Food for Thought ... on the Evolution of Toxicology and the Phasing out of Animal Testing

Thomas Hartung[1] and Marcel Leist[2]
[1]EC Joint Research Centre, IHCP/ECVAM, Ispra, Italy; [2]Doerenkamp-Zbinden Chair for alternative in vitro methods, Faculty of Natural Sciences and Mathematics, University of Konstanz, Germany

There is something brewing in the field of toxicology: Last year's vision and strategy document published by the US National Academy of Sciences (NRC, 2007) has excited many toxicologists on both sides of the Atlantic. In February 2008 several American agencies announced a coalition to set this into practice (www.sciencemag.org/content/vol319/issue5865/index.dtl): "We propose a shift from primarily *in vivo* animal studies to *in vitro* assays, *in vivo* assays with lower organisms, and computational modeling for toxicity assessments". In USA Today of the same day we find a comment by Francis Collin, Director of the National Human Genome Research Institute: "*[Toxicity testing] was expensive, time-consuming, used animals in large numbers and didn't always work*". In the same article, Elias Zerhouni, Director of the NIH, is cited: "*Animal testing won't disappear overnight, but the agencies' work signals the beginning of an end*." We have never heard anything like that from US federal agency representatives before. What is going on? What can we really expect and when?

Hypothesis 1: Like every science, toxicology is evolving – the pressing need for adaptation to scientific progress in regulatory toxicology might make the next step a revolutionary change

Science develops in waves (Kuhn, 1970), often prompted by new technologies or societal needs. Toxicology is in no way different, and risk assessments evolve over time (Sexton et al., 1995; Henry,

2003). Toxicology has seen both these engines at work, e.g. with the introduction of small laboratory animal research almost a hundred years ago or on account of scandals like around thalidomide (Contergan). Regulatory toxicology is still at an evolutionary stage that largely reflects the state of the art of its creation, with the addition of only some more modern "patches" since. This series has discussed earlier (Bottini et al., 2007) that, unique to a scientific discipline, regulatory toxicology has been frozen in time by creating (test) guidelines. There is no doubt about the importance of this process to achieve security of the regulated community and to allow international harmonisation (mutual acceptance of data), but we observe more and more disconnect between the development of a mechanistic, modern toxicology and the rigid approaches in the

regulatory field. The international coordination of regulatory testing requirements were the most efficient measures to reduce animal testing in the past by avoiding duplicate testing. Ironically, this has now turned into an obstacle to introducing new (animal-free) methods. Figure 1 tries to capture the evolvement of toxicology, starting with the historic "art of poisoning", since already the old Romans collected very detailed information on the effects of poisons; it also incorporates more recent developments and possible future directions, such as evidence-based toxicology and systems toxicology. We have elaborated on the concept of an evidence-based toxicology earlier (Hoffmann and Hartung, 2006) and will thus limit that topic here to a small summary and an update on some recent developments. In contrast, the concept of a systems toxicol-
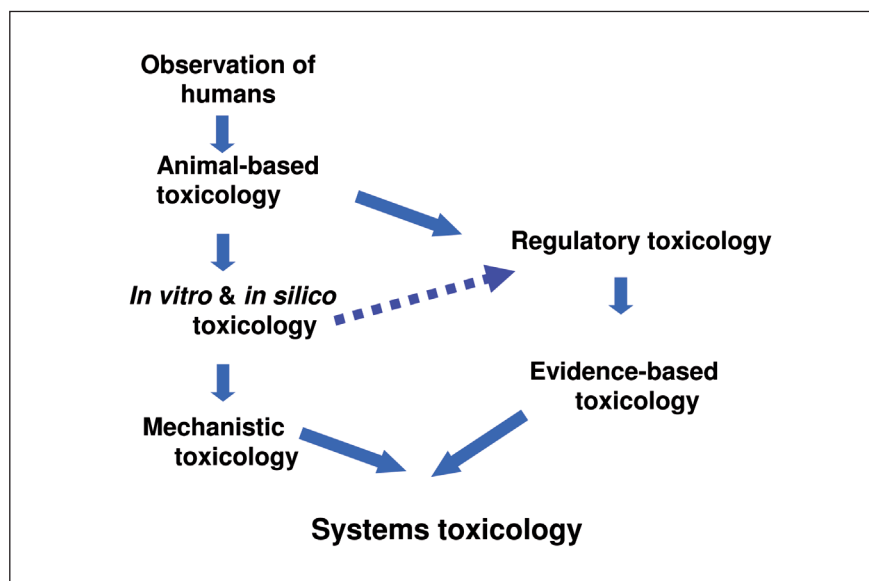

Fig. 1: The evolution of toxicology

ogy and its possible role in the regulatory arena will require more detailed discussion (see also Leist et al., this issue).

In this article we develop the vision that modern, mechanistic toxicology, which became separated from empirical/descriptive toxicology (Fig. 2), can be reunited with the latter; the emerging opportunities are a critical self-appraisal (evidence-based toxicology) and the introduction of modern techniques (systems toxicology).

Hypothesis 2: Toxicology can learn from clinical medicine and borrow some concepts from evidence-based medicine

The desire to understand the risk posed by food and substances around us is as old as the desire to cure diseases. Thus, many attempts and traditions have evolved and continue to coexist, although science has taken over for the last two centuries in both disciplines. Prejudice and tradition are difficult to eradicate, especially if the new scientific approaches are emerging slowly and are often only understood by specialists.

Even for specialists it can be hard to extract the correct information from the large amount of available information: This problem of information flooding has been described by Hall 1997 and Davidoff 1995 for the field of medicine in general. More than 10,000 medical journals publish more than two million articles per year. An internist needs to read 17 articles per day to be able to practice up to date diagnostics and therapy, but according to surveys he or she has only 30-60 minutes per week available for reading. A reliable condensation of knowledge is therefore required. Setting the highest standards of quality assurance and transparency, the Cochrane collaboration (www.cochrane.org) aims to do exactly this, structure the available evidence for a given treatment or diagnostic problem and make it available in a database. Involving about 16,000 physicians, in the meantime about 5,000 guideline documents of an evidence-based medicine (EBM) have been developed (Mayer, 2004).

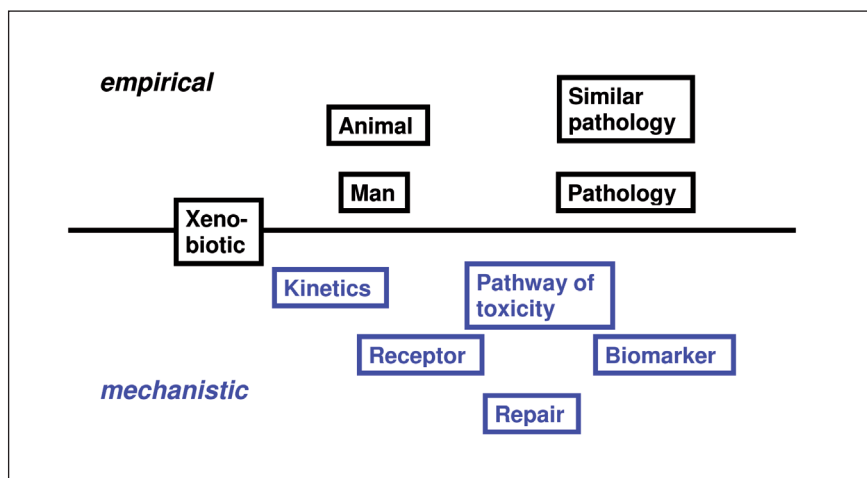My (T.H.) personal thought starter to translate this to toxicology was a book

Fig. 2: Empirical versus mechanistic toxicology

written by my friend Eddy Neugebauer "Handbook of mediators in septic shock" (Neugebauer and Holaday, 1993). Together with his co-author he, to my knowledge for the first time, applied the approach of EBM not just to clinical data, as is common practice, but also to animal data and even to *in vitro* data. Since then, the idea of translating this to toxicology has grown, and Sebastian Hoffmann and I have further elaborated the idea (Hoffmann and Hartung, 2006) and furthered it with the ECVAM team and collaborators of the 1st International Forum "Toward an evidence-based toxicology" (www.ebtox.org). The proceedings of this forum are well on their way, and we can restrict ourselves here to a few notions.

EBTox is not alternative methods with another name. The principal problem of regulatory toxicology is not taking up new methods but replacing the existing ones. In the absence of mechanisms of critical review, we tend to form traditions of thinking, which no longer question our approach. Without a thorough analysis of our toolbox, we will only see toxicology develop like patchwork art. EBM teaches us that a rigorous fact-finding and interpreting process can result in consensus for change where necessary. No process can substitute for good science, but we can systematically assess what is good science, what is not and where we cannot judge yet. One of my favorite sayings is "you do not have to be able to lay an egg to taste a bad one"; it is not necessary to have something better or to be

able to propose it to identify limitations and shortcomings (as well as strengths) of current approaches.

Hypothesis 3: New technologies allow a new approach to toxicology

Cell culture methodology has been steadily improving over the last decades, making most cells of the organism available today, often even in organotypic culture systems. Still, primary cells isolated directly from tissue samples represent the closest approximation to the physiologic condition. For human cells, this means restrictions with regard to availability, usually limiting the isolation of cells to places closely linked to surgical facilities. It also implies that often diseased organs or old donors are the source of cells. Due to cryopreservation and an increasing number of providers, at least frozen cells are now available for many systems, however, limited by costs and quantity. The most exciting prospects are currently emerging from stem cell technologies, which promise to make (human) cells available in high quality and sufficient quantity (Bremer et al., 2004; Stumann et al., 2008). The appropriate protocols to reliably generate differentiated cells showing the characteristics of the tissue situation are still rare. Especially, the purity and defined differentiation status of the generated cells is still a problem. Still, in a near future, high-quality cell

preparations of many human cell types should become available.

At the same time, cell culture is moving from simple endpoints and manual handling to high-throughput and high-content measurements. This means that automation is allowing the parallel or repeated execution of experiments on often miniaturised cell models, which allows testing many substances on the same system or one substance on various variations of the cell model. This technology is most advanced for pharmaceutical screening; practically every major company now has a substance bank of hundreds of thousands of substances, which are tested in cellular and sub-cellular screening assays of interest. In toxicology, the dimensions of through-put are considerably lower. Major activities include the ToxCast programme (www.epa.gov/comptox/toxcast) and the Tox Testing Collaboration of the National Toxicology Program (www.ntp.niehs.nih.gov/) in the US, and the InViTech action between ECVAM and the Nanotechnology and Molecular Imaging unit at the EU Joint Research Centre (http://nmi.jrc.it/projects/InViTech.htm). It appears timely to create the necessary substance repositories to fully exploit these approaches, e.g. a collection of a large number of REACH substances would allow making available data for the preliminary risk assessment within the REACH process before animal tests are even suggested. The prospects of high-throughput testing shall be the subject of a future "Food for thought …" article.

In contrast to high-throughput testing, which allows obtaining information on many substances, high-content methods produce a lot of data from one test system for each substance. The most prominent examples are certainly the omics technologies, especially genomics, proteomics and metabolomics. While the three most advanced omics technologies differ in what and how things are measured, typically gene chips, gel electrophoresis, mass spectroscopy and nuclear magnetic resonance (NMR), they have in common that multiple endpoints (usually of the same kind) are recorded, requiring biostatistical approaches and data-mining. Noteworthy, these technologies can also be combined with *in vivo* test systems,

opening up refinement and reduction opportunities. For example, early predictive derangements might allow shorter treatment schemes, more sensitive measurements might allow low-dose approaches, and multiple redundant endpoints reduce variability to reduce group sizes.

At the moment three principal uses of omics, which bring new qualities into toxicology, are obvious:

1. The parallel measurement of many endpoints might be used to identify those which are predictive and can be used further on as biomarkers, requiring in case validation; the identified biomarker might then be measured with more selective methods, such as PCR or ELISA, in the future.

2. Typical patterns of toxicity might be derived, which characterise pathologies, without even necessarily identifying what the individual signal represents (e.g. by using the mathematical procedure of principal component analysis); in this case, the omics approach would be used to screen for similar patterns induced by new substances.

3. Increasingly the interpretation of omics data is facilitated by knowledge of pathways, how for example different genes and gene products interact. The term "systems biology" was coined for this type of guided data integration, sometimes integrating even different omics technologies. There are expectations that this can be translated to toxi-

cology to form a "systems toxicology" approach. However, this would require identifying the underlying pathways of toxicity.

Similarly, we see an enormous development of other high-content methods due to automated image analysis. This can be done on cell or tissue level. In general, standardisation and wide-spread use are more limited here than for omics technologies, but further developments can be expected in the short term. Figure 3 captures how these technologies interact and could eventually lead to the development of a "systems toxicology". In essence, various new, information-rich technologies, combined with established scientific knowledge (knowledge of biochemical pathways, knowledge of patterns/toxicity signatures; knowledge of biomarkers; knowledge of pharmacokinetic and chemical properties) using computational approaches.

The reductionistic approaches of first generation *in vitro* methods (single cell systems of often limited functionality) has frequently been criticised to fail because of principles of chaos theory. Can information on toxicity to a whole organism realistically be obtained by dissecting it into components and then reassembling it? The new field of systems biology has shown promising results that this may indeed be feasible.

Rapid developments are taking place for all steps of the diagram, but a ma-
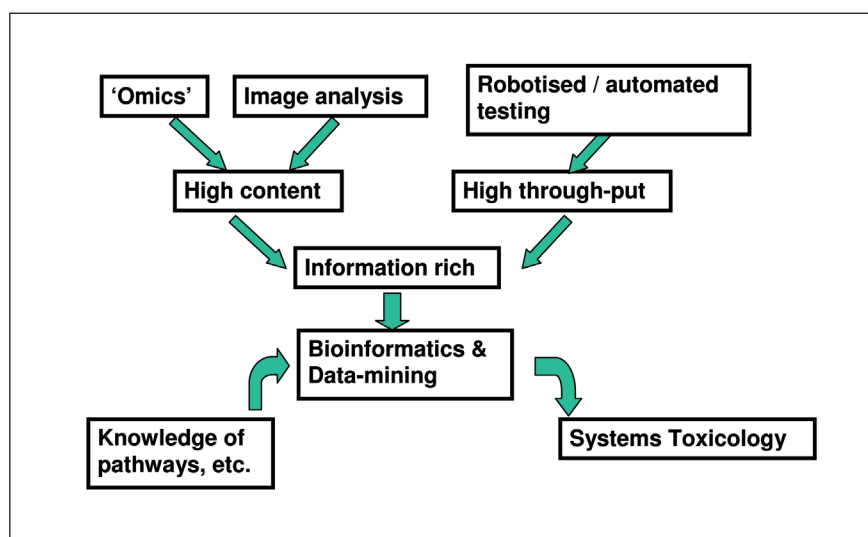

Fig. 3: Toward systems toxicology

jor concern is the lack of coordination of developments. This often means that comparisons and combined analysis of techniques is hampered. It is most important that agreed priority substances are primarily used by the different developer groups. Similarly, expert consensus could give guidance on how the best cell systems can be combined with the tests and on the various opportunities for data analysis.

Hypothesis 4: The synergy of different scientific and political developments on both sides of the Atlantic gives hope for a global project on "Human Response to Chemicals"

Our knowledge of how small molecules affect cell, organ and organism function is limited to very few substances. Of an estimated about 140,000 chemicals on the market we have broad toxicological data for less than 10% and know beneficial effects for even less. The mode of action of most of these is an enigma. Our knowledge on biochemistry and cellular players has exploded during the last century. A systematic analysis of pathways of interaction for these effectors would give us the tools for research to switch on and off certain cellular functions, identify the targets for drug action that facilitate specific drug development, and, last but not least, characterise the pathways of toxicity that need to be known to move to a new kind of toxicology as described the in the vision and strategy document of the US National Academy of Sciences cited earlier.

We should be clear that such an approach requires means similar to the human genome project and can only be achieved in a similar international collaboration. The US initiative to form a coalition of major agencies to set the NAS vision into practice represents a very promising top-down approach. In Europe, a variety of activities aim for similar goals in a more bottom-up approach of many non-coordinated projects. Various EU sponsored projects have started to explore omics technologies for toxicology (e.g. www.Carcinogenomics.eu, www.predictomics.com,

http://comics.vitaminb.com). More recently, the Netherlands have taken initiative by creating a National Toxicogenomics Centre (www.toxicogenomics.nl) and are in an ongoing discussion about a possible ASAT (Assuring Safety without Animal Testing) project (www.asat-initiative.eu). Together with smaller approaches by COLIPA, ECETOC, etc. – some still at the planning stage – all in all a substantial research potential exists, which could form the nucleus for such a global project. It will be essential to develop a platform for the coordination of setting up and steering such collaboration. Due to the lack of European coordination, it looks at this moment as if the discussion will be dominated by the US, which is surprising, since the EU had a certain opinion leadership for a long time due to the larger efforts and political driving forces. However, it has to be noted that the US initiative has not yet safeguarded major funding, and the dimension of the challenge calls for a global effort. It is also important to note that the combination of bottom-up and top-down approaches has many advantages, as has the combination of massive technological power and organisational capacity with the creativity and innovation capacity of smaller decentralised units. Finally, it does not matter where a scientific effort is carried out, but a sensitive area such as environmental health and public safety should be high up on our priority list, not least for capacity and competence building. Last but not least, we must avoid disconnect between efforts on both sides of the Atlantic in the interest of the global industry. When moving from a globally largely standardised level of safety assessments at low-tech level to a high-tech level, we must not fall back to nationalised standards. We have to admit that some of the drive of European legislative pushes, such as the 7th amendment of the Cosmetics Directive and the REACH legislation, is bringing us back to separate paths, and it would be detrimental if now Europe missed the opportunity to join the emerging technological revolution (Leist et al., this issue).

Hypothesis 5: The different scientific approaches must be used in combination to phase out animal experimentation

No single approach will overcome a given regulatory animal testing approach alone; even for the established replacement methods we increasingly learn about non-applicability for certain substances or problems of either sensitivity or specificity, which require combination with other methods. Figure 4 aims to show a certain hierarchy of approaches, going from simple (top) to more complex (bottom), i.e. to derive judgement from structure is least difficult, while combining multiple endpoints with kinetic modelling is most difficult. The hierarchy results because it does not seem reasonable to use sophisticated and expensive methods if simpler approaches can do the job. As indicated, this needs to be done for every health or ecological effect of interest. The contribution of the different approaches to actually reduce animal tests will differ for every endpoint and over time with the advancement of the technologies and their integration. Therefore, no relative contribution is indicated for each approach on purpose, using the same height for each step of this staircase. The figure first displays technologies that are to different extents suitable for making positive (toxic) and negative (non-toxic) judgements. This is indicated by placing them either to the left (positive call) or right (negative call) of the stair. Given that the majority of substances has no toxic effect (Hartung, 2008a), the negative calls are actually more important for reducing animal numbers.

The different individual steps are:
1. Structural considerations, i.e. some substances can already be identified as chemically inert or too large for uptake (bioavailability). Noteworthy, this might be complemented with systems testing to determine whether the substance can be metabolised or degraded, e.g. in acidic conditions of the stomach. If there is no indication for either, testing should be waived.
2. No exposure, no risk. This is a hot topic between different geographical and product areas, between industry

and regulators: Should our safety assessment be driven by hazard or by risk. Do we really need to bother about possible hazardous properties of substances when exposure is negligible? There are arguments for either side: We can save a lot of testing, money and time if we only test what humans and the environment are actually exposed to. However, use and exposure patterns might change once a substance is on the market, and we also have to foresee accidental exposure, for example at workplaces. Still, when giving priority, there is good reason why for example REACH asks to waive testing if exposure is negligible. The approach has been formalised and further developed under the name "thresholds of toxicological concern (TTC)" (Kroes et al., 2000; Kroes et al., 2007). This suggests not only empirical levels below which no toxicological activity can be reasonably expected, but also suggests that these levels might differ for chemical classes, mainly because different chemistry represents different bioavailability.

3. The idea might be carried further by identifying internal TTCs, i.e. threshold peak plasma levels of toxicological concern, by measuring actual bioavailability. Increasingly, it is recognised how valuable toxicokinetic information for chemicals would be in order to carry out risk assessments and especially to integrate *in vitro* and *in vivo* data. Micro-dosing approaches allow gaining such information with minimal animal numbers and suffering and can even be performed in humans. It would be even more logical to define plasma levels of substances that are too low to be harmful instead of looking at external exposure levels. To the best of our knowledge, no such analysis has so far been carried out, probably also due to the limited data available for kinetics of general chemicals, with few exceptions of e.g. dioxins, phatalates, and heavy metals such as lead in occupational medicine, where the concept of "internal dose" has been proposed.

4. Classical *in vitro* and *in silico* alternatives qualify best to identify hazards (positive calls) in contrast to the pre-
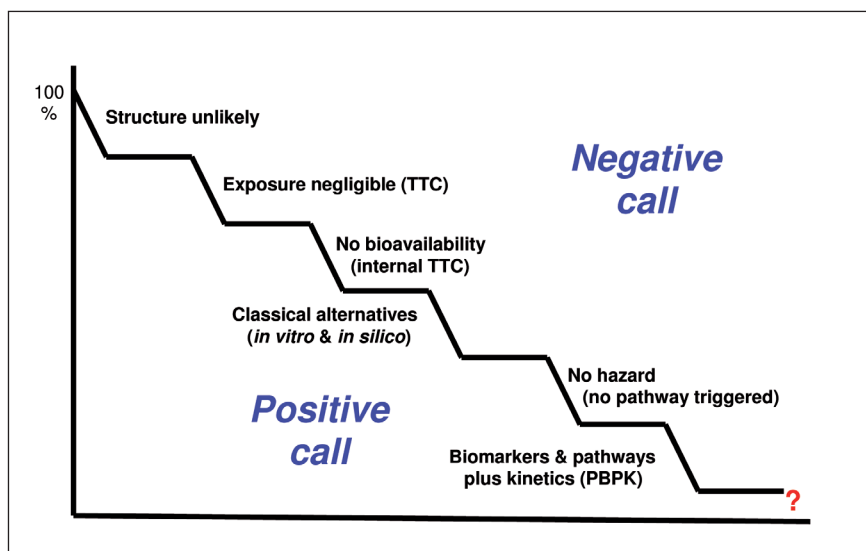

Fig. 4: The use of various approaches to phase out animal testing

vious approaches. Due to the typical limitations (metabolism, contribution from immune reactions such as inflammation) it is often difficult to exclude a mode of action not represented in this reduced system. How much uncertainty this actually leaves (Hartung, 2007b) and how much uncertainty there is in the animal test (Hartung, 2008), with possibly misleading different metabolism or defence systems, can be argued, but the principal concern remains. We have to accept that the reduction from organism to cell culture with all its artefacts or even to an algorithm gives an incomplete picture. A central problem for regulatory use will be that hazards can be identified but an extrapolation to dose can typically not be done (with the possible exception of acute toxicity). These methods are thus more suitable for classification and labelling, but will often trigger the animal test for quantitative assessments. The way forward is the combination with biokinetic modelling (see 6).

5. If we are able to identify the array of relevant pathways of toxicity for a given endpoint, it might be possible to create a test battery which represents these. For example, it has been suggested that embryotoxicity occurs by interfering with 17 different pathways (NRC, 2000). It is tempting to set up a battery of tests which assemble these and to

demonstrate with prototypic substances the competence of the test strategy ("mechanistic validation"), as we suggested earlier also for neurodevelopmental toxicology (DNT) (Coecke et al., 2007), where, however, we are talking about different cellular processes and not biochemical pathways in a narrower sense. We should be aware that, again, such an approach would primarily qualify to exclude hazard; a quantitative risk assessment cannot be done using such an approach.

6. The positive identification of toxicants, by pathways triggered, hazard identified by classical alternatives or other biomarkers, will often require a translation to relevant dose. Even though the uncertainty of such translations of doses to humans has an enormous uncertainty (Hartung, 2008), the traditional risk assessment is based on such information, e.g. the NOAEL (no observed adverse effect level). The only promising non-animal approach here is reverse biokinetic modelling, i.e. instead of calculating the fate of a substance and the resulting tissue level, it has to be calculated, which dose might cause a tissue level equivalent to the effect concentration *in vitro*. There are promising first approaches (Corvi et al., 2006; Bouvier d'Yvoire et al., 2007), but they are far from general applicability. They represent, however,

the key opportunity to broaden the applicability of *in vitro* methods.

On purpose, the staircase in Figure 4 is not drawn to reach to the ground, and there is a question mark for the extent of animal use remaining. This will depend on the scientific progress of each component for every endpoint. This includes also limitations of applicability domain for each approach, i.e. if certain approaches cannot be used for a part of the chemical universe, we might have problems to offer an alternative. It also depends on our belief in current approaches and what constitutes an adequate replacement. The importance of the EBT approach to further clarify the limitations of current approaches cannot be overestimated. It also has a societal component of how much risk we are willing to accept. Erich Kästner's famous quote "Leben ist immer lebensgefährlich" (Life is always life-threatening, own translation) nicely expresses that there is no zero-risk option. The illusion of perfect risk assessment might be as dangerous, as it prevents us from carefully following up the effects of substances once the risk assessment has been completed.

References

Bottini, A. A., Amcoff, P., Hartung, T. (2007). Food for thought … on globalisation. *ALTEX 24*, 255-261.

Bouvier d'Yvoire M., Prieto P., Blaauboer BJ., Bois FY., Boobis A., Brochot C., Coecke S., Freidig A., Gundert-Remy U., Hartung T., Jacobs MN., Lavé T., Leahy DE., Lennernäs H., Loizou GD., Meek B., Pease C., Rowland M., Spendiff M., Yang J. and Zellmarker M. Physiologically-based Kinetic Modelling (PBK Modelling): Meeting the 3Rs Agenda. The Report and Recommendations of ECVAM Workshop 63. *Altern. Lab. Anim. 2007, 35*, 661-671.

Bremer, S. and Hartung, T. (2004). The use of embryonic stem cells for regulatory developmental toxicity testing in vitro – the current status of test development. *Curr. Pharm. Des. 10*, 2733-2747.

Coecke S., Goldberg AM., Allen S., Buzanska L., Calamandrei G., Crofton K., Hareng L., Hartung T., Knaut H., Honegger P., Jacobs M., Lein P., Li A., Mundy W., Owen D., Schneider S., Silbergeld E., Reum T., Trnovec T., Monnet-Tschudi F. and Bal-Price A. (2007). Incorporating in vitro alternative methods for development neurotoxicity into international hazard and risk assessment strategies. *Environ Health Persp., 115*, 924-931

Corvi, R., Ahr, H.-J., Albertini, S. et al. (2006). Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/ NICEATM considerations for regulatory use. *Environ. Health Persp. 114*, 420-429.

Davidoff, F., Haynes, B., Sacket, D. and Smith, R. (1995). Evidence based medicine. *Br. Med. J. 310*, 1085-1086.

Hall, J. C. and Platell, C. (1997). Half-life of truth in surgical literature. *Lancet 350*, 1752.

Hartung, T. (2007a). Food for thought on … validation. *ALTEX 24*, 67-73.

Hartung, T. (2007b). Food for thought on … cell cultures. *ALTEX 24*, 3-9.

Hartung, T. (2008). Food for thought on … animal tests. *ALTEX 25*, 3-10.

Henry, C. J. (2003). Evolution of toxicology for risk assessment. *Int. J. Toxicol. 22*, 3-7.

Hoffmann, S. and Hartung, T. (2006). Towards an evidence-based toxicology. *Human Exp. Toxicol. 25*, 497-513.

Kroes, R., Galli, C., Munro, I. et al. (2000). Thresholds of toxicological concern for chemical substances present in the diet: a practical tool for assessing the need for toxicity testing. *Food Chem. Toxicol. 38*, 255-312.

Kroes, R., Renwick, A. G., Feron, V. et al. (2007). Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. *Food Chem. Toxicol. 45*, 2533-2562.

Kuhn, T. S. (1970). The structure of scientific revolutions. In O. Neurath, R. Carnap and C. Morris (eds.). *Foundations of the unity of science. Toward an International encyclopedia of unified science*. (53-272). Vol. 2. Chicago: Chicago University Press.

Leist, M., Hartung, T., Nicotera, P. (2008). The dawning of a new age of toxicology. *ALTEX 25*, (this issue).

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

Neugebauer, E. and Holaday, J. W. (1993). *Handbook of mediators of septic shock*. 1st edition. Boca Raton, FL: CRC-Press.

Mayer, D. (2004). *Essential evidence-based medicine*. Cambridge: Cambridge University Press.

NRC, Committee on developmental toxicology, Board on environmental studies and toxicology, National Research Council (2000). Recent advances in developmental biology. In NRC, *Scientific Frontiers in Developmental Toxicology and Risk Assessment* (127, Tab. 6-2). Washington, DC: The National Academies Press.

NRC, Committee on Toxicity Testing and Assessment of Environmental Agents, National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: The National Academies Press; http://www.nap.edu/catalog.php?record_id=11970

Sexton, K., Reiter, L. W. and Zenick, H. (1995). Research to strengthen the scientific basis for health assessmnts: a survey of the context and rationale for mechanistically models. *Toxicol. 102*, 3-20.

Stumann, T. C. and Bremer, S. (2008). The possible impact of human embryonic stem cells on safety pharmacological and toxicological assessments in drug discovery and drug development. *Curr. Stem Cell Res. Therap. 3*, 117-130.

Acknowledgements

Correspondence to

Prof. Dr. Dr. Thomas Hartung
EU Joint Research Centre
Institute for Health and Consumer Protection
European Centre for the Validation of Alternative Methods (ECVAM)
21020 Ispra
Italy
e-mail: thomas.hartung@ec.europa.eu