The Dawning of a New Age of Toxicology

Marcel Leist[1,2], Thomas Hartung[2] and Pierluigi Nicotera[3]
[1]Doerenkamp-Zbinden Chair for alternative in vitro methods, University of Konstanz, Germany; [2]ECVAM/ IHCP, EC-Joint Research Centre, Ispra, Italy; [3]MRC Toxicology Unit, University of Leicester, Leicester, U.K.

Summary
Toxicology faces enormous challenges in a world in which we are exposed to thousands of chemicals and millions of mixtures thereof. Radically new approaches to this problem need to be developed. A milestone in this direction is the vision of the US National Research Council (NRC) "Toxicity testing in the 21st century: a Vision and a Strategy". Currently, an alliance formed by the National Toxicology Program (NTP) and the Chemical Genomics Centre (NCGC) of the National Institutes of Health (NIH) and the Computational Toxicology Centre (NCCT) of the Environmental Protection Agency (EPA) is testing whether this new strategy can realistically form the basis of future public health decisions. The vision requires a radical paradigm shift in the approach to safety assessments, and turns the traditional procedures upside down. Where animal experiments used to be the most important technology, the future is seen in the strength of in vitro and in silico approaches based on human material. Today's toxicity testing starts with an initial black box screen on animals, sometimes followed by mechanistic studies, while the new vision approaches hazard assessment bottom-up. The procedure would begin with in vitro tests to define the affected pathways. To fill remaining gaps of knowledge, limited and targeted testing in animals would then be performed as a possible second step. This means nothing less than changing toxicology from being a predominantly observational craft and regulatory support discipline back to a natural science with all its dimensions. The background and the implications are discussed here in particular for a readership with interest also in parallel European trends.

Zusammenfassung: Toxikologie im Aufbruch
Die Toxikologie steht in einer Welt, in der wir Tausenden von Chemikalien und Millionen von deren Mischungen ausgesetzt sind, vor enormen Herausforderungen. Als Antwort darauf müssen radikal neue Ansätze entwickelt werden, um die Sicherheit der Bevölkerung zu gewährleisten. Ein Meilenstein in dieser Richtung ist die Vision des Nationalen Forschungsrates der USA „Toxizitätstestung im 21. Jahrhundert: Eine Vision und eine Strategie". Gegenwärtig testet eine Allianz, die zwischen dem NTP und dem NCGC des NIH sowie dem NCCT der EPA gebildet wurde, ob diese neue Strategie realistischerweise eine Basis für künftige Entscheidungen zum Schutz der öffentlichen Gesundheit sein kann. Die Vision setzt einen Paradigmenwechsel im Ansatz von Sicherheitsevaluationen voraus und stellt die traditionell angewandten Verfahren auf den Kopf. Wo bisher Tierexperimente die wichtigste Technologie waren, setzt die Zukunftsvision auf in vitro und in silico Ansätze, die auf menschlichem Material beruhen. Toxizitätstests beginnen heutzutage relativ blind mit einem Tierexperiment, dem dann nur manchmal mechanistische Studien folgen, während die neu vorgeschlagene Strategie die Sicherheitsevaluation von unten her beginnt: Zunächst würden mit in vitro Tests die durch Giftstoffe gestörten Stoffwechsel- und Regulationswege identifiziert; erst in einem möglichen zweiten Schritt kämen dann begrenzte und streng fokussierte Tierversuche dazu, um eventuell noch offene Wissenslücken zu schließen. Dies bedeutet nicht weniger als eine Umkrempelung der Toxikologie von einer hauptsächlich beschreibenden Tätigkeit und Hilfsdisziplin für Behörden wieder hin zu einer Naturwissenschaft mit all ihren Dimensionen. Der Hintergrund und die Konsequenzen werden hier insbesondere für Leser mit einem Interesse auch an parallelen Europäischen Trends beschrieben.*

*Die Abkürzungen sind in der englischen Zusammenfassung ausgeschrieben

Keywords: in vitro toxicology, REACH, toxicity pathways, environmental agents

1 Introduction

Toxicology is an exciting discipline that brings together specialists from vastly different areas. A picture (Fig. 1) that springs to mind is one of a body with three souls: As for many other medical disciplines, one important aspect of toxicology is that its procedures and the specific knowledge are applied like a craft. In this first domain, which contains the translational aspects of the science, careful documentation, process-optimisation and routine are of high importance. A second focus area is regulatory toxicology at the interface of industry and authorities, involved in setting and meeting guidelines and providing a basis for political decisions and legal requirements concerned with environmental health and consumer safety. The third soul of toxicology is its scientific basis. This area is concerned with the generation of new knowledge

Invited paper, received 20th March 2008, received in final form and accepted for publication 4th April 2008

and is linked to other natural sciences. It appears as if the three souls have lost connection over the past decades and that a large part of toxicology became frozen in time, using and accepting the same old animal models again and again, often without stringent examination of their validity (Hartung, 2008a,b). In this situation, the overall discipline is strongly driven by the demand for protocols and data for regulatory action. Only few resources remain for generation of fresh, fundamental toxicological knowledge and scientific output. A lot of the remaining scientific progress of toxicology depends strongly on import from other biomedical fields (Lotti and Nicotera, 2002). The consequences are reduced innovation, followed by a loss of attractiveness of the field for talented workers, and finally an inability to meet newly arising challenges.

Such new challenges are for instance the safety evaluation of compound mixtures, of biologics, of nanomaterials, of irradiated or genetically-altered food or of mobile phone radiation. None of them can be tackled adequately by classical animal-based methods. Huge challenges lie also in finding more predictive systems for developmental neurotoxicants (Grandjean and Landrigan, 2006) or non-genotoxic carcinogens (Ashby, 1996; Trosko and Upham, 2005; Williams and Whysner, 1996). However, these current problems are also a huge opportunity for the future, to bring the domains of toxicology together again, to link the field more closely to progress in other areas of biomedical sciences, and to give it a new basis (Fig. 1).

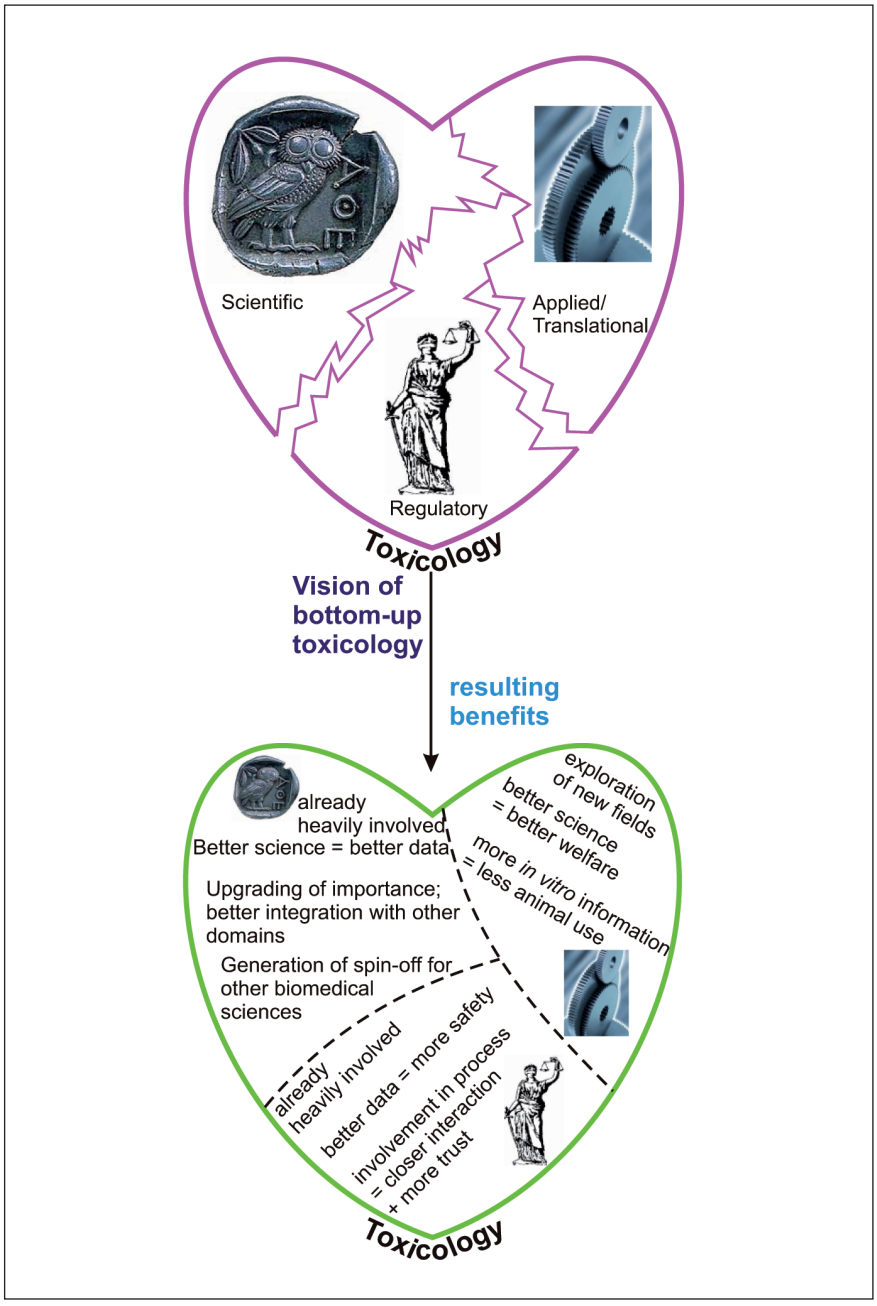

Fig. 1: Good prospects for heartbroken toxicology
Top: The three souls of toxicology (scientific, applied, regulatory) are not optimally connected and the discipline is suffering. Decidedly following the vision for toxicity testing in the 21st century, presented by the National Research Council (NRC, USA), toxicology can pull together the disciplines, strengthen each of them and put safety assessment on a new basis with less requirement for animal experimentation.

There is a vast body of evidence from mechanistic toxicology studies suggesting that the thousands of known noxious substances act by interfering with only a few (i.e. dozens) regulatory pathways of cells (NRC, 2007). For instance, a variety of hepatotoxins act by enhancing TNF-induced apoptosis (Leist et al., 1997), various compounds are neurotoxic because of perturbed cellular calcium metabolism (Nicotera, 1996; Orrenius et al., 2003; Leist and Nicotera, 1998), various immunotoxicants affect the cell cycle of lymphocytes via the Ah receptor (Kolluri, 1999), endocrine disrupters often bind to steroid receptors (Vedani et al., 2005, 2007; Waring and Harris, 2005), and interaction with the P450 system has been extensively examined as the basis of the toxicity of thousands of diverse compounds (Krebsfaenger et al., 2003; Nussler et al., 2001; Ioannides and Lewis, 2004). Information on such affected pathways can nowadays be obtained rapidly by high-throughput screening systems using human cells, and then be further analysed with modern methods of systems biology and bioinformatics.

Such a new approach has recently been suggested by the US National Toxicology Roadmap "A national toxicology program for the 21st century" (http://ntp.niehs.nih.gov/files/NTPrdmp.pdf) and the NRC (NRC, 2007), and testing of its feasibility by major safety authorities has begun (Collins et al., 2008).

2 A new vision of toxicity testing

The NRC, the most prestigious scientific council of the USA, was funded some years ago by the EPA and the NTP to develop a long range vision and implementation strategy for modern toxicology (Fig. 2; see Box 1 for explanation of the abbreviations). The heart of the new vision of toxicity testing proposed by the NRC is the concept of "toxicity pathways" (Fig. 3). As shown in Figure 3a, the vision takes its starting point from the presumption that most toxicants will eventually act by interfering with pivotal cellular structures and regulatory pathways. This would result in a limited number of toxicity pathways (e.g. disturbed calcium regulation, triggering of apoptosis, cell cycle derangement, …). It is then further presumed that knowledge of these pathways and knowledge of the action of toxicants on these pathways would allow predictions of toxicity on the level of the whole organism. This is a simple concept, but with huge implications. The practical consequence for toxicity testing would be no less than a turn-around of the currently used process from top to bottom (Fig. 3b, Fig. 4). Currently, animal models are frequently used as black box system to identify problematic compounds. Only in few cases (e.g. for valuable compounds, or compounds leading to high human exposure) will toxicity data ever be followed up to understand why a compound is toxic and whether the effect is relevant to humans. The vision laid out by the NRC suggests a radical paradigm shift. The start of a safety evaluation would begin with the chemical properties of a compound and then proceed to the biological characterisation in multiple *in vitro* systems (Fig. 3). Bioinformatic procedures would transform this information into a hazard estimate. This procedure would prioritise a few compounds (e.g. unclear hazard estimate or biokinetic predictions and high exposure) for further animal testing, and be sufficient on its own to eliminate many compounds and mixtures.

This would be a revolutionary approach if it was actually applied in practice, but is the idea really new? There is a saying that "success and good news have many parents, uncles, godfathers…., once they are apparent to everybody, while failure is

Box 1: Glossary of terms and abbreviations

(Q)SAR: (quantitative) structure-activity relationship. A way to correlate chemical structural information with biological endpoints (e.g. receptor binding or toxicity).
ASAT: EU initiative on "assuring safety without animal testing"; http://www.asat-initiative.eu
CEFIC: European Chemical Industry Council
DG RESEARCH: Directorate General of the EU for research (http://ec.europa.eu/dgs/research/index_en.html). In national terms this would correspond to the Ministry for Research. It is the major funding body for the large EU framework programme research projects.
ECVAM: European Centre for the Validation of Alternative Methods (http://www.ecvam.jrc.it)
EPA: Environmental Protection Agency (of the USA)
EPAA: European Partnership (of the European Commission and industry organisations) for Alternative Approaches to Animal Testing
InViTech: the EU high-throughput-high content centre; http://bms.jrc.ec.europa.eu/projects/InViTech.htm
MEIC: In 1989, Björn Ekwall and the Scandinavian Society for Cell Toxicology organised the Multicentre Evaluation of In Vitro Cytotoxicity (MEIC). Fifty compounds were evaluated in dozens of cytotoxicity assays and the results were published in a series of papers in 1998 in ATLA.
MRC: Medical Research Council of the UK; runs own research institutes, e.g. MRC Toxicology Unit in Leicester
NCBI: National Center for Biotechnology Information, a division of the National Library of Medicine (NLM) at the NIH
NCCT: National Center for Computational Toxicology (of the USA)
NCGC: NIH Chemical Genomics Centre
NIH: National Institutes of Health (of the USA)
NRC: National Research Council (of the USA), the principal operating agency of the National Academies of Sciences of the USA, the National Academy of Engineering and the Institute of Medicine. The National Academy of Sciences is known by many as publisher of the Proceedings of the National Academy of Sciences, USA.
NTC: The Netherlands Toxicogenomics Centre; http://toxicogenomics.nl
NTP: National Toxicology Program (of the USA)
PubChem: PubChem provides information on the biological activities of small molecules. It is a component of the NIH's Molecular Libraries Roadmap Initiative.
PubMed: Biomedical literature database at the NCBI
qHTS: quantitative high throughput screening. This technology allows the testing of thousands to ten-thousands of compounds in a single experiment. This compound number is 1-2 orders of magnitude lower than what would be used in industrial drug discovery screens. However, the data output is relatively rich, as compounds are screened at about 10 different concentrations and the shape of the resultant response curves yields additional information.
REACH: European regulation (EC) No 1907/2006 on the Registration, Evaluation, Authorisation and Restriction of Chemicals, which entered into force on the 1st of June 2007.
ZEBET: Zentralstelle zur Erfassung und Bewertung von Ersatz- und Ergänzungsmethoden zum Tierversuch am BfR/ Centre for Documentation and Evaluation of Alternatives to Animal Experiments at the BfR (Federal Institute for Risk Assessment)

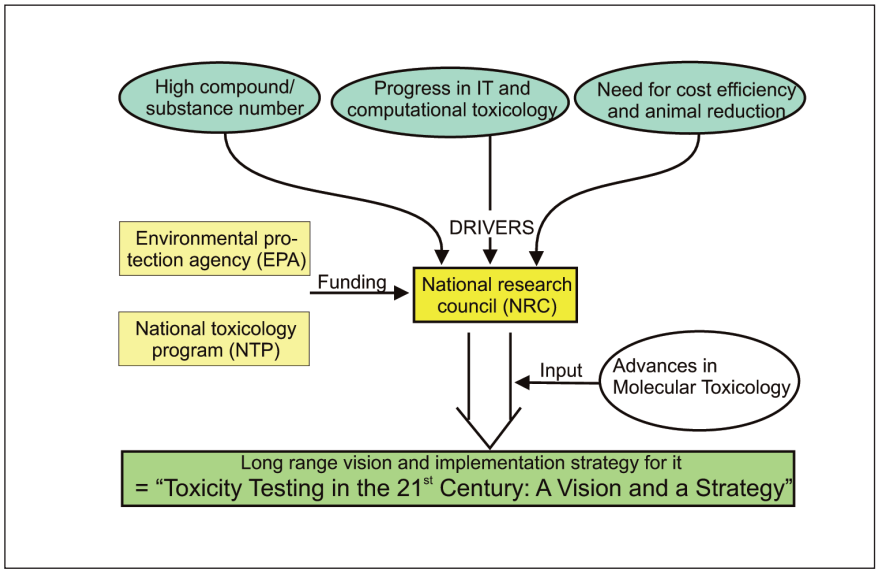

Fig. 2: Parents and godfathers of the vision
At the end of 2007, the NRC published its report after the initial trigger by two important regulatory agencies a couple of years earlier. A pivotal strength of the procedure, compared to similar approaches, is the early involvement of and support by major stakeholders (academia, regulators, industry) and the coupling of a vision to an implementation strategy.

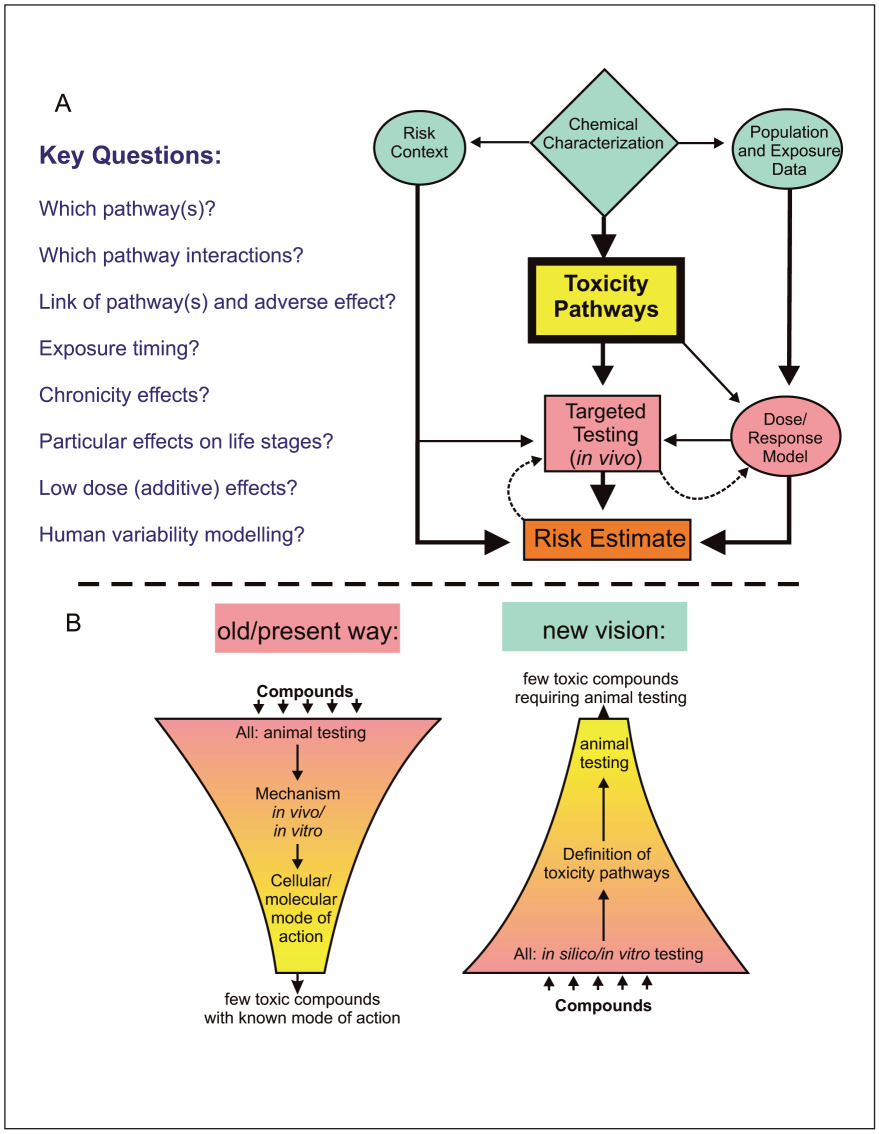

Fig. 3: Approach to toxicity testing suggested by the NRC (USA)
A. Toxicity pathways lie at the heart of the approach of hazard evaluation and are examined with the help of *in vitro* models. Gaps of knowledge and uncertainties are addressed by targeted animal testing. Risk estimates are then based on the hazard evaluation, exposure data and the risk context. For evaluation of this approach, a number of important questions need to be addressed. B. The new vision follows a bottom-up approach in contrast to the present approach.

an orphan, with an ugly mother-in-law, at best". Accordingly, many will claim now, that they have worked on the same idea as promoted by the NRC (NRC, 2007) for years, or even decades. It is indeed true that *in vitro* toxicology is a firmly established and well-organised discipline which has produced similar ideas and also already some applications in the regulatory field and in applied research (Andersen et al., 2005; Hendriksen, 2006; Gruber and Hartung, 2004; Hartung, 2001; Seiler et al., 2006; Whitlow et al., 2007). There has been a continuously good output over decades from laboratories interested in mechanistic toxicology,

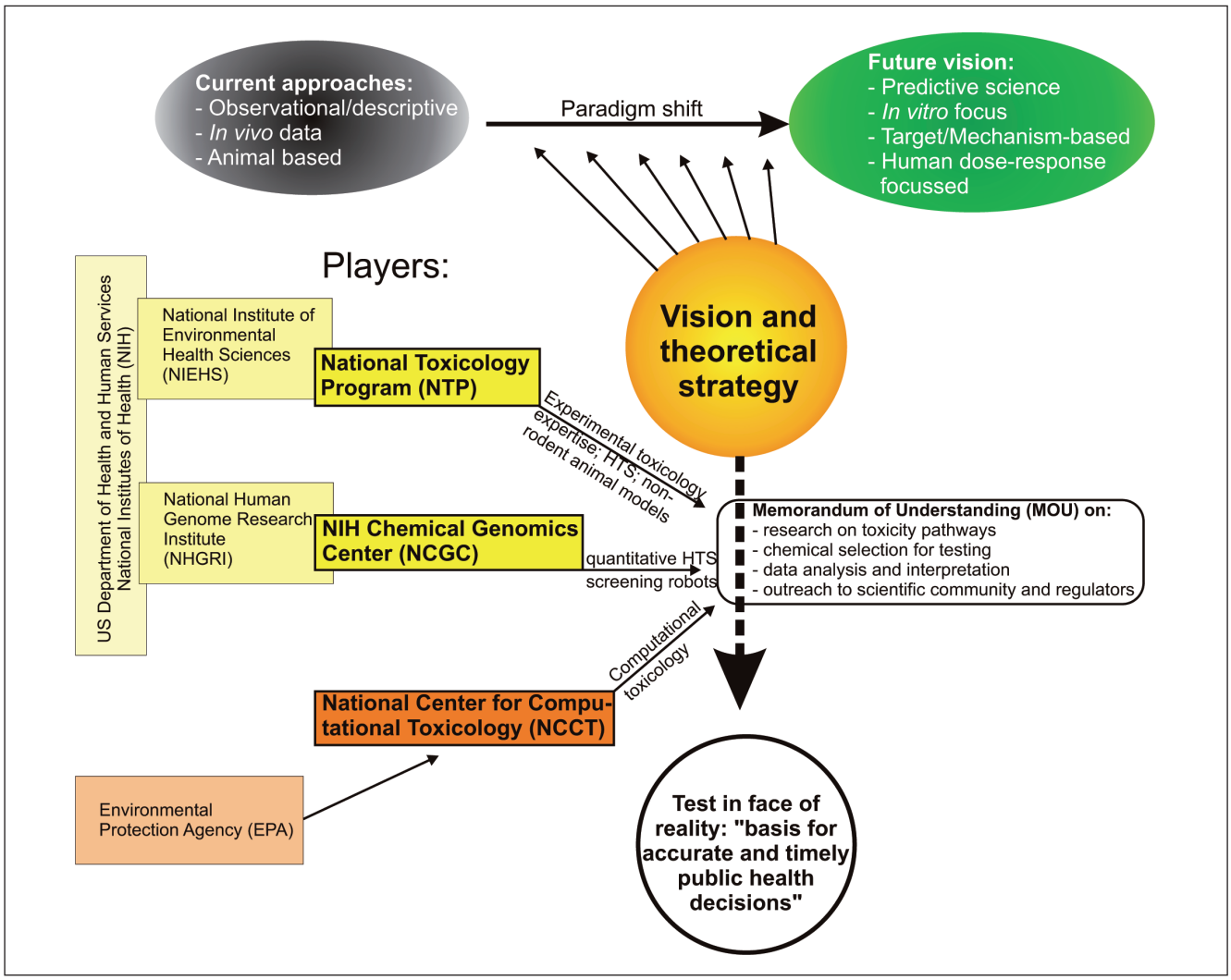

Fig. 4: Testing the feasibility of a new way of toxicity testing and reduction to practice
The vision and theoretical strategy were laid down by the NRC. Top: the paradigm shift according to this vision is outlined. Centre + bottom: In order to test whether the vision holds in the face of reality three major players agreed in a memorandum of understanding on a common strategy. The three players are institutes and programs of the EPA and the NIH, and contribute expertise as indicated.

and many companies and regulatory toxicologists are deeply involved in the development of alternative methods, such as *in silico* and *in vitro* screens. For instance, Gerhard Zbinden showed already 20 years ago the trend towards mechanistic models and the necessity for international regulators to follow this line and incorporate the ideas into the regulatory context (Zbinden, 1988, 1990). So again: what's new? It is the way it is done. The determination to "think big", the broad basis, the wide scope, the involvement of many stakeholders and drive by major authorities, the generation of open interfaces to the interested public (including accessible data bases) and the coupling of the vision to an implementation strategy that is robust enough to have a chance for success.

3 Testing the vision

A condensed overview of the initial phases of the implementation strategy was given recently by the involved US authorities (Collins et al., 2008). Here, we want to outline the essential features (Fig. 4), mainly as stimulation for the interested European readers and to provide a basis for potential interactions.

Presently, the implementation strategy is being explored by three major players on the basis of a memorandum of understanding clarifying the roles and duties (Fig. 3). One of the contributing institutions is the NCCT (Kavlock et al., 2007) under the roof of the EPA. The two other players are funded by the NIH: The NCGC contributes its screening infrastructure (robots, compound management, high-throughput measurement devices) and performs quantitative high-throughput screens (qHTS). The final player is the NTP which contributes with classical toxicological expertise, non-rodent animal models (for instance

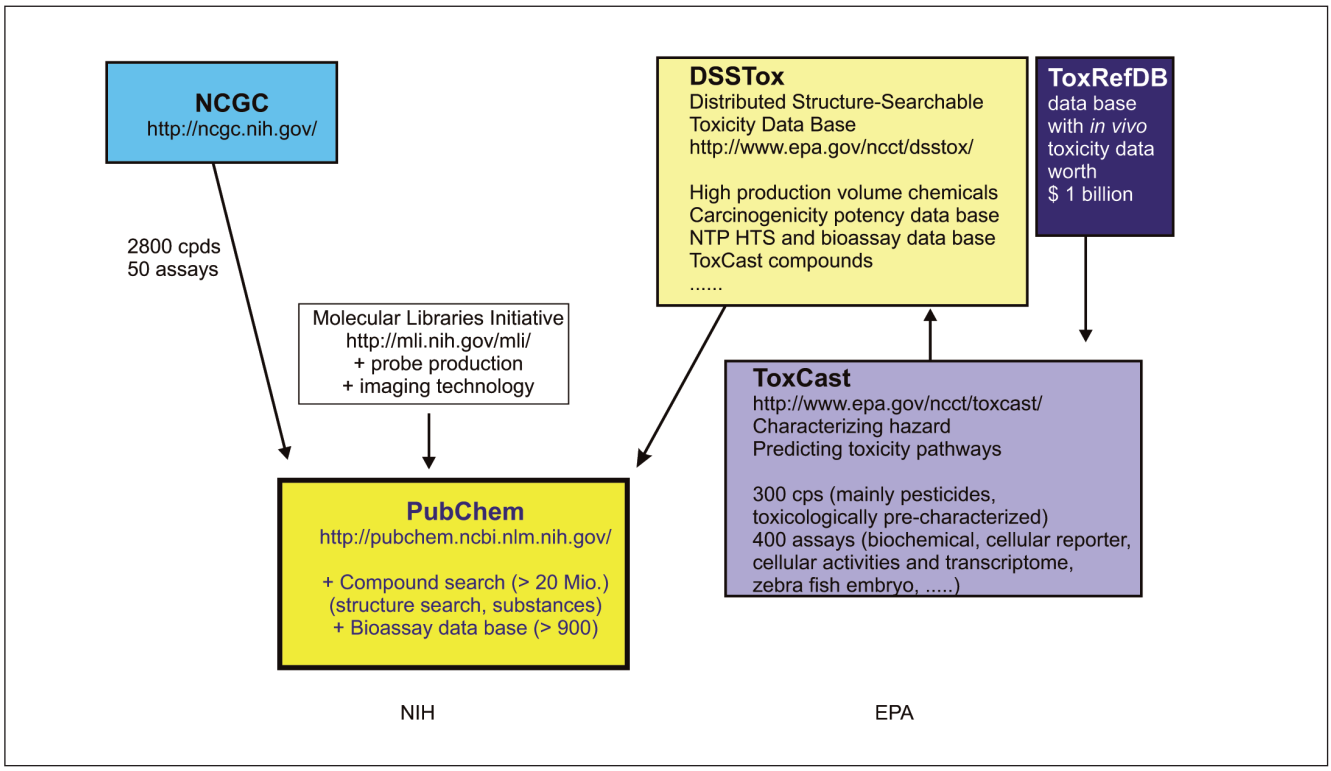

Fig. 5: Databases to help computational toxicology and *in vitro* toxicity testing
Different interlinked databases allow public access to cpd (compound) and assay information. RefToxDB is presently not publicly accessible. Links to the other databases are indicated.

zebra fish embryos) and especially a screen programme for about 300 selected compounds run through hundreds of assays (Fig. 4).

4 Steps toward a new toxicology

What happens with the data obtained? Here the idea of open public interfaces and generally-accessible databases comes into play. This sounds like a relatively trivial issue, but it should by no means be underestimated. We all have witnessed how the free internet availability of literature references via NCBI's PubMed has revolutionised the way scientific information is retrieved, and how Google has entirely changed the way general information is retrieved.

Toxicology urgently needs a parallel effort. At present a number of interconnected databases is being developed (Fig. 5) and expanded, but their user-friendliness is far from perfect. Much of the screening data will eventually end up in PubChem, which already harbours over 900 bioassays and will be fed directly with data from screens of the NCGC. Some of this data will also appear in classical journal publications, but in order to understand such publications one will have to be able to retrieve information from PubChem. An example is a publication describing the test of the cytotoxicity of about 1,400 compounds on 13 different cell lines (Xia et al., 2008a). The publication compiles data from different screens and extracts information from comparisons of cell lines and compounds. However, the compounds themselves and the original data from the screens will have to be extracted from the database (Xia et al., 2008b) – and, conversely, the database information may eventually be used again for new analyses and journal publications.

DSSTox is another database with generally richer data sets than PubChem. Here, reviewed and quality-controlled classical toxicological information is added to the compounds. The DSSTox website provides a public forum for publishing downloadable, structure-searchable, standardised chemical structure files associated with toxicity data (Houck et al., 2008).

One important input for DSSTox is ToxCast (Dix et al., 2007). This programme was designed to predict toxicity pathways and to characterise the hazard of a relatively small learning set of tool compounds (n=300) run through 400 different assays. ToxCast™ signatures will be evaluated by their ability to predict outcomes from existing mammalian toxicity testing and to identify toxicity pathways that are relevant to human health effects. High added value will be generated when this is linked to ToxRefDB, a database designed to contain data from huge historical animal testing efforts, including compounds selected for ToxCast. ToxRefDB is integrated into a more comprehensive data management system developed by NCCT called ACToR (Aggregated Computational Toxicology Resource) that manages the large-scale datasets of ToxCast™ .

The above databases are mainly compound and assay focussed. For hazard assessment further dimensions are essential. We need to understand how the human body handles a given chemical, what the important toxicity pathways are, and how we deal with human genetic variability (Fig. 6).

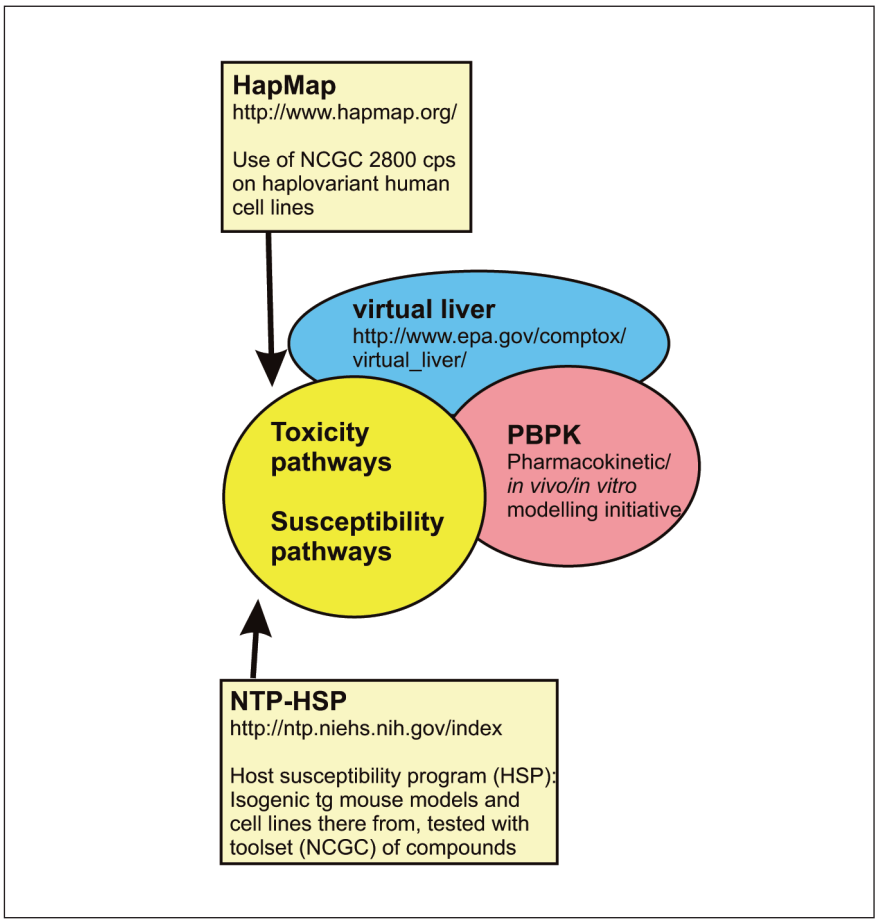

Fig. 6: Identification of toxicity pathways and *in vitro-in vivo* extrapolation
To support the overall project to test a new vision for toxicity testing, compound information alone is not sufficient. An important accessory programme is the initiative to model pharmacokinetics and *in vitro-in vivo* extrapolations and dose-response relationships with the help of physiology-based pharmacokinetic modelling (PBPK). Another initiative makes use of the HapMap project. This is a multi-country effort to identify and catalogue genetic similarities and differences in human beings. Using this information, researchers will be able to find genes that affect health and individual responses to environmental factors. The tool library used by the NCGC (about 2,800 compounds (cps) will be screened on cell lines with known haplotypes (i.e. known genetic variation). These compounds will also be tested within the HSP on transgenic (tg) mice and cells derived from them. This is expected to yield information on which genes have a major impact on adverse effects of environmental agents. The biological approach is complemented by the Virtual Liver Project, which plans to develop a database and algorithms able to predict liver toxicity and forms of liver carcinogenicity.

An *in vitro* test strategy requires more than the test system and data analysis. It cannot function without a prediction model to make use of the data. This also applies to complex integrated test strategies, and here pharmacokinetic information and dose-response modelling become highly important issues for the construction of prediction models. During establishment of the test strategy, variations of the following problem are frequently encountered: "pesticide X induces signs of toxicity (e.g. muscle paralysis) at a dose of Y mg/kg. Which concentrations should induce a positive readout in a corresponding *in vitro* toxicity test system in order to consider the test system relevant. In other words, which *in vitro* cytotoxic concentration would one predict from the *in vivo* data? Which would be a biologically relevant prediction model for *in vitro* concentrations, when *in vivo* doses are given?" Databases that translate such information are urgently required. During the application of an established test strategy to unknown compounds, a related problem occurs: "compound A triggers toxicity *in vitro* at concentrations higher than B micromolar. How much of the compound can be ingested safely?" PBPK databases will need to contain all the essential data on metabolism, protein binding and barrier permeation of compounds, in addition to suitable algorithms that will allow at least rough conversions of *in vitro* concentrations to *in vivo* doses. The setup of these databases is still in a very early phase.

The HapMap project is attempting to map and understand human haplotypes (i.e. variants of a given gene that are found in different proportions of the population). This project can also be linked to toxicity testing strategies. Interesting information is expected from testing a set of 2,800 compounds on human cell-lines with known haplovariants. In a parallel approach taken by the host susceptibility program (HSP), compounds are compared through a large number of transgenic mouse models and derived cell lines. These two programmes can contribute to the clarification of toxicity pathways and susceptibility genes and their respective effects. A different approach is taken by the EPA with the VirtualLiver project, which attempts to model the most important target organ of toxicity in its interaction with compounds. On its website it is stated ambitiously that "….the 5-year plan for the Virtual Liver Project is to develop a knowledgebase for qualitatively describing species-specific toxicity pathways due to exposure to chemicals, and to develop a virtual liver tissue that lays the foundation for quantitatively predicting the risk of non-genotoxic neoplastic lesions due to activation of certain genetic regulatory elements (i.e., nuclear receptors and other transcription factors) in humans".

5 The precautionary principle

Toxicological studies are designed to provide a basis for consumer protection by identifying hazardous compounds. The test systems will necessarily also produce false positives (compounds that are not hazardous to humans, but look hazardous in the

test system) and false negatives (compounds that are hazardous to humans, but are not correctly identified by the test system) (Leist et al., 2008). The latter class has been of particular concern. Therefore, the test systems and prediction models were tuned in a way to minimise this class as far as possible at cost of a largely increased class of false positives. This tuning of toxicity testing is called the precautionary principle and is one of the corner stones of toxicological thinking. Major changes in toxicity testing will always provoke fears in the public, in regulatory authorities and in other stakeholders that the precautionary principle may be violated. Therefore one of the major tasks of the implementation strategy of a new vision is to address these worries and to generate confidence that the safety level will not be compromised.

A first important issue to be considered is the understanding of the concept of "applicability domains". All toxicological methods are not generally applicable, but have applicability domains, i.e. limitations as to for which part of the chemical universe their predictive value has been shown. For instance, "drugs" or "pesticides" are typical applicability domains. Test guidelines, legislation, authorities, and the questions asked are vastly different in these areas. Other applicability domains would be industrial chemicals, cosmetics, biologics, and food additives. The concept was taken from the field of (Q)SAR and translated to test methods first in ECVAM's Modular Approach (Hartung et al., 2004). The vision discussed here applies mainly to the domain of environmental agents (i.e. pesticides or chemicals with relevant human exposure, for instance through the food chain). This is also reflected by different risk context scenarios that are explored and that are an important feature of the implementation strategy. Whether it can be translated to other domains without compromising the precautionary principle is one of the open questions for the future, and will certainly involve additional stakeholders.

The key issue to consider is, what new methods of toxicity testing should be used for comparison? Can we expect a 100% failsafe method? We know that present animal-based testing does not guarantee absolute safety (Zbinden, 1991). This is an obvious fact that is often forgotten in discussions on new approaches. New alternative tests are validated stringently (Hartung, 2007a), while many animal tests have never been formally validated (Hartung, 2008a,b). Even studies that address the question whether animal studies are of any toxicological use at all with respect to human safety are extremely scarce (Mathews, 2008). At least some doubt comes from the extreme variation of results when one and the same compound is used in different animal studies, and from the partially poor correlations between one species and the next, for instance between mouse and rat (Hartung, 2008a). Thus, a fair and honest approach to alternative testing strategies would imply that one does not require a 100% safety level, but rather a safety level that is in the range of (or at least as good as) that of standard animal experimentation. This also implies that showing the one or other insufficiency of *in vitro* approaches, and of the cell culture technology in particular (Hartung, 2007b), does not invalidate the usefulness of a technology. The strengths and weaknesses of animal and non-animal test approaches will just lie in different areas. Only looking at the comparison of the overall performance with regards to human safety will allow a reasonable judgement of the value.

In this context it is important to reconsider what the ultimate aim of the precautionary principle is: human safety. Sometimes, more exact knowledge on toxicity does not contribute to higher safety, but precautionary measures, e.g. regarding the transport of chemicals, take this function. Extensive animal testing will often generate redundant information, and, in addition, we accumulate more false-positive results (Bremer et al. 2007). To trigger a certain and adequate set of measures, sometimes limited *in vitro* and *in silico* information may be sufficient (Rogers et al., 2003).

6 The European side

The 3R principle (reduce, replace, refine), which already envisaged a combination of *in vitro* and *in vivo* approaches in the 1950's was originally developed in Europe (Russell and Burch, 1959). Is European toxicology less visionary now? What could be learned from the NIH/EPA approach?

Europe has a different, more diversified, but also more fragmented political landscape and different countries have found their own ways. For instance, the MRC in the UK decided almost 10 years ago to restructure its entire central toxicology institute in Leicester. Already at that time the guiding principle was to promote research on bottom-up toxicology, taking its starting point from understanding toxicity pathways and common processes like apoptosis. In Germany, ZEBET, a federal institution, was established nearly 20 years ago to develop, test and validate alternative methods to animal experimentation, and has been a major driver in the design of the first OECD toxicity testing guidelines based on *in vitro* testing only (for phototoxicity) and skin corrosion.

On the EU level, the first major driver for a new vision of toxicity testing comes from a different applicability domain than in the US – from cosmetics products. Here, the vision was immediately reduced to practice by law. The 7th amendment of the Cosmetics Directive set a strict timeline, finally banning the use of cosmetics if their ingredients were tested on animals. The implementation strategy implies that industry will need to establish animal-free test methods or change the business model. This is an interesting test case for the whole world to follow. In order to guide the development of methods and to ensure their validity, the EU founded ECVAM in 1992, a research institute entirely devoted to the validation of alternative methods. Now corresponding agencies and institutes are also found in the USA, Japan and other countries (Bottini et al., 2007). ECVAM harbours also an important database, DB-Alm, which is a high-quality source for *in vitro* test protocols and alternative methods (DB-Alm, 2007).

At present, the major driver for a rethinking of toxicity testing in Europe, is the REACH legislation (REACH, 2006). Over the last two years a revolution of the concept of how safety of chemicals is evaluated took place in Europe in this context: While in the past a (tonnage-triggered) set of mainly animal tests had to be provided in a tick-box manner, now (for both existing and new chemicals) integrated testing strategies making use of all information opportunities must be ap-

plied. A group of more than 200 experts from regulatory bodies, European Commission and industry developed these strategies (http://ecb.jrc.it/reach/rip/) in REACH Implementation Project 3.3 under the coordination of CEFIC and ECVAM. New and existing approaches were combined in order to optimise information generation for REACH, making use also of *in vitro, in silico* and read-across data from similar compounds. This law is at the basis of an enormous effort to re-evaluate about 30,000 chemicals already marketed in the EU and generates major financial and logistic pressures in addition to the ethical problem of the requirement for millions or tens of millions of animals to fulfil the test requirements. Faced with this enormous challenge, industry and the European Commission formed a partnership in the form of the EPAA (EPAA, 2006), that is working on new visions and implementation strategies. In parallel, the Directorate General of Research (DG Research) is heavily funding research consortia within the sixth and seventh framework programme to develop new *in vitro* test systems and strategies.

The key feature of REACH in the context of new visions of toxicology is that it has been influenced by an important postulate of the European animal legislation from 1986 (Directive 609/86), which can be summarised as "when alternatives to animal experimentation are available, they must be used"; "more of these alternatives need to be developed". More precisely, article 7.2. states: *"An experiment shall not be performed if another scientifically satisfactory method of obtaining the result sought, not entailing the use of an animal, is reasonably and practicably available."* And in Article 23.1.: *"The Commission and Member States should encourage research into the development and validation of alternative techniques which could provide the same level of information as that obtained in experiments using animals but which involve fewer animals or which entail less painful procedures, and shall take such other steps as they consider appropriate to encourage research in this field."*

Article 1.1 of the REACH regulation reads: *"Aim and scope 1. The purpose of this regulation is to ensure a high level of protection of human health and the environment, including the* ***promotion of alternative methods for assessment of hazards*** *of substances, as well as the free circulation of substances on the internal market while enhancing competitiveness and innovation."* (our high-lighting). REACH is thus the first major legislation in the huge application domain of industrial chemicals that gives some space for "intelligent test strategies", read-across between different information domains, the use of validated alternative methods, and also the use of non-validated alternative methods at least in a preliminary hazard evaluation (Bremer et al. 2007; Combes et al., 2008; Grindon et al., 2008a,b).

Nevertheless, REACH will still require millions of animal experiments, and the free space given by legislation is still far away from the vision of toxicity testing in the 21st century laid out by the NRC. Whether this heavy animal testing effort will lead to a parallel increase of human safety with respect to chemicals already on the market has been doubted (Knight, 2007). Thus, a new movement is presently forming that focuses on a more stringent validation of animal models and promotes an evidence-based toxicology, in which the best given test strategy is used instead of stringent adherence to only historically-legitimated animal models (Hoffmann and Hartung, 2006; Guzelian et al., 2005; Hartung, 2008b). A multitude of bottom-up movements are emerging at present, which include for instance ASAT, the NTC, and InViTech, to name a few.

7 Tasks ahead

We have tried here to survey exciting new developments and movements. Proof-of-concept studies need to clearly demonstrate the predictive power gained from

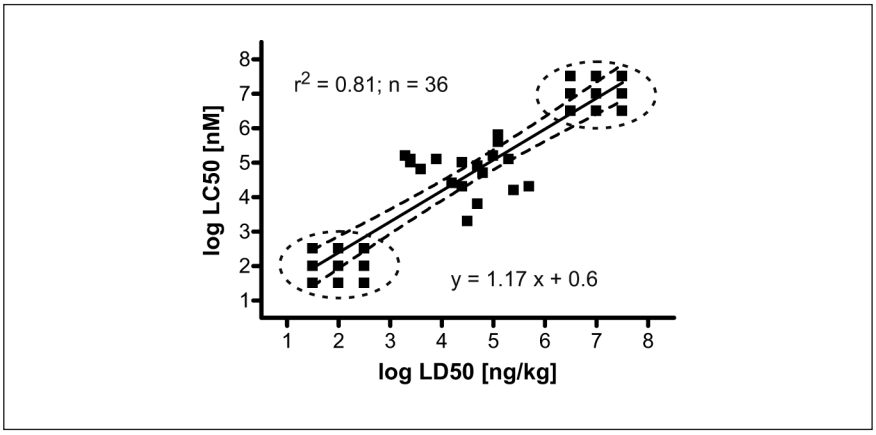

Fig. 7: Selection of libraries for *in vitro vs. in vivo* correlations
The graph shows an apparent correlation of *in vitro* toxicity data (half-maximal effective concentration = LC_{50} in nM) with acute toxicity animal data (half-maximal lethal dose = LD_{50} in ng/kg). The example shows the danger of obtaining meaningless correlations, when extreme positive and negative controls are chosen (as is common practice in the literature), and the number of samples in between is not relatively high (as is also frequently observed). A scientific approach with regards to library design, administration and validation is one essential prerequisite for new approaches to toxicology. Nine compounds were selected as "positive controls" (lower circle) with high toxicity, and nine compounds as "negative controls" with low toxicity (upper circle). For easier identification on the graph, controls were chosen with all combinations of toxicity of 2 ± 0.5 or 7 ± 0.5 log [nM or ng/kg], respectively, under the assumption that very toxic compounds *in vitro* will also be very toxic *in vivo,* and that compounds of very low toxicity will show that low effect in both systems. Further, 18 compounds were randomly assigned a toxicity on a scale of 3 - 6 (i.e. toxicities in the middle range varying by a factor of thousand, and with no correlation at all of *in vitro* and *in vivo* data) by Monte Carlo simulations. A representative example is shown. Under these conditions, "apparently" extremely good correlations are observed (r^2 = 0.8 and higher). Correlations remain still reasonably good (range of 0.52 - 0.61), when the number of compounds with random properties is doubled to 36.

these new approaches. More researchers need to be attracted to join the efforts, and regulatory authorities must show a willingness to embrace the new approaches as they gain scientific acceptance. The next few years should witness the early fruits of such efforts, but the paradigm shift will require a long-term investment and commitment to reach full potential. In a brief last paragraph we want to summarise critical issues to be addressed by the scientific community, granting agencies and authorities (Fig. 8):

Databases: These require a change of attitude as they move more into the centre of the process instead of being a final end product. It often appears from the lack of care and the limited analysis and accessibility options that they are more or less considered a tiresome duty to those who have generated the data. It is not sufficient to simply "dump" the data somewhere, even if they are flexibly retrievable and adequately quality-controlled. The science of visualisation of data and especially visualisation of large complex data spaces needs to be applied much more strongly here. The importance of this process and the need for users and developers to hold a constant dialogue during the design of analysis and visualisation algorithms are still heavily underestimated. Another important issue is the cross-linking of information. For instance a number of databases have been generated in Europe on *in vitro* acute toxicity data. For instance the MEIC data base (see Box 1) covers a very large number of toxicity assays, and the Halle Registry (Halle, 2003) over 300 *in vitro-in vivo* comparisons, but cross-linking is limited, as is the general and easy accessibility.

Libraries: To fill the databases with information, real compounds are required. Especially in the proof-of-principle phase of testing, the selection of these compound libraries plays an important role (Fig. 7) and contributes to the success, the validity and the general acceptance of validation efforts. Not only will the right "theoretical" composition of the libraries be of high importance, but also the physical composition and availability. Compound stability and purity, the general accessibility and continuous quality control are non-trivial issues, especially in the field of environmental chemicals, industrial chemicals and pesticides. Here one solution to be considered is chemical reference laboratories making defined library copies available to others. This has been conceptualised on the European level in form of CORRELATE (Correlate, 2007) and should also be considered as a great opportunity in the context of REACH (see Hartung, 2008b).

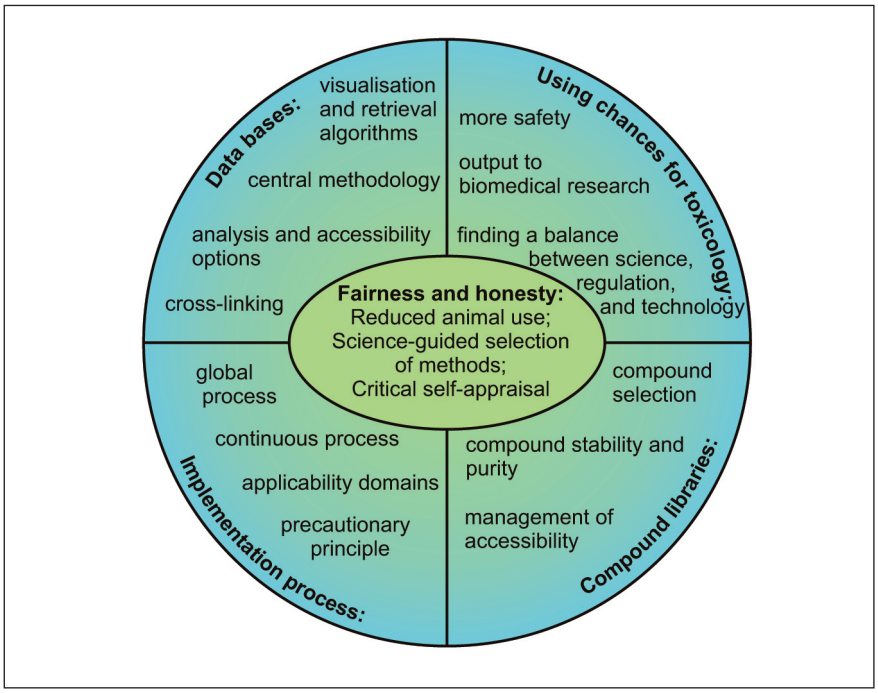

Fig. 8: Compilation of tasks ahead and associated key issues

Process: Many areas of basic biomedical research have experienced bumpy rides with periods of hype and disappointment. Toxicology has a continuous high responsibility for human safety and cannot, even transiently, simply drop the precautionary principle. However, it can ask critical questions on how it should best be applied in different situations, exposure scenarios and applicability domains. This provides a basis for a continuous, long-term effort to let toxicology evolve to a higher level than now. This process needs essentially to be global and involve all stakeholders (Bottini et al., 2007). Despite all enthusiasm, rapid success is not to be expected and all hype had rather be avoided as initial setbacks are likely to happen. This has to be accepted in the strategy. However, the determination to move on needs to be strong enough to attack problems with the right critical mass and impact right from the beginning and as they emerge.

Chances: The process of putting regulatory toxicology and the process of toxicity testing on a more mechanistic basis provides a chance for toxicology to evolve as a discipline, and also contribute general biomedical knowledge. This closes the circle started at the beginning of this article (Fig. 1). In the past, toxicology had the chance to promote the advance of biomedical sciences in general, for instance by discovering and driving the fields of apoptosis, toxinology or stress response. However, these opportunities were not seized, and other sciences drove these fields instead. Now, new chances are arising, possibly in the fields of systems biology, DNA repair or pathological aging. Possibly, also in the fields of chemical genetics and the introduction of chemical screens to non-pharmaceutical areas. To grasp any of these chances, it is important to dare to take the lead and not to lose touch with basic science. Application of HTS or qHTS as described above sounds fancy, but it is at the moment only a technology, not a science. This technology has brought a lot of disappointment in drug discovery, which one can learn from. It will be important in the future to avoid the mistakes of the past, and to incorporate the "technology" into a robust "scientific concept", which combines brain with the muscles.

Fairness and honesty: An unbiased approach, based on scientific evidence only, will be the best way to find solutions acceptable for all stakeholders. Presently one may wonder what the scientific basis for some animal experiments is. The lousy output and poor information from acute toxicity studies with lethality endpoint has been criticised for a long time (Tamborini et al., 1990; Zbinden, 1986; Paget, 1983; Zbinden and Flury-Roversi, 1981), and now, at least in the application domain of drugs, there seems to be a broad agreement that the assay could easily have been abolished (Robinson et al., 2007). Why hasn't this already happened? A similar situation can be found for two-generation studies for developmental toxicity testing, where the second generation apparently does not contribute with significant information (Janer et al., 2007). Here, non-scientific reasons seem to prevail, and the argument may be expanded to more examples of animal toxicity testing (Hartung, 2008a). It is also a sign of poor science that so little pharmacokinetic information is available from acute toxicity tests. This makes the present *in vivo-in vitro* comparisons very difficult and thus prevents a potential substitution of animal experiments by alternative methods. To be honest, the field of alternative methods also needs to look at obvious weaknesses of its own methods and establish itself as an academic discipline (Leist, 2006). Many assays are still just as much black box systems as animal experiments and pharmacokinetic information has been terribly neglected. If all sides focus on a vision of best science for best toxicology, then the sun will indeed rise on a new era.

References

Andersen, M. E., Dennison, J. E., Thomas, R. S. and Conolly, R. B. (2005). New directions in incidence-dose modeling. *Trends Biotechnol. 23(3)*, 122-127.

Ashby, J. (1996). Alternatives to the 2-species bioassay for the identification of potential human carcinogens. *Hum. Exp. Toxicol. 15*, 183-202.

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

Bremer, S., Pellizzer, C., Hoffmann, S. et al. (2007). The development of new concepts for assessing reproductive toxicity applicable to large scale toxicological programmes. *Curr. Pharm. Des. 13(29)*, 3047-3058.

Collins, F. S., Gray, G. M., Bucher, J. R. (2008). Toxicology. Transforming environmental health protection. *Science 319*, 906-907.

Combes, R., Grindon, C., Cronin, M. T. et al. (2008). Integrated decision-tree testing strategies for acute systemic toxicity and toxicokinetics with respect to the requirements of the EU REACH legislation. *ATLA 36(1)*, 45-63.

Correlate (2007). http://projects-2007.jrc.ec.europa.eu/show.gx?Object.object_id=PROJECTS0000000003008C51

DB-Alm (2007). Invittox protocol number 101. http://ecvam-dbalm.jrc.ec.europa.eu/public_view_doc2.cfm?id=6E7E72104B2DEFD6BE979B3B139176C67180BB0BC12CB10496CDA74B54630A05A3291B895581F634

Dix, D. J., Houck, K. A., Martin, M. T. et al. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Tox. Sci. 95*, 5-12.

EPAA (2006). European Partnership to Promote Alternative Approaches to Animal Testing http://ec.europa.eu/enterprise/epaa/conf_2006_presentation-vdgraaf_unilever.pdf

Grandjean, P., Landrigan, P. J. (2006). Developmental neurotoxicity of industrial chemicals. *Lancet 368*, 2167-2178.

Grindon, C., Combes, R., Cronin, M. T. et al. (2008a). An integrated decision-tree testing strategy for repeat dose toxicity with respect to the requirements of the EU REACH legislation. *ATLA 36(1)*, 93-101. PMID: 18333717

Grindon, C., Combes, R., Cronin, M. T. et al. (2008b). Integrated decision-tree testing strategies for developmental and reproductive toxicity with respect to the requirements of the EU REACH legislation. *ATLA 36(1)*, 65-80.

Gruber, F. P., Hartung, T. (2004). Alternatives to animal experimentation in basic research. *ALTEX 21 Suppl. 1*, 3-31.

Guzelian, P. S., Victoroff, M. S., Halmes, N. C. et al. (2005). Evidence-based toxicology: a comprehensive framework for causation. *Hum. Exp. Toxicol. 24*, 161-201.

Halle, W. (2003). The Registry of Cytotoxicity: toxicity testing in cell cultures to predict acute toxicity (LD_{50}) and to reduce testing in animals. *ATLA 31*, 89-198.

Hartung, T. (2001). Three Rs potential in the development and quality control of pharmaceuticals. *ALTEX 18 Suppl. 1*, 3-13.

Hartung, T., Bremer, S., Casati, S. (2004). A Modular Approach to the ECVAM Principles on Test Validity. *ATLA 32*, 467-472.

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

Hartung, T. (2007b). Food for thought on …..cell culture. *ALTEX 24*, 143-147.

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

Hartung, T. (2008b). Food for thought….. on the evolution of toxicology and phasing out of animal testing. *ALTEX 25*, (this issue, page 91).

Hendriksen, C. F. (2006). Towards eliminating the use of animals for regulatory required vaccine quality control. *ALTEX 23*, 187-190.

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

Houck, K., Dix, D., Judson, R. et al. (2008). DSSTox EPA ToxCast High Throughput Screening Testing Chemicals Structure-Index File: SDF File and Documentation, Updated version: TOXCST_v2b_320_08Feb2008, http://www.epa.gov/ncct/dsstox/sdf_toxcst.html

Ioannides, C., Lewis, D. F. (2004). Cytochromes P450 in the bioactivation of chemicals. *Curr. Top. Med. Chem. 4*, 1767-1788.

Janer, G., Hakkert, B. C., Slob, W. et al. (2007). A retrospective analysis of the two-generation study: what is the added value of the second generation? *Reprod. Toxicol. 24*, 97-102.

Kavlock, R. J., Ankley, G., Blancato, J. et al. (2007). Computational Toxicology – A State of the Science Mini Review. ToxSci. Advance Access published on December 7, 2007. doi:10.1093/toxsci/kfm297

Knight, A. (2007). Animal experiments scrutinised: systematic reviews demonstrate poor human clinical and toxicological utility. *ALTEX 24(4)*, 320-325.

Kolluri, S. K., Weiss, C., Koff, A., Göttlicher, M. (1999). p27(Kip1) induction and inhibition of proliferation by the

intracellular Ah receptor in developing thymus and hepatoma cells. *Genes Dev. 13*, 1742-1753.

Krebsfaenger, N., Mürdter, T. E., Zanger, U. M. (2003). V79 Chinese hamster cells genetically engineered for polymorphic cytochrome P450 2D6 and their predictive value for humans. *ALTEX 20*, 143-154.

Leist, M., Gantner, F., Naumann, H. et al. (1997). Tumor necrosis factor-induced apoptosis during the poisoning of mice with hepatotoxins. *Gastroenterology 112*, 923-934.

Leist, M, Nicotera, P. (1998). Calcium and neuronal death. *Rev. Physiol. Biochem. Pharmacol. 132*, 79-125.

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

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

Lotti, M., Nicotera, P. (2002). Toxicology: a risky business. *Nature 416*, 481.

Mathews, R. A. J. (2008). Medical progress depends on animal models - doesn't it? *J. R. Soc. Med. 101*, 95-98.

Nagel, R. (2002). DarT: The embryo test with the Zebrafish Danio rerio - a general model in ecotoxicology and toxicology. *ALTEX 19 Suppl. 1*, 38-48.

Nicotera, P. (1996). The Gerhard Zbinden Memorial Lecture. Alteration of cell signalling in chemical toxicity. *Arch. Toxicol. 18 Suppl.*, 3-11.

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

Nussler, A. K., Wang, A., Neuhaus, P. et al. (2001). The suitability of hepatocyte culture models to study various aspects of drug metabolism. *ALTEX 18*, 91-101.

Orrenius, S., Zhivotovsky, B., Nicotera, P. (2003). Regulation of cell death: the calcium-apoptosis link. *Nat. Rev. Mol. Cell Biol. 4*, 552-565.

Paget, E. (1983). The LD_{50} test. *Acta Pharmacol. Toxicol. (Copenh.) 52 Suppl. 2*, 6-19.

REACH (2006). REACH legislation under directive (EC) No 1907/2006, http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2006:396:SOM:en:HTML

Robinson, S., Delongeas, J. L., Donald, E. et al. (2007). A European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development. *Regulatory Toxicol. Pharmacol.* (in press). doi:10.1016/j.yrtph.2007.11.009

Rogers, M. D. (2003). Risk analysis under uncertainty, the precautionary principle, and the new EU chemicals strategy. *Regul. Toxicol. Pharmacol. 37*, 370-381.

Russell, W. M. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London: Methuen.

Seiler, A., Buesen, R., Hayess, K. et al. (2006). Current status of the embryonic stem cell test: the use of recent advances in the field of stem cell technology and gene expression analysis. *ALTEX 23 Suppl.*, 393-399.

Tamborini, P., Sigg, H., Zbinden, G. (1990). Acute toxicity testing in the nonlethal dose range: a new approach. *Regul. Toxicol. Pharmacol. 12*, 69-87.

Trosko, J. E., Upham, B. L. (2005). The emperor wears no clothes in the field of carcinogen risk assessment: ignored concepts in cancer risk assessment. *Mutagenesis 20,* 81-92.

Vedani, A., Dobler, M., Lill, M. A. (2005). Virtual test kits for predicting harmful effects triggered by drugs and chemicals mediated by specific proteins. *ALTEX 22,* 123-134.

Vedani, A., Dobler, M., Spreafico, M. et al. (2007). VirtualToxLab - in silico prediction of the toxic potential of drugs and environmental chemicals: evaluation status and internet access protocol. *ALTEX 24*, 153-161.

Waring, R. H., Harris, R. M. (2005). Endocrine disrupters: a human risk? *Mol. Cell Endocrinol. 244*, 2-9.

Whitlow, S., Bürgin, H., Clemann, N. (2007). The embryonic stem cell test for the early selection of pharmaceutical compounds. *ALTEX 24*, 3-7.

Williams, G. M., Whysner, J. (1996). Epigenetic carcinogens: evaluation and risk assessment. *Exp. Toxicol. Pathol. 48*, 189-195.

Xia, M., Huang, R., Witt, K. L. et al. (2008a). Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ. Health Perspect. 116*, 284-291.

Xia, M. et al. (2008b). The 1408 compounds of Environ Health Perspect 116, 284. http://www.epa.gov/ncct/dsstox/sdf_ntphts.html#DownloadTable

Zbinden, G., Flury-Roversi, M. (1981). Significance of the LD50-test for the toxicological evaluation of chemical substances. *Arch. Toxicol. 47,* 77-99.

Zbinden, G. (1986). Invited contribution: acute toxicity testing, public responsibility and scientific challenges. *Cell Biol. Toxicol. 2*, 325-335.

Zbinden, G. (1988). Reduction and replacement of laboratory animals in toxicological testing and research. Interim report 1984-1987. *Biomed. Environ. Sci. 1*, 90-100.

Zbinden, G. (1990). Alternatives to animal experimentation: developing in-vitro methods and changing legislation. *Trends Pharmacol. Sci. 11*, 104-107.

Zbinden, G. (1991). Predictive value of animal studies in toxicology. *Regul. Toxicol. Pharmacol. 14*, 167-177.

Acknowledgement

We gratefully acknowledge the valuable input and proofreading of Suzanne Kadereit and Stefan Schildknecht, the secretarial help by Brigitte Schanze and grant support by the Doerenkamp-Zbinden Foundation, the State of Baden-Württemberg and the European Union (ESNATS).

Correspondence to:

Prof. Dr. Marcel Leist
Doerenkamp-Zbinden Chair for alternative in vitro methods
Faculty of Natural Sciences and Mathematics
University of Konstanz
PO Box M657
78457 Konstanz
Germany
e-mail: marcel.leist@uni-konstanz.de
Tel.: +49-7531-885037
Fax: +49-7531-885039