Food for Thought ... Considerations and Guidelines for Basic Test Method Descriptions in Toxicology

Marcel Leist, Liudmila Efremova and Christiaan Karreman
Doerenkamp-Zbinden Chair for in vitro Toxicology and Biomedicine, University of Konstanz, Germany

Summary
The development and application of new test systems and test methods is central to the progress of in vitro toxicology. In order to live up to the future challenges, it is important to use the vast knowledge of adjoining fields, such as cell biology or developmental biology, and to attract specialists from such areas to develop new methods. Not all of them bring with them the necessary experience and training required for the development of toxicological test systems. Thus, promising new biological test systems sometimes still require additional considerations to become successful. Sometimes even the referees of scientific journals and their editors seem to lack judgement guidelines for minimum performance standards. Here we provide a list of points to be considered for the establishment of a test method. The chapters range from the explanation of the meaning of positive controls, performance standards or signal-noise ratios to a discussion of statistical considerations, suitable solvents and data display formats. The considerations are simple and expressed in a non-mathematical format, with a strong focus on plausibility and common sense. The major intention of this article is to provide a compilation of important issues requiring consideration. Whether they apply to a specific system and whether action is required must be determined by individual judgement.

Keywords: test system, validation, quality control, data presentation, assay design

1 Introduction

Alternatives to animal tests are often "*in vitro* assays". These often use cell cultures or subcellular fractions as the "test system", and measure multiple biological or chemical "endpoints" (Tab. 1). This vocabulary, and the corresponding scientific approaches, is well-established in the field of *in vitro* toxicology, but less so in other biological areas that contribute importantly to the development of new testing methods. The establishment of a new method and its description to others has become a discipline in itself, with rules and procedures that may be difficult to understand for those new to the field. Accordingly, presentations and publications often have catastrophic quality deficits. The relevant theoretical framework and guiding principles have been developed not only within toxicology, but are used in similar ways in other fields, such as clinical chemistry, hygiene, pharmacy and clinical microbiology. The establishment, validation and documentation of test methods in these different areas are extensively covered in the specialist literature, also including specific recommendations published by regulatory bodies. For instance, OECD TG34 (OECD, 2005) gives guidance for validation of toxicological assays.

Instead of referring to large handbooks and hard-to-read primary literature, we want to give a short, and knowingly incomplete, overview of the major issues requiring attention. This should help to give an orientation to both presenters and reviewers, and may be referred to by editors to prevent publication of the worst papers. Admittedly, also incomplete or bad data sometimes trigger good ideas and approaches, and the sometimes published "one-point correlations" (challenge of a new test with one single compound) can provide first evidence that an assay could work. Therefore, the recommendations given below are rather meant as guidance and aide-memoire than as strict filters, and their respective applicability depends on the specific questions. While the establishment of a measurement method, e.g. for the cellular content of a certain protein or for the measurement of neurite length, may indeed not require the use of a large set of chemicals (or any chemical at all), the requirements become very different (and certainly higher), when this test system is used as the basis for a toxicological test method, e.g. to predict chemical toxicity for cardiac development or axonopathies. Thus, it is highly important for test developers to distinguish between toxicological test systems, test methods and biochemical assays. For the field of *in vitro* toxicology, maintenance of high quality standards, as described below, is essential for upholding its reputation and, in the long run, also for regulatory acceptance.

Tab. 1: Glossary

Term	Definition
Test system	Cellular (or biochemical) system used in a study (e.g. "proliferating hESC", or "neuronally-differentiating PC-12 cells", or "organotypic lung slices"). The term is often used interchangeably with "*in vitro* system". The test system is only one component of a test or test method. Good performance of a test system does not imply good functioning of a test method.
In vitro system	A cell/tissue culture system used as the basis for the development of a test method. Often also called "test system". N.B.: In biochemistry, the term is often used for cell-free systems, as opposed to cellular (living) systems. Cell culture assays, i.e. *in vitro* assays in a toxicological sense, are often called "*in vivo* systems" in biochemistry.
Test method	A procedure based on a *test system*, used to obtain information on the biological effects of a substance. It is characterised by a definitive procedure that produces a test result. Toxicological test methods generate information regarding the ability of a substance or agent to produce a specified biological effect under specified conditions. The term is used interchangeably with "test" and "assay". It can have several *endpoints*.
Endpoint	The biological or chemical process, response, or effect assessed in a *test system by a specific assay*, e.g. "viability" as measured by LDH-release, expression of a marker as measured by PCR, or beating of cardiomyocytes evaluated by an imaging system.
Assay	This term is used in a broader or narrower sense depending on the field, similar to *test method*. In a narrower sense, it can refer to an analytical procedure (e.g. protein determination, PCR). In a wider sense, it implies the use of a defined analytical procedure to determine an endpoint within a test system. A classical example is the Ames assay, which comprises a complex test system of growing and plating bacteria under different conditions together with an analytical procedure based on the counting of colonies. In the narrower sense, many assays yielding multiple endpoints can be performed using one test system. The test method may rely on a complex combination of such endpoints (e.g. a combination of different PCR markers with data obtained by imaging and Western blotting).
Positive/negative control (PC/NC)	An NC for a test is a compound or condition that should not trigger a response, i.e. it should not change the endpoint from baseline. A PC is a compound or condition that triggers a response, i.e. a change of the endpoint from baseline in the right direction and to a certain defined extent. The performance of PC and NC can be used as acceptance criteria of a test.
Acceptance criteria	Criteria defined before performing an assay to determine whether it is "valid", i.e. whether the data can be used. Typical requirements comprise: is the test method functioning (e.g. are the endpoint values for PC and NC in the right range), is the test method performing within the desired range of variability (e.g. are the standard deviations of PC and NC in the right range).
General cytotoxicity (GC)	The term is used when a compound triggers cell death that is not specific for the cell type used in the assay but would occur in most cells at the same concentration and within a similar time frame. For many test methods it is important to measure specific adverse effects that occur at concentrations below those triggering cell death in the test system. Therefore, the verification of test conditions not triggering GC is important for many tests.
Unspecific controls (UC)	This term often refers to compounds displaying GC. For some test systems, it is sufficient to work with PC and NC. For other test systems, it is important to demonstrate a difference between compounds that act specifically, and compounds that lead to changes of the endpoint because they trigger GC. For instance, a test may be designed to determine the metabolic fingerprint of cell cycle blockers. Such a test would require the examination of UC and the comparison of their profile with PC compounds.
Highest non-cytotoxic concentration (HNCC)	This term refers to the highest concentration of a compound that does not trigger GC. The HNCC is important, as it allows the detection of specific adverse effects with highest likelihood. It defines the highest concentration to be used in test systems examining particular toxic effects independent of GC. Testing at concentrations higher than the HNCC may lead to artefacts.

Term	Definition
Replicates within one experiment	These are also called "technical replicates" and can take two different forms: A: the repeated performance of an analysis on the same sample, e.g. duplicate PCR, Western blot or FACS determinations. B: the determination of an endpoint from more than one culture well, with all these wells being incubated in parallel/on the same day/in the same experiment.
Independent experiments	These are also called "biological replicates" and should not be confused with technical replicates in different dishes. A biological replicate is a separate experiment, i.e. on another day, with other cells, new test solutions, etc. A biological replicate can comprise several technical replicates.
Robustness/Ruggedness	Is a measure of a methods' capacity to remain unaffected by small variations in method parameters and environmental conditions. Testing of robustness provides an indication of a test's reliability during normal usage. Sometimes, a distinction is made between robustness and ruggedness. The latter focuses on the reproducibility of the test results obtained for identical samples under normal test conditions that underlie unintentional changes (room temperature, source of human sample material, lot variation of reagents, operator-dependent variables, weather conditions, etc.) Robustness testing would explore the insensitivity of a test to deliberate variations in the test environment or setup (incubation time, temperature, cell passage number, sample storage, cell density, type of culture dish, etc.)

2 Definition of the test method

Recently, criticism has been voiced that animal experiments are often poorly documented (Kilkenny et al., 2009, 2010). Unfortunately, alternative method descriptions often also require more care and consideration of the items compiled below.

2.1 Biological basis of the test system

The description of a test system does not necessarily require the frequently cited standard operation procedure (SOP). However, it is also unlikely that a sufficient description can be given in five lines. The option to deposit online supplementary material upon publication should be used to provide sufficient documentation of the work. Frequent mistakes: lack of source and characterisation of cells; insufficient description of culture conditions for maintenance and experiment; no information on which parameters are critical and what affects them.

2.2 Technical basis of test system and method

This refers to measurement methods, essential instrumentation, important manipulation steps, details on the determination of endpoints and description of the data processing. Frequent mistakes: Descriptions of material incomplete or absent; DNA constructs used to modify cells poorly described and characterised; rules of good cell culture practice (Coecke et al., 2005) neglected and/or not documented.

2.3 Test procedure

A test method does not only consist of a (biological) test system, but also requires definition of when compounds are added, when effects are measured, when samples are taken, how these samples are stored and aliquoted, etc. Recommendation: the standard test procedure should be represented in graphic form, including information on the type(s) of medium, coatings and other manipulations (Fig. 1).

2.4 Rationale for the relevance of the results

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? N.B.: comprehensive answers to these questions are not possible before the completion of a long validation process, but intentions and background considerations should be shared at an early stage.

2.5 Decision criteria for interpretation of the results

This issue is one of the most frequently forgotten. 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)?

3 Definition of basic response characteristics of the test method

The issue of response dynamics has two dimensions: the test method as such, as well as the behaviour of the test in the presence of chemicals. These are separate issues which require inde-

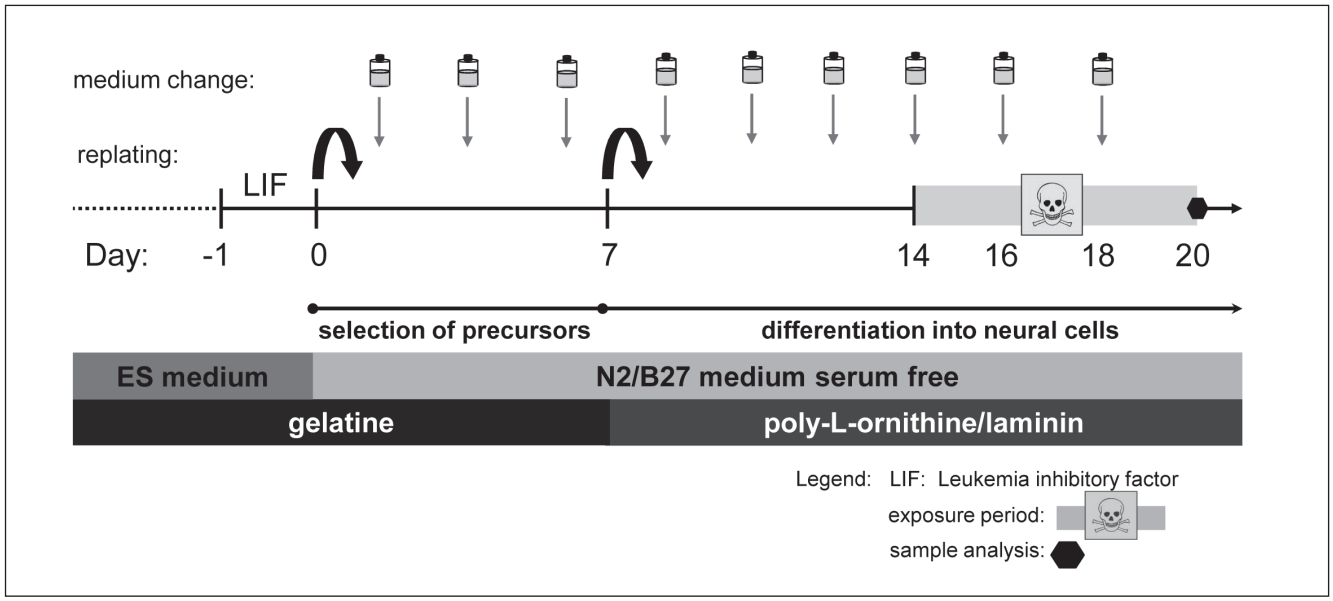

Fig. 1: Example for graphic representation of a test system
The graphic representation indicates the types of substrates used, the types of media used in different phases of the experiment, and it indicates exactly the time points of experimental manipulation, such as medium change, replating of the cells or addition and removal of the toxicant. The presentation of tests in this way creates transparency and facilitates discussion and troubleshooting.

pendent optimisation and characterisation. For instance, measurement of your body weight can be done well on scales (to give a good readout on your general growth characteristics), but this endpoint will hardly respond to acute poisoning. Instead, blood pressure or vomiting activity may be good measures of human poisoning, but they in turn give little information on the growth activity over time.

The minimum information usually required 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 detection limit (e.g.: B + (5 x N)). Another quality parameter of the test system (independent of any test compound) is the z' factor (Zhang et al., 1999), which should ideally be >0.5 and indicates the detection power of the system (z' = 1 - ((3 x (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.

3.1 Plausibility, consistency

These parameters are not as easy to quantify as the z' factor, but they are certainly not less important. Information on consistency can only be obtained by repetition of the experiment. Recommendation: Plausibility should provide a guide as to the required number of repetitions and further investigations. Adaptive responses and other discontinuous phenomena (e.g. two peaked concentration response curves) may indeed occur, but because they are rare, they should be more intensively scrutinised than "typical" curve behaviour. In most cases "strange curves" will turn out to be artefacts of the experiment or of data handling.

3.2 Curve shapes

Everybody has spent years at school on curve discussion. Now, this knowledge can be applied. In pharmacology, the curve slope has mechanistic implications. In toxicology, the underlying biology is much more complex. Most importantly, different mechanisms may apply at different concentration ranges, and two compounds may act very differently although they may appear equipotent. The obvious (the good): 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. The implication: 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}.

The dogma (the bad): Curves always start at 100% (e.g. viability), and mathematical equations must be forced this way. The disappointment: many curves in the literature have no compound concentration at baseline (a no effect concentration, i.e. a concentration of toxicant that still allows 100% function/viability) although they have undergone a peer-review process that should eradicate such fundamental mistakes. The superstition (the ugly): that a compound may NOT show a concentration-

response behaviour. Even though a curve may be steep, it will in most cases still be a curve. What often differs from pharmacological curves: the lower end is not necessarily sigmoidal, but may hit the x-axis at a steep angle.

3.3 Use of semi-qualitative data and scoring methods
This is perfectly acceptable if handled correctly. Scoring and expert judgment have for a long time formed the basis of quantitative histopathology or neurological assessment and still do so in toxicological experiments. However, a transparent presentation of the decision criteria and scoring rules is often neglected. In many cases this requires graphic representation and example images! Morphological classifications must be demonstrated by well-chosen and sufficiently extensive examples. Space limitations are no excuse for neglect of such practice!

4 Biological specificity and mechanistic rationale

4.1 Moving away from the black box
A priori, it may not seem necessary to understand an assay as long as it delivers good (= predictive and reproducible) results. In medicine, this principle is called "he who heals is right". 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* assays (Leist et al., 2008c), not just out of academic interest, but as a basis for a new approach to risk assessment AND 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; similarly, characterisation of new chemotherapeutics based on such high-end methods may simply measure alterations of the cell cycle, which could just as easily be accessed by traditional methods.

4.2 Distinction of cytotoxicity and other effects
Frequently, test methods should assess specific adverse effects (SAE), independent of general cytotoxicity (GC). For instance, inhibition of neurite outgrowth can only be measured meaningfully in a concentration range that does not kill the cells. 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 (3-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} x 0.02). Mistakes and dangers 1: Determination of general cytotoxicity under different experimental conditions than used for SAE determination. Ideally, GC should be determined in parallel/simultaneously with SAE. Mistakes and dangers 2: Inability to measure GC does not mean that it does not occur. This applies in particular to short term assays (few hours), as most GC endpoints require several hours to become manifest.

5 Data on compound response dynamics

Here, the dimension of compound testing is addressed, and not that of the test system as such (dealt with in 2.). Minimum set of requirements: does a compound that should change the endpoint do this – and by how much does it do this (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). The response dynamics of a PC, and thus the performance of the test method, cannot be qualified without assessing the response to UC. Advice: after initial use of a few PC, NC and UC, and optimisation of the test method with their help (e.g. choice of different readouts or timing), it may be useful to re-challenge the test method with a new set of PC and NC to assess its performance with respect to unknown compounds.

6 Statistical considerations

6.1 Statistics in a non-mathematical sense
Statistics deals with identifying the source and extent of variance in the test method. Information on this must be provided throughout every paper section and figure. Although its theoretical background may be hard to understand, the intentions and results of statistics are of the utmost importance. Even more important than the use of the most appropriate method is the creation of transparency, i.e. allowing the readers themselves to estimate the type and extent of variance of the data. If there is a genuine wish to disclose this information, the suitable form and mathematics will follow. In the same vein, it may not be so important whether standard deviations or standard errors are shown, or whether 2 or 4 replicates were done. The major issue is that all the essential information is disclosed so that the readers can judge the data themselves.

6.2 What are replicates?
Talking about replicates touches a sore spot of the majority of published work. Again, the major issue is not mathematics or a set of strict rules. As above, it is a matter of transparency and common sense. The experimenter needs to know where the main source of variance lies and then to demonstrate the extent of this variance. If the major source of variation is the performance of separate experiments on different days, then the most meaningful replicates are the number of separate experiments done on different days. In such cases, it is little meaningful to pipet a PCR reaction three times from the same sample or to have eight parallel wells incubated the same way to demonstrate that the

experimenter was able to dispense with 98% accuracy. In other cases (more rarely), variation between different experiments may be minimal, while different technical replicates are hard to standardise. In this case, a low number of biological replicates and a high number of technical replicates are advisable. These basic considerations are important both for experimental design and for reporting the results. As a simple rule of thumb for most test methods the experiment should be performed completely independently three times to obtain an estimate of the variance of the results.

6.3 Acceptance criteria

Acceptance criteria become important when we transition from simple, extremely robust systems to the commonly used complex test systems with endpoints that can be affected by multiple known and unknown parameters. In simple words: when we measure the weight of a chemical or the temperature of a room, we usually accept the outcome of this experiment without further acceptance criteria. Under slightly more complex or stringent or dangerous or expensive conditions, we automatically start questioning whether we can trust the data and asking what would make us trust the data. The scales or pH electrode would be exposed to a test sample of known weight or pH, the thermometer responsiveness would be controlled by brief exposure to ice or something warm, and the exact readout would be tested by some reference method. These activities must not be confused with calibrations! They are first and foremost 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, but still manuscripts are sometimes submitted that largely neglect this concept. 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. Only if these controls fulfil the acceptance criteria, can the other experimental data be used. The reversal of this rule is often hard to accept, but still mandatory: data from an experiment that did not fulfil the acceptance criteria cannot be used.

6.4 Robustness and ruggedness

Especially complex systems require that a method is not only described but that robustness testing is also performed and steps and conditions of low robustness are identified. This may sound very technical, and may by some be considered an issue reserved for professionals developing methods for commercialisation. However, considerations of robustness are in fact known by most biologists who read articles in Nature Methods now and again. The descriptions in this journal, in contrast to many others, include extensive troubleshooting, which also involves a robustness analysis of the different steps.

6.5 Correlation and causality

Some short notes on a big issue: we all know (but still need to be reminded sometimes) that correlation does not mean causality, even if the correlation is very good. A less trivial point is that a correlation in an experimental system does not necessarily mean a correlation in the real world. On the one hand, the correlation may be 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 (Leist et al., 2008b). 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 are comprised of 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 the many 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 one of them. Typical examples are -omics studies suggesting a correlation between some metabolite or protein modification with toxicity (Groebe et al., 2010; West et al., 2010). For such studies, well-chosen statistics use measures to counteract the effect of multiple endpoints on apparent significances of effects. Without multiple endpoint correction, and assuming a 5% significance level, five out of 100 endpoints will show up as false-positive correlations. Omission of all negative data will then give the false impression that a new biomarker has been discovered.

7 Solvents and compound quality

If one examines the *in vitro* literature describing the developmental toxicity of lead, one will find that a sizeable fraction of the work does not give information on what type of lead ion was used. Similar problems are found in the literature on arsenic toxicity. Such information should be an absolute minimum requirement, but, in addition, it may also be important to give information on contaminants. For many small, organic molecules, especially of (semi)-natural origin, the composition of isomers and enantiomers varies from supplier to supplier and from lot to lot, as does the degree of purity. It should be a matter of course that such data are included in the materials and methods, i.e. definition of the chemical, at least by the CAS number, and/or the supplier's exact catalogue number. N.B.: the latter information is only apparently exact and reproducible. At the current frequency of supplier mergers and changes of the product spectrum, such information may sometimes not be useful at a later stage.

Once the compounds have been well-defined, various other problems remain, and the most important one involves the solvents used for assays. A simple rule of thumb for best practice is the restriction to a small set of solvents, preferably DMSO only. However, it should be noted that DMSO ages

(change of smell), and that it can be obtained in largely differing qualities (with more than 1000-fold price difference). Moreover, it can be used at different concentrations, and many publications do not include the final concentration in each assay, but rather only report the highest concentration used in the whole study.

If the few small issues named above are addressed, a large number of experiments will already be reported in sufficient detail. However, for some compounds, additional problems can be encountered. For instance, the testing of organic solvents (e.g. toluene) is hampered by their tendency to evaporate and by their poor solubility in cell culture medium. Both need to be taken into account in the experimental design and the documentation of the results.

8 Data presentation

The presentation of data in graphic form has the purpose of generating a transparent and easily understandable overview of a larger set of results. This frequently necessitates manipulation of the primary data, e.g. performing normalisations. With all such procedures, it should still be possible for the reader to deduce the underlying primary data. A typical example is the release of lactate dehydrogenase (LDH) from cells as a measure of toxicity. This is sometimes indicated as % increase compared to untreated controls. Such data presentations are not acceptable if they do not indicate the extent of release in the control cultures (if it is 2%, then doubling does not mean a high toxicity; if it is 50%, then more than a doubling cannot be expected, even for the strongest toxicant).

A bad habit is the presentation of concentrations on a weight/volume basis instead of as molarity. Under such conditions the chloride vs. iodide salts of compounds show apparently largely different toxicity, and homologues within a family of related but differently-sized compounds also appear to act differently. In this context it is also important to distinguish between dose and concentration. For most *in vitro* assays, the concentration-response curve is the desired outcome. The dose is only rarely meaningful. It is a measure of the amount or weight of a compound applied to an organism. The general concept of concentration vs. dose implies that if the volume in a cell culture dish is doubled under otherwise identical conditions, then the dose per well doubles, while the concentration stays the same. This may illustrate that the dose-concept is in most cases not useful for *in vitro* experiments.

Of course, some important exceptions apply to the above rules under non-equilibrium conditions and under conditions of repeated exposure. Compounds may distribute non-homogeneously in cell culture dishes, and e.g. accumulate in certain parts of the cell. For instance, the exposure to low concentrations of a lipophilic compound may lead to different active site concentrations, i.e. a different cellular dose depending on the time of exposure and, depending on such phenomena, the overall dose may play a role in addition to the actual average concentration in an assay (Blaauboer, 2010).

9 Conclusions

We hope that the information given here is useful for many entering the field of *in vitro* methods, and that it also may serve as orientation for those who judge the quality of manuscripts describing new test systems. The points mentioned above should be seen as guidance on all the issues that need consideration. Whether they then apply to a respective system, and whether action is required will require individual judgement. A series of follow-up articles may also initiate a transition to a more interactive form of publishing, allowing comments and suggestions of the reader on web-based tools.

Stimulation of discussion has always been a major intention of ALTEX, for instance in the food for thought articles (Leist et al., 2008b; Hartung and Leist, 2008; Hartung, 2010a,b, 2009, 2008a,b, 2007; Bottini and Hartung, 2009; Rovida, 2010; Hartung and Koëter, 2008), in addition to the traditional provision of information on diverse topics related to animal welfare (Animal welfare perspective: Kolar and Rusche, 2008; Schmidt, 2008; Sauer, 2009b; Kamber, 2009, Eskes et al., 2009; Stingl et al., 2009) and the teaching of 3R methods (Jukes, 2008, 2009; Leist, 2006; Hartung et al., 2009; Daneshian et al., 2010). ALTEX has welcomed articles on test method validation (Rovida and Hartung, 2009; Vogel, 2009; Hartung and Rovida, 2009), the regulatory context for such methods (Hartung et al., 2010) and also on funding for such method development (Devolder, 2008). A focus has always been on new technologies to be developed by specialists in other fields (Mitterhauser and Toegel, 2008; Favre, 2008; Knight, 2008) and on addressing specific technical problems within already established concepts (e.g. Rothen-Rutishauser, 2008; Heindl et al., 2008; Wanner and Schreiner, 2008; Li, 2008a,b; Hagelschuer et al., 2009; Bahramsoltani et al., 2009; Manzer et al., 2009; Pelkonen et al., 2009; Sauer et al., 2009; Kamusella, 2009; Ullrich et al., 2009). This also involved the presentation of new test systems based on cells (Henn et al., 2009; Leist et al., 2008a; Kuegler et al., 2010; Schindler et al., 2009) in addition to *in silico* approaches (Vedani et al., 2009; Sauer, 2009a; Goris et al., 2009; Hartung and Hoffmann, 2009).

References

Bahramsoltani, M., Plendl, J., Janczyk, P. et al. (2009). Quantitation of angiogenesis and antiangiogenesis in vivo, ex vivo and in vitro – an overview. *ALTEX 26*, 95-107.

Blaauboer, B. J. (2010). Biokinetic modeling and in vitro-in vivo extrapolations. *J. Toxicol. Environ. Health B Crit. Rev. 13(2-4),* 242-252.

Bottini, A. A. and Hartung, T. (2009). Food for thought … on the economics of animal testing. *ALTEX 26*, 3-16.

Coecke, S., Balls, M., Bowe, G. et al. (2005). Second ECVAM taskforce on good cell culture practice. *ATLA 33(3),* 261-287.

Daneshian, M., Leist, M. and Hartung, T. (2010). The center for alternatives to animal testing – Europe (CAAT-EU): a transatlantic bridge for the paradigm. *ALTEX 27*, 63-69.

Devolder, T., Reid, K., Rogiers, V. et al. (2008). A review of national public funding programmes in European countries. *ALTEX 25*, 233-242.

Eskes, C., Moura Sá-Rocha, V., Nunes, J. et al. (2009). Proposal for a Brazilian centre on alternative test methods. *ALTEX 26,* 295-298.

Favre, D. (2008). Reverse transcriptase activity of hepatitis B virus polymerase in eukaryotic cell extracts in vitro. *ALTEX 25,* 197-211.

Goris, N. E., Eblé, P. L., Jong, M. C. M. and Clercq, K. (2009). Quantification of foot-and-mouth disease virus transmission rates using published data. *ALTEX 26*, 52-54.

Groebe, K., Hayess, K., Klemm, M. et al. (2010). Protein biomarkers for in vitro testing of embryotoxicity. *J. Proteome Res*. (Epub ahead of print).

Hagelschuer, I., Hauff, P., Strube, A. et al. (2009). Luminescent imaging technology as an opportunity to reduce and refine animal experiments: light at the end of the tunnel? *ALTEX 26*, 177-185.

Hartung, T. (2010a). Food for thought … on alternative methods for chemical safety testing. *ALTEX 27*, 3-14.

Hartung, T. (2010b). Food for thought ... on alternative methods for nanoparticle safety testing. *ALTEX 27*, 87-95.

Hartung, T., Bruner, L., Curren, R. et al. (2010). 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 27*, 43-51.

Hartung, T. (2009). Food for thought … on evidence-based toxicology. *ALTEX 26*, 75-82.

Hartung, T. and Hoffmann, S. (2009). Food for thought … on in silico methods in toxicology. *ALTEX 26*, 155-166.

Hartung, T. and Rovida, C. (2009). That which must not, can not be…a reply to the echa and edf responses to the reach analysis of animal use and costs. *ALTEX 26*, 307-311.

Hartung, T., Blaauboer, B. and Leist, M. (2009). Food for thought … on education in alternative methods in toxicology. *ALTEX 26*, 255-263.

Hartung, T. (2008a). Food for thought … on animal tests. *ALTEX 25*, 3-9.

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

Hartung, T. and Koëter, H. (2008). Food for thought … on food safety testing. *ALTEX 25*, 259-265.

Hartung, T. and Leist, M. (2008). Food for thought … on the evolution of toxicology and the phasing out of animal testing. *ALTEX 25*, 91-96.

Hartung, T. (2007). Food for thought … on validation. *ALTEX 24*, 67-72.

Heindl, C., Hess, A. and Brune, K. (2008). Refinement and reduction in animal experimentation: options for new imaging techniques. *ALTEX 25*, 121-125.

Henn, A., Lund, S., Hedtjärn, M. et al. (2009). The suitability of BV2 cells as alternative model system for primary microglia cultures or for animal experiments examining brain inflammation. *ALTEX 26*, 83-94.

Jukes, N. (2009). Alternatives across Latin America: catalysing change in the curriculum. *ALTEX 26*, 55-57.

Jukes, N. (2008). Russia: Update on animal experiments and alternatives in education. *ALTEX 25*, 56-62.

Kamber, M. G., Steiger, A., Aus der Au, C. et al. (2009). Interspecies crosses: aspects of animal protection. *ALTEX 26,* 227-231.

Kamusella, P. C., Wissgott, C., Grosse-Siestrup, C. et al. (2009). A model of isolated, autologously hemoperfused porcine slaughterhouse lungs. *ALTEX 26*, 279-284.

Kilkenny, C., Browne, W. J., Cuthill, I. C. et al. (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PloS Biol*. *8(6),* e1000412.

Kilkenny, C., Parsons, N., Kadyszewski, E. et al. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One*. *4(11),* e7824.

Knight, A. (2008). Non-animal methodologies within biomedical research and toxicity testing. *ALTEX 25*, 213-231.

Kolar, R. and Rusche, B. (2008). An animal welfare perspective on animal testing of GMO crops. *ALTEX 25*, 127-130.

Kuegler, P. B., Zimmer, B., Waldmann, T. et al. (2010). Markers of murine embryonic and neural stem cells, neurons and astrocytes: reference points for developmental neurotoxicity testing. *ALTEX 27*, 17-42.

Leist, M., Bremer, S., Brundin, P. et al. (2008a). The biological and ethical basis of the use of human embryonic stem cells for in vitro test systems or cell therapy. *ALTEX 25*, 163-190.

Leist, M., Kadereit, S. and Schildknecht, S. (2008b). Food for thought … on the real success of 3R approches. *ALTEX 25*, 17-24.

Leist, M., Hartung, T. and Nicotera, P. (2008c). The dawning of a new age of toxicology. *ALTEX 25,* 103-114.

Leist, M. (2006). What can a chair on alternatives to animal experimentation effectuate? *ALTEX 23*, 211-213.

Li, A. P. (2008a). Human hepatocytes as an effective alternative experimental system for the evaluation of human drug properties: General concepts and assay procedures. *ALTEX 25*, 33-42.

Li, A. P. (2008b). In vitro evaluation of human xenobiotic toxicity: scientific concepts and the novel integrated discrete multiple cell co-culture (IdMOC) technology. *ALTEX 25*, 43-49.

Manzer, A. K., Lombardi-Borgia, S., Schäfer-Korting, M. et al. (2009). SV40-transformed human corneal keratocytes: Optimisation of serum-free culture conditions. *ALTEX 26*, 33-39.

Mitterhauser, M. and Toegel, S. (2008). An in vitro model for the comparative evaluation of bone seeking pharmaceuticals. *ALTEX 25*, 51-55.

OECD (2005). OECD environment, health and safety publications series on testing and assessment No. 34. Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. ENV/JM/MONO (2205) 14. Available at: http://www.oecd.org/document/30/0,3343,en_2649_34377_1916638_1_1_1_1,00.html

Pelkonen, O., Tolonen, A., Rousu, T. et al. (2009). Comparison of metabolic stability and metabolite identification of 55 ECVAM/ICCVAM validation compounds between human and

rat liver homogenates and microsomes – a preliminary analysis. *ALTEX 26*, 214-222.

Rothen-Rutishauser, B., Mueller, L., Blank, F. et al. (2008). A newly developed in vitro model of the human epithelial airway barrier to study the toxic potential of nanoparticles. *ALTEX 25*, 191-196.

Rovida, C. (2010). Food for thought ... why no new in vitro tests will be done for REACH by registration. *ALTEX 27*, 175-183.

Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals – a report by the transatlantic think tank for toxicology (t^4). *ALTEX 26*, 187-208

Sauer, U. (2009a). Animal and non-animal experiments in nanotechnology – the results of a critical literature survey. *ALTEX 26*, 109-128.

Sauer, U. (2009b). Political incentives towards replacing animal testing in nanotechnology? *ALTEX 26*, 285-294.

Sauer, U. G., Wächter, T., Grune, B. et al. (2009). Go3R – semantic internet search engine for alternative methods to animal testing. *ALTEX 26*, 17-31.

Schindler, S., Aulock, S., Daneshian, M. and Hartung, T. (2009). Development, validation and applications of the monocyte activation test for pyrogens based on human whole blood. *ALTEX 26*, 265-277.

Schmidt, K. (2008). Dignity or integrity – Does the genetic modification of animals require new concepts in animal ethics? *ALTEX 25,* 313-320.

Stingl, L., Völkel, M. and Lindl, T. (2009). 20 years of hypertension research using genetically modified animals: no clinically promising approaches in sight. *ALTEX 26*, 41-51.

Ullrich, A., Stolz, D. B., Ellis, E. C. et al. (2009). Long term cultures of primary human hepatocytes as an alternative to drug testing in animals. *ALTEX 26*, 295-302.

Vedani, A., Smiesko, M., Spreafico, M. et al. (2009). VirtualToxLab™ – in silico prediction of the toxic (endocrine-disrupting) potential of drugs, chemicals and natural products. Two years and 2,000 compounds of experience. *ALTEX 26*, 167-176.

Vogel, R. (2009). Alternatives to the use of animals in safety testing as required by the EU-Cosmetics Directive 2009. *ALTEX 26*, 233-236.

Wanner, R. and Schreiner, M. (2008). An In vitro assay to screen for the sensitizing potential of xenobiotics. *ALTEX 25*, 115-120.

West, P. R., Weir, A. M., Smith, A. M. et al. (2010). Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. *Toxicol. Appl. Pharmacol. 247(1),* 18-27.

Zhang, J. H., Chung, T. D. and Oldenburg, K. R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen 4(2),* 67-73.

Acknowledgement

This work was supported by the Doerenkamp-Zbinden Foundation, the FP7 ESNATS project, the DFG international research training school IRTG1331, and the Konstanz Research School Chemical Biology of the German Excellence Initiative.

Correspondence to

Prof. Marcel Leist, PhD
Doerenkamp-Zbinden Chair for in vitro Toxicology
and Biomedicine
University of Konstanz
Box M657
78457 Konstanz
Germany
e-mail: marcel.leist@uni-konstanz.de