quality assurance and good laboratory practice, *Methods Mol. Biol. (Totowa, N. J.)*, 2011, **731**, 115–124.
7 T. Hartung, Evidence-based toxicology – the toolbox of validation for the 21st century?, *ALTEX*, 2010, **27**(4), 253–263.

8 T. Hartung, *et al.*, A modular approach to the ECVAM principles on test validity, *Altern. Lab. Anim.*, 2004, **32**(5), 467–472.
9 S. Creton, *et al.*, Cell transformation assays for prediction of carcinogenic potential: state of the science and future research needs, *Mutagenesis*, 2011.
10 R. L. Kodell, *et al.*, Hierarchical models for probabilistic dose-response assessment, *Regul. Toxicol. Pharmacol.*, 2006, **45**(3), 265–272.
11 P. B. Kuegler, *et al.*, Markers of murine embryonic and neural stem cells, neurons and astrocytes: reference points for developmental neurotoxicity testing, *ALTEX*, 2010, **27**(1), 17–42.
12 N. V. Stiegler, *et al.*, Assessment of chemical-induced impairment of human neurite outgrowth by multiparametric live cell imaging in high-density cultures, *Toxicol. Sci.*, 2011, **121**(1), 73–87.
13 C. Volbracht, M. Leist and P. Nicotera, ATP controls neuronal apoptosis triggered by microtubule breakdown or potassium deprivation, *Mol. Med.*, 1999, **5**(7), 477–489.
14 J. Lotharius, *et al.*, Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed-lineage kinase pathway, *J. Neurosci.*, 2005, **25**(27), 6329–6342.
15 O. Hansson, *et al.*, Overexpression of heat shock protein 70 in R6/2 Huntington's disease mice has only modest effects on disease progression, *Brain Res.*, 2003, **970**(1–2), 47–57.
16 M. Leist, L. Efremova and C. Karreman, Food for thought … considerations and guidelines for basic test method descriptions in toxicology, *ALTEX*, 2010, **27**(4), 309–317.
17 M. Balls, *et al.*, The principles of weight of evidence validation of test methods and testing strategies, The report and recommendations of ECVAM workshop 58, *Altern. Lab. Anim.*, 2006, **34**(6), 603–620.
18 M. Leist, T. Hartung and P. Nicotera, The dawning of a new age of toxicology, *ALTEX*, 2008, **25**(2), 103–114.
19 Y. Benjamini, *et al.*, Controlling the false discovery rate in behavior genetics research, *Behav. Brain Res.*, 2001, **125**(1–2), 279–284.
20 T. Hartung, Comparative analysis of the revised Directive 2010/63/EU for the protection of laboratory animals with its predecessor 86/609/EEC – a t4 report, *ALTEX*, 2010, **27**(4), 285–303.
21 C. Rovida, Food for thought … why no new *in vitro* tests will be done for REACH by registrants, *ALTEX*, 2010, **27**(3), 175–183.
22 M. Bahramsoltani, *et al.*, Quantitation of angiogenesis and antiangiogenesis *in vivo*, *ex vivoin vitro* – an overview, *ALTEX*, 2009, **26**(2), 95–107.
23 T. Lindl and M. Voelkel, No clinical relevance of approved animal experiments after seventeen years, *ALTEX*, 2011, **28**(3), 242–243.
24 M. E. Andersen, Calling on science: making "alternatives" the new gold standard, *ALTEX*, 2010, **27**, 29–37.
25 N. Hasiwa, *et al.*, Critical evaluation of the use of dogs in biomedical research and testing in Europe, *ALTEX*, 2011, **28**(4), 326–340.
26 J. Zurlo, Critical evaluation of the use of dogs in biomedical research and testing, *ALTEX*, 2011, **8**(4), 355–359.
27 T. Hartung, Food for thought … on animal tests, *ALTEX*, 2008, **25**(1), 3–16.
28 C. Westmoreland, *et al.*, Assuring safety without animal testing: Unilever's ongoing research programme to deliver novel ways to assure consumer safety, *ALTEX*, 2010, **27**(3), 61–65.
29 S. Hoffmann, *et al.*, Points of reference in the validation process: the report and recommendations of ECVAM Workshop 66, *Altern. Lab. Anim.*, 2008, **36**(3), 343–352.
30 M. Daneshian, M. Leist and T. Hartung, The center for alternatives to animal testing – Europe (CAAT-EU): a transatlantic bridge for the paradigm shift in toxicology, *ALTEX*, 2010, **27**(1), 63–69.
31 D. A. Basketter, *et al.*, A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing – t4 report*, *ALTEX*, 2012, **29**(1), 3–91.
32 C. Rauch, *et al.*, Alternatives to the use of fetal bovine serum: human platelet lysates as a serum substitute in cell culture media, *ALTEX*, 2011, **28**(4), 305–316.
33 D. Brunner, *et al.*, Serum-free cell culture: the serum-free media interactive online database, *ALTEX*, 2010, **27**(1), 53–62.
34 J. van der Valk, *et al.*, Optimization of chemically defined cell culture media—replacing fetal bovine serum in mammalian in vitro methods, *Toxicol In Vitro*, 2010, **24**(4), 1053–1063.
35 M. Hamann, *et al.*, Non-invasive genotyping of transgenic mice: comparison of different commercial kits and required amounts, *ALTEX*, 2010, **27**(3), 185–190.
36 J. Stracke, *et al.*, Assessment of stress in laboratory beagle dogs constrained by a Pavlov sling, *ALTEX*, 2011, **28**(4), 317–325.
37 OECD, OECD Guidelines for the Testing of Chemicals, Homepage. 2012 [cited 2012; Available from: http://www.oecd.org/document/7/0,3343,en_2649_34377_37051368_1_1_1_1,00.html
38 T. Hartung, Food for thought ... on alternative methods for chemical safety testing, *ALTEX*, 2010, **27**(1), 3–14.
39 EDQM. European Pharmacopeia 7th Edition, Homepage. 2012; Available from: http://www.edqm.eu/en/European-Pharmacopoeia-1401.html
40 EPA. About the Office of Chemical Safety and Pollution Prevention (OCSPP), Homepage. 2012; Available from: http://www.epa.gov/aboutepa/ocspp.html
41 EPA. Health Effects Test Guideline, OPPTS 870.2600. Skin Sensitization EPA 712-C-03-197. Washington, DC: U.S. Environmental Protection Agency, 2003 [cited 2012; Available from: http://iccvam.niehs.nih.gov/SuppDocs/FedDocs/EPA/EPa870r_2600.pdf
42 H. Kandarova and S. Letasiova, Alternative methods in toxicology: prevalidated and validated methods, *Interdiscip. Toxicol.*, 2011, **4**(3), 107–113.
43 OECD, Test No. 203: Fish, Acute Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, 1992, DOI: 10.1787/9789264069961-en.
44 T. H. Hutchinson, *et al.*, A strategy to reduce the numbers of fish used in acute ecotoxicity testing of pharmaceuticals, *Environ. Toxicol. Chem.*, 2003, **22**(12), 3031–3036.
45 S. Jeram, *et al.*, A strategy to reduce the use of fish in acute ecotoxicity testing of new chemical substances notified in the European Union, *Regul. Toxicol. Pharmacol.*, 2005, **42**(2), 218–224.
46 ECVAM, ESAC Statement on the scientific validity of the upper threshold concentration (UTC) step-down approach – A new testing strategy to reduce the use of fish in acute aquatic toxicity testing, 2006.
47 I. W. Selderslaghs, *et al.*, Locomotor activity in zebrafish embryos: a new method to assess developmental neurotoxicity, *Neurotoxicol. Teratol.*, 2010, **32**(4), 460–471.
48 I. W. Selderslaghs, *et al.*, Development of a screening assay to identify teratogenic and embryotoxic chemicals using the zebrafish embryo, *Reprod. Toxicol.*, 2009, **28**(3), 308–320.
49 T. Seidle, P. Prieto and A. Bulgheroni, Examining the regulatory value of multi-route mammalian acute systemic toxicity studies, *ALTEX*, 2011, **28**(2), 95–102.
50 OECD, Test No. 420: Acute Oral Toxicity – Fixed Dose Procedure, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, 2002, DOI: 10.1787/9789264070943-en.
51 OECD, Test No. 423: Acute Oral toxicity – Acute Toxic Class Method, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, 2002, DOI: 10.1787/9789264071001-en.
52 M. J. van den Heuvel, *et al.*, The international validation of a fixed-dose procedure as an alternative to the classical LD_{50} test, *Food Chem. Toxicol.*, 1990, **28**(7), 469–482.
53 E. Schlede, *et al.*, The international validation studyof the acute toxic class method (oral), *Arch. Toxicol.*, 1995, **69**(10), 659–670.
54 W. Diener and E. Schlede, Acute toxic class methods: alternatives to LD/LC_{50} tests, *ALTEX*, 1999, **16**(3), 129–134.
55 E. Schlede, *et al.*, A national validation study of the acute-toxic-class method – an alternative to the LD_{50} test, *Arch. Toxicol.*, 1992, **66**(7), 455–470.
56 M. D. Bethesda, Guidance document on using in vitro data to estimate starting doses for acute toxicity, in *01-4500*, ICCVAM, Editor, NIH, 2001.
57 W. S. Stockes and L. Schechtman, *In vitro* cytotoxicity test methods for estimating starting doses for acute oral systemic toxicity tests, in *ICCVAM Test Method Evaluation REPORT*, ICCVAM, 2006.
58 A. Schrage, *et al.*, Refinement and reduction of acute oral toxicity testing: a critical review of the use of cytotoxicity data, *Altern. Lab. Anim.*, 2011, **39**(3), 273–295.
59 OECD, Test No. 403: Acute Inhalation Toxicity, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, 2009, DOI: 10.1787/9789264070608-en.
60 S. Bitz, The botulinum neurotoxin LD_{50} test – problems and solutions, *ALTEX*, 2010, **27**(2), 114–116.
61 M. D. Bethesda, Report on the ICCVAM-NICEATM/ECVAM scientific workshop on alternative methods to refine, reduce or replace the mouse LD_{50} assay for botulinum toxin testing, in *08-6416*, NIH, Editor 2008, NIH.

62 T. A. Ekong, I. M. Feavers and D. Sesardic, Recombinant SNAP-25 is an effective substrate for Clostridium botulinum type A toxin endopeptidase activity *in vitro*, *Microbiology*, 1997, **143**(Pt 10), 3337–3347.
63 R. E. Gaines Das, *et al.*, Validation of *in vitro* assays for botulinum toxin: a case study, *Dev. Biol. Stand.*, 1999, **101**, 267–276.
64 D. Sesardic, *et al.*, Detection of antibodies against botulinum toxins, *Mov. Disord.*, 2004, **19**(Suppl 8), S85–S91.
65 T. Hartung, Three Rs potential in the development and quality control of pharmaceuticals, *ALTEX*, 2001, **18**(Suppl 1), 3–13.
66 P. P. Pastoret, *et al.*, New vaccine production technologies and their impact on the use of animals, *Dev. Biol. (Basel)*, 2002, **111**, 219–220.
67 U. Rosskopf-Streicher, *et al.*, Quality control of inactivated erysipelas vaccines: results of an international collaborative study to establish a new regulatory test, *Vaccine*, 2001, **19**(11–12), 1477–1483.
68 U. Rosskopf-Streicher, *et al.*, Potency testing of swine erysipelas vaccines by serology results of a pre-validation study, *ALTEX*, 1999, **16**(3), 123–128.
69 M. Balls and E. Hellsten, Statement on the application of the ELISA procedure for batch potency testing of tetanus vaccines for human use, *I.f.H.C.P. Joint Research Centre*, ECVAM Unit, Editor 2000, European Commission, Ispra.
70 M. Balls and E. Hellsten, Statement on the application of the TOBI test for batch potency testing of tetanus vaccines for human use, *I.f.H.C.P. Joint Research Centre*, ECVAM Unit, Editor 2000, European Commission, Ispra.
71 Council_of_Europe, ed. 2.7.6 Assay of diphtheria vaccine (adsorbed), 01/2005:20706. *European Pharmacopoeia*, 6th edn, Council of Europe, Strasbourg, France, 2008, vol. 2.7.6.
72 M. Schwanig, *et al.*, Elimination of abnormal toxicity test for sera and certain vaccines in the European Pharmacopoeia, *Vaccine*, 1997, **15**(10), 1047–1048.
73 WHO, Final Report: IABS Scientific Workshop on Neurovirulence Tests of Live Virus Vaccines, Geneva, Switzerland, 2005, p. 7.
74 E. Dragunsky, *et al.*, Transgenic mice as an alternative to monkeys for neurovirulence testing of live oral poliovirus vaccine: validation by a WHO collaborative study, *Bull World Health Organ.*, 2003, **81**(4), 251–260.
75 OECD, Test No. 409: Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264070721-en, 1998.
76 OECD, Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying Ocular Corrosives and Severe Irritants, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264076303-en, 2009.
77 OECD, Test No. 438: Isolated Chicken Eye Test Method for Identifying Ocular Corrosives and Severe Irritants, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264076310-en, 2009.
78 OECD. Draft OECD Guideline for the Testing of Chemicals: The Cytosensor Microphysiometer Test Method: An *In Vitro* Method for Identifying Chemicals Not Classified as Irritant, as well as Ocular Corrosive and Severe Irritant Chemicals, 2010, Available from: http://www.oecd.org/dataoecd/17/43/46784020.pdf
79 OECD. Draft OECD Guideline for the Testing of Chemicals: Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants, 2010, Available from: http://www.oecd.org/dataoecd/17/44/46784112.pdf
80 T. Hartung, *et al.*, First alternative method validated by a retrospective weight-of-evidence approach to replace the Draize eye test for the identification of non-irritant substances for a defined applicability domain, *ALTEX*, 2010, **27**(1), 43–51.
81 P. McNamee, *et al.*, Update on the Colipa research programme for development of in vitro alternative methods for eye irritation, in *6th World Congress on Alternatives & Animal Use in the Life Sciences*, Tokyo, Japan, 2007.
82 L. Scott, *et al.*, A proposed eye irritation testing strategy to reduce and replace *in vivo* studies using bottom-up and top-down approaches, *Toxicol. In Vitro*, 2010, **24**(1), 1–9.
83 J. F. Jellett, *et al.*, Detection of paralytic shellfish poisoning (PSP) toxins in shellfish tissue using MIST Alert, a new rapid test, in parallel with the regulatory AOAC mouse bioassay, *Toxicon.*, 2002, **40**(10), 1407–1425.
84 F. H. Mackintosh, *et al.*, Assessment of MIST Alert, a commercial qualitative assay for detection of paralytic shellfish poisoning toxins in bivalve molluscs, *J. AOAC Int.*, 2002, **85**(3), 632–641.
85 FAO/IOC/WHO, Report of the joint FAO/IOC/WHO ad hoc expert consultation on biotoxins in bivalve molluscs, Oslo, Norway, 2004, pp. 26–30.
86 P. Vanparys, *et al.*, ECVAM prevalidation of three cell transformation assays, *ALTEX*, 2011, **28**(1), 56–9.
87 R. Combes, *et al.*, Cell transformation assays as predictors of human carcinogenicity – the report of ECVAM workshop 39, *ATLA*, 1999, **27**, 745–767.
88 D. Maurici, *et al.*, Carcinogenicity, *Altern. Lab. Anim.*, 2005, **33**(Suppl 1), 177–182.
89 OECD, Detailed review paper on cell transformation assays for detection of chemical carcinogens, OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 31, ENV/JM/MONO (2007)18, 2007, 170.
90 OECD, Test No. 471: Bacterial Reverse Mutation Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071247-en, 1997.
91 OECD, Test No. 480: Genetic Toxicology: *Saccharomyces cerevisiae*, Gene Mutation Assay, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071407-en, 1986.
92 OECD, Test No. 481: Genetic Toxicology: *Saccharomyces cerevisiae*, Miotic Recombination Assay, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071421-en, 1986.
93 OECD, Test No. 473: *In vitro* Mammalian Chromosome Aberration Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071261-en, 1997.
94 OECD, Test No. 476: *In vitro* Mammalian Cell Gene Mutation Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071322-en, 1997.
95 OECD, Test No. 479: Genetic Toxicology: *In vitro* Sister Chromatid Exchange Assay in Mammalian Cells, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071384-en, 1986.
96 OECD, Test No. 482: Genetic Toxicology: DNA Damage and Repair, Unscheduled DNA Synthesis in Mammalian Cells *in vitro*, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071445-en, 1986.
97 OECD, Test No. 487: *In Vitro* Mammalian Cell Micronucleus Test; DOI: 10.1787/9789264091016-en, 2010.
98 M. Moreno-Villanueva, *et al.*, The automated FADU-assay, a potential high-throughput *in vitro* method for early screening of DNA breakage, *ALTEX*, 2011, **28**(4), 295–303.
99 A. Pessina, *et al.*, Application of the CFU-GM assay to predict acute drug-induced neutropenia: an international blind trial to validate a prediction model for the maximum tolerated dose (MTD) of myelosuppressive xenobiotics, *Toxicol. Sci.*, 2003, **75**(2), 355–367.
100 M. Schulz, *et al.*, Toxicity assessment of waste water samples with fish cell lines, *ALTEX*, 1995, **12**(4), 188–195.
101 OECD, Test No. 432: *In Vitro* 3T3 NRU Phototoxicity Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071162-en, 2004.
102 ECVAM. ESAC Statement on the validity of in-vitro pyrogen tests, 2006 [cited 2012; Available from: http://ecvam.jrc.it/ft_doc/ESAC statement pyrogenicity 20060516.pdf].
103 J. A. Troutman, *et al.*, The incorporation of lysine into the peroxidase peptide reactivity assay for skin sensitization assessments, *Toxicol. Sci.*, 2011, **122**(2), 422–436.
104 NICEATM-ICCVAM, Independent Peer Review Panel Report: Five *In Vitro* Test Methods Proposed for Assessing Potential Pyrogenicity of Pharmaceuticals and Other Products. http://iccvam.niehs.nih.gov/docs/pyrogen/PrRevPanFinRpt.pdf, 2007.
105 NICEATM-ICCVAM, ICCVAM Background Review Document: Validation Status of Five *In Vitro* Test Methods Proposed for Assessing Potential Pyrogenicity of Pharmaceuticals and Other Products, 71003, in Federal Register, 2008.
106 NICEATM-ICCVAM. *In Vitro* Pyrogen Test Methods: Overview and Recommendations. 2009; Available from: http://iccvam.niehs.nih.gov/methods/pyrogen/pyrogen.htm
107 S. Poole, *et al.*, A rapid 'one-plate' *in vitro* test for pyrogens, *J. Immunol. Methods*, 2003, **274**(1–2), 209–220.
108 S. Hoffmann, *et al.*, International validation of novel pyrogen tests based on human monocytoid cells, *J. Immunol. Methods*, 2005, **298**(1–2), 161–173.

109 S. Hoffmann, *et al.*, Optimisation of pyrogen testing in parenterals according to different pharmacopoeias by probabilistic modelling, *J. Endotoxin. Res.*, 2005, **11**(1), 25–31.
110 R. Perdomo-Morales, *et al.*, Monocyte activation test (MAT) reliably detects pyrogens in parenteral formulations of human serum albumin, *ALTEX*, 2011, **28**(3), 227–235.
111 OECD, Test No. 415: One-Generation Reproduction Toxicity Study, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264070844-en, 1983.
112 OECD, Test No. 443: Extended One-Generation Reproductive Toxicity Study, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264122550-en, 2008.
113 ECVAM. The Micromass Test – Method of Brown: INVITTOX no. 122. Embryotoxicity/Teratogenicity, 2006.
114 H. Spielmann, *et al.*, The practical application of three validated *in vitro* embryotoxicity tests. The report and recommendations of an ECVAM/ZEBET workshop (ECVAM workshop 57), *Altern. Lab. Anim.*, 2006, **34**(5), 527–538.
115 ECVAM, Embryonic Stem Cell Test (EST): INVITTOX no. 113. Embryotoxicity/Teratogenicity, 2002.
116 C. Rovida, F. Longo and R. R. Rabbit, How are reproductive toxicity and developmental toxicity addressed in REACH dossiers?, *ALTEX*, 2011, **28**(4), 273–294.
117 OECD, Test No. 455: The Stably Transfected Human Estrogen Receptor-alpha Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264076372-en, 2009.
118 OECD, Test No. 456: H295R Steroidogenesis Assay, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264122642-en, 2011.
119 EPA, OCSPP Harmonized Test Guidelines, Series 890 – Endocrine Disruptor Screening Program Test Guidelines, 2012.
120 OECD, Test No. 428: Skin Absorption: *In Vitro* Method, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071087-en, 2004.
121 OECD. OECD Guidance on dermal absorption. 2010 [cited 2012; Available from: http://www.oecd.org/dataoecd/0/1/46257610.pdf].
122 OECD, Test No. 430: *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance Test (TER), OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071124-en, 2004.
123 OECD, Test No. 431: *In Vitro* Skin Corrosion: Human Skin Model Test, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071148-en, 2004.
124 OECD, Test No. 435: *In Vitro* Membrane Barrier Test Method for Skin Corrosion, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264067318-en, 2006.
125 OECD, Test No. 439: *In Vitro* Skin Irritation Reconstructed Human Epidermis Test Method, DOI: 10.1787/9789264090958-en, 2010.
126 OECD, Test No. 429: Skin Sensitisation: Local Lymph Node Assay, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, DOI: 10.1787/9789264071100-en, 2002.
127 NICEATM-ICCVAM. The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation. 1999 [cited 2012; Available from: http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf
128 ISO, ISO 10993-10:2010, Biological evaluation of medical devices – Part 10: Tests for irritation and delayed type hypersensitivity, 2002.
129 ECVAM. ESAC Statement on the Validity of the Local Lymph Node Assay for Skin Sensitisation Testing. 2000 [cited 2012; Available from: http://ecvam.jrc.it/page_pdf.cfm?voce=s&idvoce=27].
130 ECVAM. ESAC Statement on the reduced local lymph node assay (rLLNA). 2007; Available from: http://ecvam.jrc.it/page_pdf.cfm?voce=s&idvoce=27
131 C. Rovida, Local lymph node assay: how testing laboratories apply OECD TG 429 for REACH purposes, *ALTEX*, 2011, **28**(2), 117–129.
132 G. Maxwell, *et al.*, Skin sensitisation: the Colipa strategy for developing and evaluating non-animal test methods for risk assessment, *ALTEX*, 2011, **28**(1), 50–55.
133 J. Jaworska, *et al.*, Integrating non-animal test information into an adaptive testing strategy – skin sensitization proof of concept case, *ALTEX*, 2011, **28**(3), 211–225.
134 S. Adler, *et al.*, Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010, *Arch. Toxicol.*, 2011, **85**(5), 367–485.
135 K. Taylor, C. Casalegno and W. Stengel, A critique of the EC's expert (draft) reports on the status of alternatives for cosmetics testing to meet the 2013 deadline, *ALTEX*, 2011, **28**(2), 131–148.
136 T. Hartung, *et al.*, An expert consortium review of the EC-commissioned report "alternative (non-animal) methods for cosmetics testing: current status and future prospects – 2010", *ALTEX*, 2011, **28**(3), 183–209.
137 U. G. Sauer, Animal and non-animal experiments in nanotechnology – the results of a critical literature survey, *ALTEX*, 2009, **26**(2), 109–128.
138 U. G. Sauer, Political incentives towards replacing animal testing in nanotechnology?, *ALTEX*, 2009, **26**(4), 285–294.
139 E. K. Silbergeld, *et al.*, t^4 Workshop Report: Nanotoxicology: "the end of the beginning" – signs on the roadmap to a strategy for assuring the safe application and use of nanomaterials, *ALTEX*, 2011, **28**(3), 236–241.
140 T. Hartung, Food for thought ... on alternative methods for nanoparticle safety testing, *ALTEX*, 2010, **27**(2), 87–95.
141 P. Marx-Stoelting, *et al.*, A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect workshop, *Altern. Lab. Anim.*, 2009, **37**(3), 313–328.
142 E. van Vliet, Current standing and future prospects for the technologies proposed to transform toxicity testing in the 21st century, *ALTEX*, 2011, **28**(1), 17–44.
143 T. Hartung and M. McBride, Food for thought … on mapping the human toxome, *ALTEX*, 2011, **28**(2), 83–93.
144 M. T. Cronin, *et al.*, Formation of mechanistic categories and local models to facilitate the prediction of toxicity, *ALTEX*, 2011, **28**(1), 45–49.
145 K. Boekelheide and M. E. Andersen, A mechanistic redefinition of adverse effects – a key step in the toxicity testing paradigm shift, *ALTEX*, 2010, **27**(4), 243–252.
146 M. E. Andersen, *et al.*, Can case study approaches speed implementation of the NRC report: "toxicity testing in the 21st century: a vision and a strategy?", *ALTEX*, 2011, **28**(3), 175–182.
147 S. Lund, *et al.*, The dynamics of the LPS triggered inflammatory response of murine microglia under different culture and in vivo conditions, *J. Neuroimmunol*, 2006, **180**(1–2), 71–87.
148 S. Coecke, *et al.*, Workgroup report: incorporating *in vitro* alternative methods for developmental neurotoxicity into international hazard and risk assessment strategies, *Environ. Health Perspect.*, 2007, **115**(6), 924–931.
149 D. R. Dietrich, Courage for simplification and imperfection in the 21st century assessment of "Endocrine disruption", *ALTEX*, 2010, **27**(4), 264–278.
150 J. Jaworska and S. Hoffmann, Integrated Testing Strategy (ITS) – Opportunities to better use existing data and guide future testing in toxicology, *ALTEX*, 2010, **27**(4), 231–242.
151 A. Kinsner-Ovaskainen, *et al.*, Overcoming barriers to validation of non-animal partial replacement methods/integrated testing strategies: the report of an EPAA-ECVAM workshop, *Altern. Lab. Anim.*, 2009, **37**(4), 437–444.
152 K. M. Crofton, *et al.*, Developmental neurotoxicity testing: recommendations for developing alternative methods for the screening and prioritization of chemicals, *ALTEX*, 2011, **28**(1), 9–15.
153 C. A. Boyle, P. Decoufle and M. Yeargin-Allsopp, Prevalence and health impact of developmental disabilities in US children, *Pediatrics*, 1994, **93**(3), 399–403.
154 P. Decoufle, *et al.*, Increased risk for developmental disabilities in children who have major birth defects: a population-based study, *Pediatrics*, 2001, **108**(3), 728–734.
155 T. Schettler, Toxic threats to neurologic development of children, *Environ. Health Perspect.*, 2001, **109**(Suppl 6), 813–816.
156 B. M. Kuehn, Increased risk of ADHD associated with early exposure to pesticides, *PCBs, JAMA*, 2010, **304**(1), 27–8.
157 S. K. Sagiv, *et al.*, Prenatal organochlorine exposure and behaviors associated with attention deficit hyperactivity disorder in school-aged children, *Am. J. Epidemiol.*, 2010, **171**(5), 593–601.
158 P. Grandjean and P. J. Landrigan, Developmental neurotoxicity of industrial chemicals, *Lancet*, 2006, **368**(9553), 2167–2178.
159 S. L. Makris, *et al.*, A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426, *Environ. Health Perspect.*, 2009, **117**(1), 17–25.

160 L. D. Middaugh, *et al.*, Neurobehavioral assessment: a survey of use and value in safety assessment studies, *Toxicol. Sci.*, 2003, **76**(2), 250–261.
161 P. Lein, P. Locke and A. Goldberg, Meeting report: alternatives for developmental neurotoxicity testing, *Environ. Health Perspect.*, 2007, **115**(5), 764–768.
162 A. K. Bal-Price, *et al., In vitro* developmental neurotoxicity (DNT) testing: relevant models and endpoints, *Neurotoxicology*, 2010, **31**(5), 545–554.
163 M. Leist, *et al.*, The biological and ethical basis of the use of human embryonic stem cells for *in vitro* test systems or cell therapy, *ALTEX*, 2008, **25**(3), 163–190.
164 B. Zimmer, *et al.*, Sensitivity of dopaminergic neuron differentiation from stem cells to chronic low-dose methylmercury exposure, *Toxicol. Sci.*, 2011, **121**(2), 357–367.
165 B. Zimmer, *et al.*, Coordinated waves of gene expression during neuronal differentiation of embryonic stem cells as basis for novel approaches to developmental neurotoxicity testing, *Cell Death Differ.*, 2011, **18**(3), 383–395.
166 S. Kadereit, Z. B. Zimmer, C. van Thriel, J. G. Hengstler and M. Leist, Compound selection for *in vitro* models of developmental neurotoxicity and prenatal neurotoxicity, *Frontiers in Bioscience*, 2012, **17**, 2442–2460.
167 C. van Thriel, *et al.*, Translating neurobehavioural endpoints of developmental neurotoxicity tests into *in vitro* assays and readouts, *Neurotoxicology*, 2011.
168 J. G. Hengstler, R. Marchan and M. Leist, Highlight report: towards the replacement of in vivo repeated dose systemic toxicity testing, *Arch. Toxicol.*, 2012, **86**(1), 13–5.
169 N. Wilcox and A. Goldberg, Food for thought … on validation. A puzzle or a mystery: an approach founded on new science, *ALTEX*, 2011, **28**(1), 3–8.
170 W. S. Stokes and M. Wind, Validation of innovative technologies and strategies for regulatory safety assessment methods: challenges and opportunities, *ALTEX*, 2010, **27**(3), 87–95.
171 S. Hoffmann and T. Hartung, Toward an evidence-based toxicology, *Hum. Exp. Toxicol.*, 2006, **25**(9), 497–513.
172 T. Hartung, Food for thought … on evidence-based toxicology, *ALTEX*, 2009, **26**(2), 75–82.
173 J. Zurlo, Evidence-based toxicology collaboration, kick-off meeting, *ALTEX*, 2011, **28**, 152.
174 CAAT, 21st Century Validation Strategies for 21st Century Tools. ALTEX 2010; Available from: http://caat.jhsph.edu/programs/workshops/july13validation.html
175 D. Rudacille, Summary of the July 2010 workshop, "21st century validation strategies for 21st century tools", *ALTEX*, 2010, 27.

