

FORUM SERIES, PART III

A Toxicology for the 21st Century—Mapping the Road Ahead

Thomas Hartung¹

Johns Hopkins University, Bloomberg School of Public Health, Department for Environmental Health Sciences, Center for Alternatives to Animal Testing, Chair for Evidence-based Toxicology, Baltimore, MD 21231

Received February 19, 2009; accepted March 13, 2009

The landmark publication by the National Research Council putting forward a vision of a toxicology for the 21st century in 2007 has created an atmosphere of departure in our field. The alliances formed, symposia and meetings held and the articles following are remarkable, indicating that this is an idea whose time has come. Most of the discussion centers on the technical opportunities to map pathways of toxicity and the financing of the program. Here, the other part of the work ahead shall be discussed, that is, the focus is on regulatory implementation once the technological challenges are managed, but we are well aware that the technical aspects of what the National Academy of Science report suggests still need to be addressed: A series of challenges are put forward which we will face in addition to finding a technical solution (and its funding) to set this vision into practice. This includes the standardization and quality assurance of novel methodologies, their formal validation, their integration into test strategies including threshold setting and finally a global acceptance and implementation. This will require intense conceptual steering to have all pieces of the puzzle come together.

Key Words: toxicity testing.

The willingness to accept risks in daily life is diminishing continuously. The willingness to lessen efforts of safety evaluations is therefore low. Thus, a generally accepted consensus is that new approaches must not lower the current safety standards. This leads quickly to a concept, where current methodologies are considered gold standards, which need to be met. It is therefore not sufficient to develop new approaches, but also to show their limitations in comparison to current regimes, a process normally summarized as validation. However, this typically leads to a strategy where out of the “patch-work” of the toxicological tool-box, maximally one patch is replaced by a new one. This does not really open up for a new general approach.

We might take a different view and ask ourselves, how regulatory toxicology would be done, if we had to design it from the scratch. The vision of the NRC committee (Andersen and Krewski, 2009; National Research Council, 2007) is laying out

such a new design, putting forward a new approach based on modern technologies and in a more integrated way (Hartung and Leist, 2008; Leist *et al.*, 2008). This includes likely the accumulated knowledge on pathways of toxicity, modern technologies such as (human) cell culture, omics technologies (genomics, proteomics, metabonomics), image analysis, high-through-put testing, *in silico* modeling including PBPK (physiology-based pharmacokinetics, here more toxicokinetic modeling) and QSAR (quantitative structure activity relationships). Federal agencies have already joined forces to attempt this (Collins *et al.*, 2008). For the purpose of this article let us assume the feasibility of this approach. It is unrealistic to assume that it will be a one test approach, which does the whole job—likely it will be another test battery (tool-box). However, it is the first hypothesis put forward that we can only gain if we construct this new approach from scratch and not only replace or add new “patches.” We shall explore here, which fundamental problems remain, and if such a battery of novel tests can be achieved, independent of the technologies to be applied.

CHALLENGE 1: TESTING STRATEGIES INSTEAD OF INDIVIDUAL TESTS

Today’s approach to regulatory testing is rather simple: one problem, one test. Limitations of this approach have been discussed earlier (Hartung, 2008b) especially when considering the low prevalence of most hazards (Hoffmann and Hartung, 2005). It is important to note that *in vitro* tests do not have less limitations than the *in vivo* ones (Hartung, 2007b). A toxicology based on pathways is one which is likely based on various tests, be it in a battery (i.e., where all tests are done to derive the results) or a test strategy (i.e., where tests are done based on interim decision points). Our experiences with the first approach are poor and bad: Combining typically three mutagenicity tests in a battery resulted in a disaster of accumulating false-positives (Kirkland *et al.*, 2005); only 3–20% of positive findings are real-positives—hardly an efficient strategy. We therefore need other ways to combine tests for the different pathways in a different way, but we have neither a terminology for test strategies nor tools to compose or validate them.

¹ For correspondence via Fax: +1 410 955 0617. E-mail: thartung@jhsph.edu.

CHALLENGE 2: STATISTICS AND MULTIPLE TESTING

When testing for multiple pathways, we will need to correct our statistics for multiple testing. We have to lower significance levels accordingly or we will run increasingly into false-positive findings. The proponents of the new approach assume more than a hundred, and less than a thousand such pathways. A lot of multiple testing... Assuming only 100 pathways, significance levels of $p = 0.05$ would have to be lowered to 0.006 using the most common Bonferroni correction. This—likely with sophisticated methods of high inherent variance—will result in an astronomic number of replicates necessary: For the example of $p = 0.05$ and stable noise/signal ratios, a 71-fold increase in sample size (e.g., number of animals or replicate cellular tests) is necessary to reach the same level of confidence.

CHALLENGE 3: THRESHOLD SETTING

Where does a relevant effect start? Certainly not where we can measure a significant change. What is measurable depends only on our detection limits, and in the case of multiendpoint methods a lot on signal/noise relation and the inevitable number of false-positive results. If, for example, a toxicogenomics approach is taken, several thousand genes might be measured and, especially when low thresholds of fold-induction are used, false-positive events will occur. Even if real-positive, the questions arising are then, whether this is significant with the given number of replicates, or even more important, whether this is relevant (notably completely different questions). Although the former can be tested with replicate testing and statistics (see, however, problem of multiple testing), the relevance is more difficult to establish: The more remote we are in (sub-) cellular pathways, the more difficult to extrapolate to the overall organism. The NRC vision document is not really clear here, whether we talk of cells and their signal transduction pathways or the even more complicated physiological pathways of in dynamic systems with compensatory mechanisms.

What does it mean if a pathway is triggered but if accompanied by some compensatory ones as well? We definitively have to overcome the mentality of “we see an effect, this is an effect level.” Any method, which assesses only a certain level of the organism (e.g., the transcriptome when using genomics), will be questioned whether these changes are translated to the higher integration levels (proteome, metabolism, physiology). This argues for systems biology approaches where such considerations are taken into account, but complexity of modeling increases dramatically, with impacts on standardization, costs, feasible number of replicates etc. The greater the distance from the primary measurement to the overall result in a model, the more difficult threshold setting will become because of error propagation.

Setting of thresholds or other means of deriving a test result (data analysis procedure) is a most critical part of test

development. It determines the sensitivity and specificity of the new test, that is, the proportion of false-positive and false-negative results. Noteworthy, this needs to be done before validation, but we already need a substantial “training set” of substances to derive the data analysis procedure, that is, the algorithm to convert raw experimental results into a test result (positive/negative, highly toxic, moderate, mild. . .). This raises the question, where such reference results come from?

CHALLENGE 4: WHAT TO VALIDATE AGAINST?

The first problem is that the choice of the point of reference determines where we will arrive. If the new toxicology is based on animal tests as the reference, we can only approach this “gold standard” but will not be able to overcome its shortcomings. We have suggested (Hoffmann *et al.*, 2008) the concept of composite reference points, that is, a consensus process of identifying the reference result attributed to a reference test substance. This allows at least the investigator to review and revise individual results, but does not change the main problem that mostly animal data are only available.

The second problem is that it is unlikely that we will be able to evaluate the entire pathway-based test strategy in one step. So the question becomes what to validate against, if we have only partial substitutes? If we have the perfect test for a pathway of toxicity, where are the data on substances to test against—we will typically not know which of the many pathways was triggered *in vivo*. As a way forward we have proposed a “mechanistic” validation (Coecke *et al.*, 2007), where it is shown that the prototypic agents affecting a pathway are picked up while others not expected to do so are not.

There are various challenges to the validation process as it is right now (Hartung *et al.*, 2004), which we have discussed elsewhere (Ahr *et al.*, 2008; Hartung, 2007a). Especially for the complex omic technologies and other information-rich and demanding technologies, we are only starting to see the challenges for validation (Corvi *et al.*, 2006).

CHALLENGE 5: HOW TO OPEN UP REGULATORS FOR CHANGE?

We have coined the term “postvalidation” (Bottini *et al.*, 2008) to describe the cumbersome process of regulatory acceptance and implementation. It is increasingly recognized that it is neither the lack of new approaches nor their proven reliability by validation, but that translation into regulatory guidelines and use is now the bottle-neck of the process.

Change requires giving up on something not to add to it. As long as most new approaches are considered “valuable additional information,” the incentive to drive new approaches through technical development, validation and acceptance is rather low, given 10–12 years of work of large teams and costs of several hundred thousand dollars. The process is so demanding

because regulatory requirements often mandate virtually absolute proof that a new method is equal to or better than traditional approaches. Most importantly, to let go from tradition requires seeing the limitations of what is done today. This discourse was too long dominated by animal welfare considerations. This has been convincing for parts of the general public, but the scientific and regulatory arena is much less impressed by this argument.

Especially, if personal responsibility and liability come into play, traditional tests are rarely abandoned. Costs are not too much of an issue, because they are the same for competitors and become simply part of the costs of the products. In general, costs are less than 1%—often 0.1%—of the turnover of regulated products (Bottini and Hartung, 2009). Thus, the major driving force for change would be that we can do things better. However, limitations are not very visible in this field and the interest to expose them is low. In order to identify limitations, we would need, first of all, transparency of data and decisions, and establishment of reproducibility and relevance of our approaches. Neither of this is given: data are typically not published and/or are considered proprietary; repeat experiments are often even excluded by law and data on the human health effects are rare for comparison (Hartung, 2008a). This leaves us in a situation, where “expertise,” that is, the opinion of experts rules, whereas hard data (“evidence”) are short in supply.

Interestingly, clinical medicine is in a similar situation, that is, facing the coexistence of traditional and novel scientific approaches. Here, however, information overload rather than lack of data is the problem. A remarkable process has taken place over the last two decades, which is called Evidence-based Medicine (EBM). The Cochrane Collaboration includes more than 16,000 physicians and has so far produced about 5000 systematic reviews compiling the available evidence for an explicit medical question in a transparent and rigorous process. Among others this has stimulated the development of meta-analysis tools in order to combine systemically information from various studies.

We (Hoffmann and Hartung, 2006) and others (Guzelian *et al.*, 2005) have put forward the idea to initiate a similar process for toxicology. Consequently, the first International Forum toward Evidence-based Toxicology (EBT) was held in 2007 bringing together more than 170 stakeholders from four continents (www.ebtox.org, Griesinger *et al.*, in press). Three main areas of interest emerged (1) a systematic review of methods (similar to the review of diagnostic methods in EBM), (2) the development of tools to quantitatively combine results from different studies on the same or similar substances (analogous to meta-analyses); and (3) the objective assessment of causation of health as well as environmental effects. This movement is still in its infancy. However, it promises to help with a key obstacle, that is, identifying the limitations of current approaches, and thus might be the door-opener for any novel approach. Due to its transparency and rigor in approach, judgments are likely more convincing than classical scientific

studies and reviews. At the same time, the objective compilation of conclusions from existing evidence requires the development of tools, which will have broader impact, especially the meta-analysis type of methods or scoring tools for data quality. The latter has been furthered as a direct outcome of the EBT forum in a contract and expert consultation by the European Centre for the Validation of Alternative Methods (ECVAM) (Schneider *et al.*, unpublished data). It aims to base the well-known “Klimisch scores” (Klimisch *et al.*, 1997) for data quality on a systematic set of criteria, one set each for *in vivo* and for *in vitro* data. This might have enormous impact for the requested systematic use of existing data, for example, for the European REACH registration process of existing chemicals. It provides the means to include, reject or weigh existing information before decisions on additional test needs are taken, or for combined analysis.

With regard to the novel toxicological approaches, however, most important will be that existing and new ways are assessed with the same scrutiny. Sound science is the best basis for the selection of tools. Validating against methods believed to do a proper job is only betting and will always introduce uncertainty about the compromise made while forgetting about the compromise represented by the traditional method.

The term of evidence-based toxicology must not be confused with weight-of-evidence approaches, which describe an often personal judgment of the different information available to come to an overall conclusion, for example, in genotoxicity. Such approaches have also been suggested in the validation process (Balls *et al.*, 2006), but they represent rather compromise solutions in the absence of final proof. In contrast, EBT aims to use all evidence reasonably available to come to a judgment in a transparent and objective manner.

CHALLENGE 6: THE GLOBAL DIMENSION

A central obstacle for the introduction of new approaches is globalization of markets. Globally acting companies want to use internationally harmonized approaches. This means that change to new approaches if not forced by legislation, will occur when the last major economic region has agreed on the new one. A teaching experience (Hartung, in press) was the Local Lymph Node Assay (LLNA) in mice (notably an *in vivo* alternative method) to replace guinea pig tests such as the Buehler and guinea pig maximization test. Since 1999, the LLNA is the preferred method in Europe reinforced by an animal welfare legislation, which requests such reduction and refinement methods to be used whenever possible, and since 2001, the test is OECD-accepted. Still, in 2008, we found that out of about 1450 new chemicals registered with skin sensitization data since 1999 in Europe, only about 50 had LLNA data, whereas the rest was still tested in the traditional tests. This illustrates the resistance to change even when “only” exchanging one animal test by another. We can imagine, how much more difficult this

will be for *in vitro* or *in silico* approaches, or the complex new approaches aimed for now.

This means that efforts to create a novel approach need global buy-in. National solutions will quickly encounter nontechnical problems for implementation and acceptance. When lobbying for programs to identify pathways of toxicity (or more general of interaction of small molecules with cells) project, the aim should be for a global program from the start, for example, similar to the human genome project.

We should, however, not be too negative about the impact of globalization. As discussed earlier (Bottini *et al.*, 2007), this might as much constitute a driver for change as it is now an obstacle. International harmonization is an opportunity to export standards of safety assessments to trade partners (Bottini and Hartung, 2009).

CHALLENGE 7: QUALITY ASSURANCE FOR THE NEW APPROACH

For the global use of methods, it does not suffice to agree on how to test. If we want to accept approaches executed at other places, challengeable quality standards for performance and documentation of tests must exist, as they have been developed as OECD Good Laboratory Practice (GLP) or various ISO standards. GLP was, however, developed mainly for the dominating *in vivo* tests. Building on a workshop to identify the gaps for a GLP for *in vitro* approaches (Cooper-Hannan *et al.*, 1999), and the parallel development of Good Cell Culture Practices (Coecke *et al.*, 2005; Hartung *et al.*, 2002), now some OECD guidance for *in vitro* toxicology is available (OECD, 2004). However, for the complex methods envisaged for the new type of toxicity tests will require a much more demanding quality assurance. We must learn how to report properly the results from new methods like genomics or QSARs—we have to imagine, how difficult it will be to report

in a standardized manner the whole process. A key problem will be the fluid nature of the new methodologies: standardization and validation requires freezing things in time, every change of method requires re-evaluation not possible for the complex methodologies. On the contrary, we see continuous amendments of *in silico* models or new technologies (e.g., gene chips). Shall we validate and implement a certain stage of development and close the door for further developments? This is exactly what is required for international agreements on methods—and it is difficult to imagine for complex methods still under development.

CHALLENGE 8: HOW TO CHANGE WITH STEP BY STEP DEVELOPMENTS BECOMING NOW AVAILABLE?

Things would be easy if a new regulatory toxicology would become available at once—we might then compare old and new and decide to change. But we will continue to receive bits and pieces (Fig. 1) as we have already experienced for a while. When should we make a major change and not just add and replace patches? What is not clear is, where the mastermind for change will come from. Which group or institution will lead us through the change? Given the substantial efforts for the development of each piece, we can not wait for their implementation until everything is ready.

The first of two solutions is to implement the new methods in parallel to gain experience with the new without abandoning the old. Beside the costs, this will create the problem of what to do with discrepant information. Although we need this on the one hand (or we will not result in something new), we will not be able to neglect any indications of hazard from new tests (we have lost our innocence), even though they have not yet taken over. The second opportunity is to start with those areas where we have new problems and explore the new opportunities. This might include new health endpoints such as endocrine disruption, developmental neurotoxicity, respiratory sensitization or new products such as biologicals, genetically modified organisms, nanoparticles, or cell therapies. But in both cases we might fall into the trap of just adding new patches without substantial change. This means we have to somehow to organize a transition. Many people working at different angles of the whole will not create the “new deal.”

CHALLENGE 9: HOW TO ORGANIZE TRANSITION?

Beside the technological challenge, we have identified the need for systematic combination of approaches (integrated testing), and a program to assess objectively current approaches, to validate them and to implement them. This program requires out-of-the-box thinking, that is, intellectual steering (Fig. 2).

As a first step, with the financial support of the Doerenkamp-Zbinden foundation (<http://www.doerenkamp.ch/en/>) which have been created in recent years five professorships for alternative

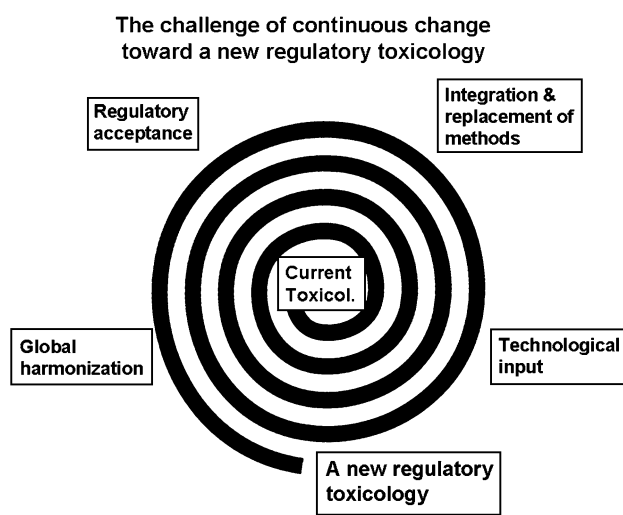


FIG. 1. The continuous process of development, integration and acceptance of new methods.

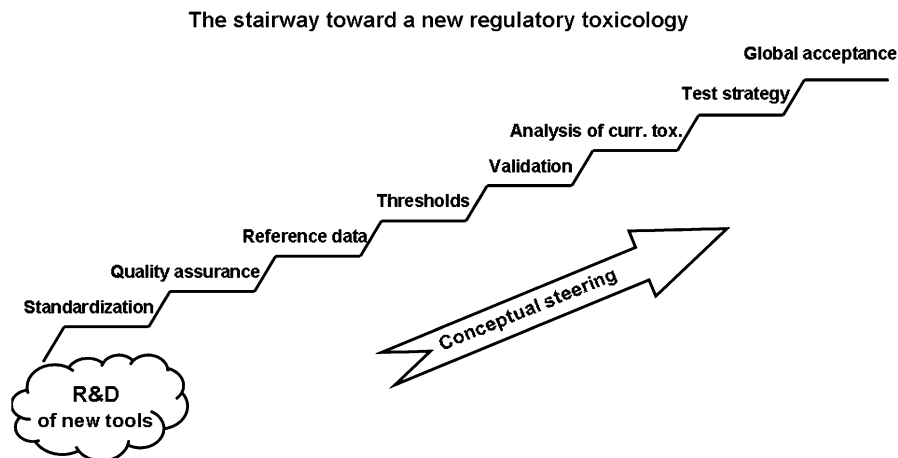


FIG. 2. The steps necessary after the technical solutions to arrive at a new approach of regulatory toxicology.

approaches (e.g., at the universities of Erlangen, Konstanz, Utrecht, Geneva, and most recently, Johns Hopkins in Baltimore). Moreover, a Transatlantik Think Tank of Toxicology (T⁴) was created (ALTEX, 2008) between the toxicological chairs of these institutions, which aims to collaborate on dedicated studies and analyses, and workshops to support the paradigm shift in toxicology.

Certainly this is only the first little step, but it might form a nucleus for further initiatives. Similarly, the Forum series in Toxicological Sciences, and discussions at the SOT meetings, furthers the shaping and sharpening of ideas.

CHALLENGE 10: MAKING IT A WIN/WIN/WIN SITUATION

Three major stakeholders will have to collaborate to create the new toxicology, that is, the academia, regulators and the regulated communities in industry. This collaboration is still more an exception (for example the European Partnership for Alternative Approaches (EPAA), between 40 companies, 7 trade associations, and the European Commission) than the rule (Bottini and Hartung, 2009; Hartung, 2008c). Academia has not been involved, although research funding and emerging technologies may help to increase academic engagement. The time for validation and acceptance of new methods of about one decade make this area only of little attraction for academics, and most are not in a dialogue thereby enabling them to understand the needs of industry and regulators. The exchange between industry and regulators is also often poor, perhaps driven by concern that providing more information is only giving more opportunity for further requests. In the United States, we lack both the public/private partnership and the research funding into alternative approaches. By basing the novel toxicology less on animal welfare and more on sound science considerations, this might change in the future. The sheer dimensions of the tasks ahead will require a trans-disciplinary, trans-national, trans-

stakeholder, and trans-industrial sectors approach. Information hubs such as AltWeb (<http://altweb.jhsph.edu/>), AltTox (<http://www.alttox.org/>), EPAA (<http://ec.europa.eu/enterprise/epaa/>), EBTox (<http://www.ebtox.org/>), ECVAM (<http://ecvam.jrc.it/>), and the Center for Alternatives to Animal Testing (<http://caat.jhsph.edu/>) have a key role here.

There is gain for all players including the following: the challenge of the development of new approaches; the better understanding of limitations of our assessments; the likely development of safer products with new test approaches; and the international harmonization prompted by a major joint effort. There is economic gain as well (Bottini and Hartung, 2009), but while we are talking broadly about science, ethics and politics, this has not been sufficiently addressed. The stairway to “Regulatory Toxicology version 2.0” is steep, but the goal merits the effort.

REFERENCES

- Ahr, H.-J., Alepee, N., Breier, S., Brekelmans, C., Cotgreave, I., Gribaldo, L., Dal negro, G., De Silva, O., Hartung, T., Lacerda, A., *et al.* (2008). Barriers to validation. A report by European Partnership for Alternative Approaches to animal testing (EPAA) working group 5. *ATLA Altern. Lab. Anim.* **36**, 459–464.
- ALTEX. (2008). EU/USA: A transatlantic think tank of toxicology (T⁴). *ALTEX Altern. Anim. Exp.* **25**, 361.
- Andersen, M. E., and Krewski, D. (2009). Toxicity testing in the 21st century: Bringing the vision to life. *Toxicol. Sci.* **107**, 324–330.
- Balls, M., Amcoff, P., Bremer, S., Casati, S., Coecke, S., Clothier, R., Combes, R., Corvi, R., Curren, R., Eskes, C., *et al.* (2006). The principles of weight of evidence validation of test methods and testing strategies. The Report and Recommendations of ECVAM Workshop 58. *ATLA Altern. Lab. Anim.* **34**, 603–620.
- Bottini, A. A., Alepee, N., De Silva, O., Hartung, T., Hendriksen, C., Kuil, J., Pazos, P., Philips, B., Rhein, C., Schifflers, M.-J., *et al.* (2008). Optimization of the post-validation process. The report and recommendations of ECVAM workshop 67. *ATLA Altern. Lab. Anim.* **36**, 353–366.

- Bottini, A. A., Amcoff, P., and Hartung, T. (2007). Food for thought . . . on globalization of alternative methods. *ALTEX Altern. Anim. Exp.* **24**, 255–261.
- Bottini, A. A., and Hartung, T. (2009). Food for thought . . . on economic mechanisms of animal testing. *ALTEX Altern. Anim. Exp.* **26**, 3–16.
- Coecke, S., Balls, M., Bowe, G., Davis, J., Gstraunthaler, G., Hartung, T., Hay, R., Merten, O.-W., Price, A., Schechtman, L., et al. (2005). Guidance on good cell culture practice. *ATLA Altern. Lab. Anim.* **33**, 261–287.
- Coecke, S., Goldberg, A. M., Allen, S., Buzanska, L., Calamandrei, G., Crofton, K., Hareng, L., Hartung, T., Knaut, H., Honegger, P., et al. (2007). Incorporating *in vitro* alternative methods for developmental neurotoxicity into international hazard and risk assessment strategies. *Environ. Health Perspect.* **115**, 924–931.
- Collins, F. S., Gray, G. M., and Bucher, J. R. (2008). Toxicology: Transforming environmental health protection. *Science.* **319**, 906–907.
- Cooper-Hannan, R., Harbell, J. W., Coecke, S., Balls, M., Bowe, G., Cervinka, M., Clothier, R., Hermann, F., Klahm, L. K., de Lange, J., et al. (1999). The principles of Good Laboratory Practice: Application to *in vitro* toxicology studies. The report and recommendations of ECVAM workshop 37. *ATLA Altern. Lab. Anim.* **27**, 539–577.
- Corvi, R., Ahr, H. J., Albertini, S., Blakey, D. H., Clerici, L., Coecke, S., Douglas, G. R., Gribaldo, L., Groten, J. P., and Haase, B. (2006). Validation of toxicogenomics- based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. *Environ. Health Perspect.* **114**, 420–429.
- Griesinger, C., Hoffmann, S., Kinsner-Ovaskainen, A., Coecke, S., and Hartung, T. Foundations of an evidence-based toxicology. Proceedings of the First International Forum Towards Evidence-Based Toxicology. Conference Centre Spazio Villa Erba, Como, Italy. 15–18 October 2007. *Hum. Exp. Toxicol.* (in press).
- Guzelian, P. S., Victoroff, M. S., Halmes, N. C., James, R. C., and Guzelian, C. P. (2005). Evidence-based toxicology: A comprehensive framework for causation. *Hum. Exp. Toxicol.* **24**, 161–201.
- Hartung, T. (2007a). Food for thought . . . on validation. *ALTEX Altern. Anim. Exp.* **24**, 67–72.
- Hartung, T. (2007b). Food for thought . . . on cell culture. *ALTEX Altern. Anim. Exp.* **24**, 43–147.
- Hartung, T. (2008a). Towards a new toxicology—Evolution or revolution? *ATLA Altern. Lab. Anim.* **36**, 633–639.
- Hartung, T. (2008b). Food for thought . . . on animal tests. *ALTEX Altern. Anim. Exp.* **25**, 3–9.
- Hartung, T. (2008c). Food for thought . . . on alternative methods for cosmetics safety testing. *ALTEX Altern. Anim. Exp.* **25**, 147–162.
- Hartung, T. Toxicology kissed awake. *Nature* (in press).
- Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Roi, A. J., et al. (2004). A modular approach to the ECVAM principles on test validity. *ATLA Altern. Lab. Anim.* **32**, 467–472.
- Hartung, T., Balls, M., Bardouille, C., Blanck, O., Coecke, S., Gstraunthaler, G., and Lewis, D. (2002). Report of ECVAM task force on good cell culture practice (GCCP). *ATLA Altern. Lab. Anim.* **30**, 407–414.
- Hartung, T., and Leist, M. (2008). Food for thought . . . on the evolution of toxicology and phasing out of animal testing. *ALTEX Altern. Anim. Exp.* **25**, 91–96.
- Hoffmann, S., Edler, L., Gardner, I., Gribaldo, L., Hartung, T., Klein, C., Liebsch, M., Sauerland, S., Schechtman, L., Stammati, A., et al. (2008). Points of reference in validation—The report and recommendations of ECVAM Workshop. *ATLA Altern. Lab. Anim.* **36**, 343–352.
- Hoffmann, S., and Hartung, T. (2005). Diagnosis: Toxic!—Trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations. *Toxicol. Sci.* **85**, 422–428.
- Hoffmann, S., and Hartung, T. (2006). Towards an evidence-based toxicology. *Hum. Exp. Toxicol.* **25**, 497–513.
- Kirkland, D., Aardema, M., Henderson, L., and Müller, L. (2005). Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity. *Mutat. Res.* **584**, 1–256.
- Klimisch, H.-J., Andreae, M., and Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul. Toxicol. Pharmacol.* **25**, 1–5.
- Leist, M., Hartung, T., and Nicotera, P. (2008). The dawning of a new age of toxicology. *ALTEX Altern. Anim. Exp.* **25**, 103–114.
- National Research Council. (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. National Research Council Committee on Toxicity Testing and Assessment of Environmental Agents, The National Academies Press, Washington, DC. Available at: http://www.nap.edu/catalog.php?record_id=11970. Accessed April 10, 2009.
- OECD. (2004). Series on Principles of Good Laboratory Practice and Compliance Monitoring, Number 14, Advisory Document of the Working Group on Good Laboratory Practice, The Application of the Principles of GLP to *in vitro* Studies. Available at: [http://appli1.oecd.org/olis/2004doc.nsf/linkto/env-jm-mono\(2004\)26](http://appli1.oecd.org/olis/2004doc.nsf/linkto/env-jm-mono(2004)26). Accessed April 10, 2009.