European Journal of Pharmaceutics and Biopharmaceutics 77 (2011) 338–349

Contents lists available at ScienceDirect

European Journal of Pharmaceutics and Biopharmaceutics

journal homepage: www.elsevier.com/locate/ejpb

Review article

# From alternative methods to a new toxicology

Thomas Hartung *

*Johns Hopkins University, Dept. Environmental Health Sciences, Baltimore, USA*
*Pharmacology and Toxicology, University of Konstanz, Konstanz, Germany*

ARTICLE INFO

*Article history:*
Available online 30 December 2010

*Keywords:*
Alternative methods
Animal testing
Toxicology
Validation
Evidence-based medicine
Omics technologies

ABSTRACT

Mechanistic toxicology has evolved by relying, to a large extent, on methodologies that substitute or complement traditional animal tests. The biotechnology and informatics revolutions of the last decades have made such technologies broadly available and useful, but regulatory toxicology has been slow to embrace these new approaches. Major validation efforts, however, have delivered the evidence that new approaches do not lower safety standards and can be integrated into regulatory safety assessments.

Particularly in the EU, political pressures, such as the REACH legislation and the 7th Amendment to the cosmetic legislation, have prompted the need of new approaches. In the US, the NRC vision report calling for a toxicology for the 21st century (and its most recent adaptation by EPA for their toxicity testing strategy) have initiated a debate about how to create a novel approach based on human cell cultures, lower species, high-throughput testing, and modeling.

Lessons learned from the development, validation, and acceptance of alternative methods support the creation of a new approach based on identified toxicity pathways. Conceptual steering and an objective assessment of current practices by evidence-based toxicology (EBT) are required. EBT is modeled on evidence-based medicine, which has demonstrated that rigorous systematic reviews of current practices and meta-analyses of studies provide powerful tools to provide health care professionals and patients with the current best scientific evidence. Similarly, a portal for high-quality reviews of toxicological approaches and tools for the quantitative meta-analyses of data promise to serve as door opener for a new regulatory toxicology.



## 1. Introduction

Over the last three decades, the development of alternative approaches based on the 3Rs principles – Reduce, Replace, Refine, as suggested in 1959 by Bill Russel and Rex Burch [1] – was very much driven by animal welfare considerations. Replacement of animal tests has been heavily promoted, especially in Europe. Major validation efforts have delivered the evidence that new approaches do not lower safety standards and can be integrated into regulatory safety assessments [2]; the inclusion of several alternative methods in OECD and ICH (International Conference on Harmonization) consensus guidelines is proof of this. More recently, in the US, the vision for "toxicology for the 21st century" (Tox-21c) [3], prompted by the National Research Council report of the same name only 3 years ago, has dominated the discussion. Europe and the US have pursued the development of new toxicological tools in very different ways [2]. The NAS/NRC Tox-21c report calls for a paradigm shift in toxicology [4–8]. In February 2008, several American agencies, recently joined by the FDA, announced a coalition to facilitate its implementation [9]: "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, Francis Collins, now Director of the National Institutes of Health, stated: "[Toxicity testing] was expensive, time-consuming, used animals in large numbers, and didn't always work." In the same article, Elias Zerhouni, then Director of NIH, said: "Animal testing won't disappear overnight, but the agencies' work signals the beginning of the end." Only three years after its publication, we have seen numerous conferences and symposia addressing the report and its implementation, the formation of an alliance of US agencies, and the development of a new EPA toxicity testing strategy [10,11]. Depending on the proponent, more or less emphasis is given to technological updates, throughput of testing, costs, replacement of animal testing, or quality of toxicological assessments. There is no doubt that all aspects synergize to bring about a potentially revolutionary change [12].

Significant political support exists on both sides of the Atlantic. In Europe, the horizontal animal welfare legislation from 1986

* Johns Hopkins Bloomberg School of Public Health, Dept. Environmental Health Sciences, Center for Alternatives to Animal Testing, 615 N. Wolfe St., Baltimore, MD 21205, USA.
*E-mail address:* THartung@jhsph.edu

0939-6411/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejpb.2010.12.027

revised in 2010, which urges the use of 3Rs methods wherever possible, and cosmetics and chemical legislation are the primary drivers. In the US, it is mainly federal agencies, most prominently the Environmental Protection Agency (EPA), that made the implementation of the NRC report their toxicity testing strategy in 2009, preempting such legislative measures as the reauthorization of the Toxic Substances Control Act (TSCA) (the US chemicals legislation).

European statistics [13] show that regulatory toxicity testing comprises about 10% of all animal use and vaccine testing represents about 15% (which is likely less in other parts of the world because of the strong role of European manufacturers). The major consumer, however, is medical product development and quality control, at 45%, and fundamental biological research, at 30%. The latter, in particular, follows completely different drivers for animal use than industry. Other areas, such as education or diagnosis of disease, are minor, with less than 2% each. Our analysis [14] suggests that these figures can most probably be translated to other industrialized countries.

Tox-21c addresses regulatory use of animals. If it is ultimately successful in phasing out all animal use, the result would be 10% fewer animals used in regulatory toxicity testing (mainly rats, mice, fish, and rabbits). How impressive this figure is depends on personal judgment [15], but fortunately, this is not the only motivation for Tox-21c, which is ultimately a revision of the toxicological toolbox that promises to improve the quality of predictions and increase throughput [16,17]. Costs are not really an issue: we have shown that regulatory toxicology represents roughly 0.03% of the turnover of and is not really a challenge to the regulated industry [14]. However, industry and consequently society are suffering from the limited quality of predictions. This holds true not only for the missed side effects of substances but also the "precautionary" elimination of valuable substances [18,14]. The low throughput of testing is a threat to society, while lack of knowledge or "toxicological ignorance" is enormous. We have only recently shown how a bottleneck for testing is created by applying traditional approaches analyzing the European REACH legislation [19,20] – and that such a program could never be executed without changing approaches.

The number of new substances and products entering our lives is growing, and the pace of their introduction is continuously accelerating, but the tools to assess their safety are barely increasing, if not slowing. Some safety concerns that have arisen are endocrine disruption (mainly in view of reduced male fertility), immunogenicity (i.e., antibody formation against biological drugs), immunotoxicity, respiratory hazards (possibly linked to childhood asthma), neurodevelopmental toxicities (possibly contributing to autism and attention deficit syndromes), etc. There is a growing gap between assessment capacities and assessment needs [19,20]. We have recently made a model calculation (using the example of REACH) which illustrates that forcing industry to apply traditional approaches to a fraction of substances results in efforts beyond what can be reasonably executed.

These data illustrate current challenges to the science of safety assessments. The central thesis of the article is that a paradigm shift in toxicology is necessary and possible. There needs to be a strategy developed in the near future that defines the metrics against which the new testing paradigm will be evaluated and develops the methods for objective assessment of the process. However, it appears very early to propose one at this stage.

## 2. The drivers of change for a paradigm shift in toxicology

According to Thomas S. Kuhn (1922–1996) [21], science develops in waves, and it seems that regulatory toxicology is undergoing a dramatic change. Kuhn describes scientific revolutions as "the tradition-shattering complements to the tradition-bound activity of normal science," pointing out that a shift in professional commitments to shared assumptions takes place when an anomaly "subverts the existing tradition of scientific practice." Several such anomalies are currently emerging at the same time, and the author has argued elsewhere [12] that this meets the characteristics of Kuhn's revolutionary change paradigm.

### 2.1. Legislation

It should be noted that politics is driving the change in toxicological science. In some instances, new methods have been added, but the core of the toolbox was never touched and there is remarkable resistance to change – it is probably the only area in the life sciences where crucial approaches have not changed for 40–80 years. These approaches are apparently difficult to revise because they represent a belief system rather than a scientific approach with self-critical renewal systems. This has several reasons:

- The concept of risk, including what is *acceptable* risk, is more a societal construct than a science, especially since, in most cases, we do not have objective measures. We typically extrapolate hazards from animal test data, for which we do not know how it translates to humans and have to combine this with extremely limited exposure information.
- The majority of toxicological results are typically not published. They are generally negative data of no harm, considered proprietary, not interesting to the broader scientific community, or possibly dangerous if misinterpreted by a lay audience.
- There is hardly any retesting, which would control earlier results; this is even frequently prohibited by legislation.

This has led to the strong beliefs among regulators (and most of the regulated community) in the standard tools, which appear to be 'useful from experience'. Petr Skrabanek and James McCormick masterfully phrased it: "Learning from experience may be nothing more than learning to make the same mistakes with increasing confidence." [22]. In many instances, it takes the combination of lay press, consumer groups, environmentalists, and animal welfare activists to convince politicians to impose change on science. For the paradigm shift in toxicology, it is less the animal welfare legislation, which impacts, but the regulations of different product categories in different parts of the world.

#### 2.1.1. Animal welfare legislation

Animal welfare legislation is perceived by many in the field as an obstacle to what they need to do (and are going to do anyway – less than 1% of the proposals are rejected by animal use committees). The vigorous discussion during the development of the new European laboratory animal welfare legislation [25] may serve as an example here. EUROHORC, the organization of European Heads of Research Councils, objected strongly to limits on research, such as restriction of use of non-human primates, limits by severity levels, restrictions on reuse of animals, extension of scope of Directive to cover invertebrates and larval forms, and increased requirements for care and accommodation [23]. Compassion for the animals is, of necessity, sublimated to endure the work. It seems to be a classical case of cognitive dissonance, i.e., the uncomfortable feeling caused by holding contradictory ideas simultaneously. This theory, taken from social psychology, suggests that people have a motivational drive to reduce dissonance by changing their attitudes, beliefs, and actions [24]. We cannot do things we dislike over long periods of time without suffering, so we adjust our attitudes. Most people carrying out animal experiments have had these experiences, and the natural inner resistance either dissipates or the person drops out of this field of work.

It is very interesting to see that the legal provisions are very different on both sides of the Atlantic. In the US, the Animal Welfare Act from 1966 (with its six amendments) explicitly excludes mice, rats, and birds and does not cover cold-blooded animals such as fish:

> §2132. Definitions
> (g) The term "animal" means any live or dead dog, cat, monkey (non-human primate mammal), guinea pig, hamster, rabbit, or such other warm-blooded animal, as the Secretary may determine is being used, or is intended for use, for research, testing, experimentation, or exhibition purposes, or as a pet; but such term excludes (1) birds, rats of the genus Rattus, and mice of the genus Mus, bred for use in research...

Thus, approximately 90% of experimental animals are excluded from legislation. It is only indirectly, via NIH funding guidelines, that most institutions are forced to comply with animal welfare standards for these species, as Institutional Animal Care and Use Committees (IACUCs) control such experimental plans.

In Europe, Directive 86/609/EEC from 1986 has several relevant aspects for the field of alternatives, harmonizing the national legislations of the 27 member states:

> Article 7
> 2. 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.
> Article 8
> All experiments shall be carried out under general or local anaesthesia.
> 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. The Commission and Member States shall monitor trends in experimental methods.

Article 7 is—in theory—a very strong driver for alternative methods, but in practice there is very little enforcement. The revision 2010/63/EU of last year introduces, however, inspections and other reinforcements from 2013 [25].

Article 23 of the 1986 directive already has resulted in a substantial funding program for alternatives and their validation by the EU. More than €250 million was spent by the European Commission, and a similar amount was estimated to have been spent by the member states [26]. Of the 40 methods validated in Europe so far, however, hardly any have resulted from this funding, though some have been supported for further development and validation.

In conclusion, animal welfare legislation (especially when taking into consideration the revised EU legislation) is a relatively weak driver of change on both sides of the Atlantic. However, it prompted substantial funding schemes, which have led to the engagement of many researchers into the field.

### *2.1.2. EU cosmetics legislation*

The 7th amendment (Directive 2003/15/EC) of the cosmetics directive (Directive76/768/EEC) is the most rigorous legislation enforcing the use of alternative methods to date [27]. It is clear that the legislators want to uncouple the cosmetics industry from the slower progress in other areas, as described in Recital 5:

> "Currently, only alternative methods which are scientifically validated by the European Centre for the Validation of Alternative Methods (ECVAM) or the Organisation for Economic Cooperation and Development (OECD) and applicable to the whole chemical sector are systematically adopted at Community level. However, the safety of cosmetic products and their ingredients may be ensured through the use of alternative methods which are not necessarily applicable to all uses of chemical ingredients. Therefore, the use of such methods by the whole cosmetic industry should be promoted and their adoption at Community level ensured, when such methods offer an equivalent level of protection to consumers."

Key features of the directive are the four deadlines for phasing out testing and the enforcement of those deadlines by marketing bans:

- A deadline of 11 September 2004 for the testing of finished products (reinforced by a marketing ban).
- An immediate testing ban for ingredients if an alternative method is "validated and adopted at Community level with due regard to the development of validation within the OECD." This leaves some room for interpretation. Depending on the view, this means either: after the validity statement by ECVAM's Scientific Advisory Committee; after this "advice" is taken and adopted by ECVAM and the European Commission; after its acceptance by DG SANCO's Scientific Committee on Consumers Safety (SCCP) or acceptance and inclusion in the EU test guideline regulation (formerly Annex V of the Dangerous Substance Directive); or even acceptance as an OECD test guideline.
- A general testing ban on cosmetic ingredients from 11 March 2009 enforced for 10 animal test requirements by an instant marketing ban. Results from tests conducted outside of the EU until 2013 for all other tests can be used for notification.
- A marketing ban from 11 March 2013 for the more complex endpoints (those requiring repeated substance application, e.g., repeat dose toxicity, sensitization, reproductive toxicity, and carcinogenicity, as well as toxicokinetics, which is not typically required).

Importantly, the 7th amendment from 2009/2013 onward permits only the use of replacement alternatives. It should be noted that the legislation does foresee a review of the feasibility of the 2013 deadline in 2011 and can further postpone this in a co-decision procedure. The legislation is in many ways unique, as it phases out essential safety tests before alternatives are available. This "incentive for change" probably reflects the legislators' discontent that the 6th amendment of 1993 led only to two postponements of the already foreseen phasing out of animal testing. The 6th amendment introduced a marketing ban on cosmetic products tested on animals from 1 January 1998, provided that alternative testing methods had been validated and accepted by that date. The marketing ban has been postponed twice by the EU, on the grounds that insufficient progress had been made in developing and validating alternatives to the animal tests used for assessing cosmetic safety.

This legislation has strongly reinforced the cosmetic industry's efforts to make alternatives available, culminating in the matching funds for a total €50 million call for research proposals for systemic toxicity with the European Commission in 2010. As Europe is an exporter and importer of cosmetics, this resonates with legislation in other economic areas. Though animal use is minimal (0.03% of all animals are used directly by the cosmetic industry, with some indirect contribution of about 1% for new chemicals testing), the legislative pressure and the commitment by major cosmetic and consumer product industries has been a driving force for change. Even if our estimates [27] were too optimistic, the progress for acute and topical toxicity, i.e., the endpoints of the 2009 deadline,

has been remarkable. The more complex hazards of the 2013 deadline, however, cannot be met.

*2.1.3. REACH and TSCA reauthorization*

Increasing pressure in the 1970s led to regulation of environmental chemicals on both sides of the Atlantic [28]. With deadlines beginning in the early 1980s, all existing chemicals were grandfathered and regulations for new chemicals introduced. The approach was very different in the EU, however – the revision of the Dangerous Substance Directive (67/548/EEC) introduced production volume dependent testing demands starting at 100 kg per year. Most new chemicals have relatively small production volumes and 300–400 chemicals were registered per year with mostly light data packages. In the US, the Toxic Substance Control Act (TSCA) requires premarketing notification for any substance without data, and typically 1500–2000 substances are notified per year; only 15% of these contain toxicity testing data. The Environmental Protection Agency (EPA) then has 90 days to request testing. This prompted a culture of structure/activity relationship-based assessments, which were the only way to respond quickly to the notification. However, only in about 200 cases out of 26.000 notifications actual testing demands resulted. Reassessments around the turn of the century, however, showed that new chemicals represent a tiny proportion of chemical commerce (1–3%) and indicated a dramatic gap in knowledge with regard to the existing "old" chemicals.

In Europe, Regulation (EC) No 1907/2006, known as REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals), first addressed this problem. The regulation from 2006 aimed to study all chemicals marketed at more than one ton per year in Europe (taking the amounts of all producers and importers together). The process began with a mandatory pre-registration deadline of December 2008. A total of 180,000 pre-registrations were expected by about 27,000 companies for 30,000 substances. The newly established European Chemicals Agency (EChA), however, received more than 2.7 million pre-registrations from about 65,000 companies for 144,000 substances. We therefore reanalyzed [19] the previous estimates and found that all calculations were based on a survey from 1991–1994. At this time, the EU had 12 instead of the current 27 member states. Furthermore, the chemical industry has grown over the almost two decades at about 5% on average. Taking these factors into account, along with some last minute changes to the legislation, and applying the testing guidelines provided by EChA, we predicted more than 68,000 chemicals, requiring the use of more than 54 million animals and testing costs of more than €9 billion [20,29]]. This is a theoretical scenario only, it is important to note, and the results will differ because we simply do not have the laboratory capacities and expertise to carry out the magnitude of work required. The scenario assumes that tonnage levels were equally distributed in the 1991–1994 survey and that the increase was homogenous – and both assumptions are certainly wrong. The dimension of the challenge, however, is very clear. EChA, in the meantime, has corrected their estimate from 2 to 9 million animals [30] without updating the 1991–1994 numbers. This has three possible consequences: (a) REACH will take much longer, (b) a lot of testing will be waived, or (c) new methods will be considered. Most probably it will be a mixture of all three, which is reasonable. We should, however, use waiving very carefully, otherwise the whole REACH exercise becomes useless, if not substantially new information is generated.

We also tend to forget that most of our regulatory toxicology was developed for drugs under development [18] – a very specific situation, in which entirely new substances (usually of a group and tailored to a certain part of the chemical universe only) are moved through a decision process that addresses enormous safety concerns for treating patients. We have copy/pasted these approaches for other products, especially chemicals, simply because new chemicals were produced at a low volume not prompting these testing requirements. It is easy to prescribe heavy testing demands for new chemicals produced at very high volumes, while it is rather unlikely that such production volumes, typical for petrochemicals or basic chemicals, will ever be required by newly introduced chemicals. That is why we find, after more than 25 years of the Dangerous Substance Directive in Europe, fewer than 20 cancer bioassays and fewer than 100 two-generation studies [18,31]. But the situation changes entirely when the same testing demands are applied to HPV chemicals. Legislators have lowered the information requirements, but only on the other side of the spectrum: there is no longer any testing below one ton (notable also for all new chemicals to which REACH applies), and below 10 tons the testing requirements are minimal, requiring just two animal tests for acute toxicity and sensitization. Some opinion leaders in toxicology have rightly suggested that new chemicals will no longer receive an adequate assessment in Europe [32].

## 2.2. Animal welfare

While there is broad consensus on the need for laboratory animal welfare, there are psychological and practical barriers to the implementation of alternatives. Creating awareness of these barriers and providing support for change to researchers are necessary. Mandatory enforcement, such as that provided by animal use committees, encourages the use of alternatives. Animal welfare is also driven by organized groups, which directly impacts policies. This is much more pronounced in Europe, where these groups participate in the consensus process and are involved as stakeholders. Through political lobbying, these groups have considerable influence, reflecting attitudes that are increasingly popular among the general public.

## 2.3. Throughput of testing

There is a dramatic disconnect between the number of substances in our environment and the number tested (Fig. 1). And many of the substances have received only a so-called base set of tests, typically the rather inexpensive and fast tests that not necessarily represent their complete safety profile.

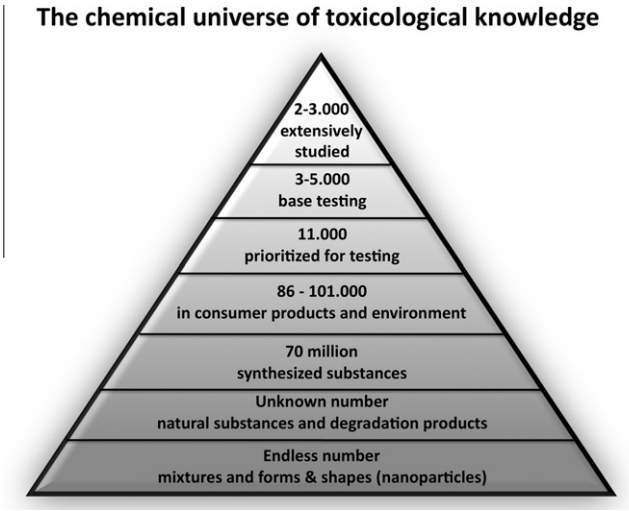


**Fig. 1.** The numbers are estimated by the author based on databases from the US National Toxicology Program, the US Environmental Protection Agency, the EU New Chemical Database, etc. The number of chemicals in use is based on the EPA TSCA and EU EINECS database. The number of substances in chemical abstracts available was estimated for the number of chemicals synthesized.

Individual tests can be extremely costly, e.g., €400,000 for a two-generation reproductive toxicity study, €1 million for a cancer bioassay, or €1.4 million for a developmental neurotoxicity study (according to OECD guidelines). Capacities in Europe for these tests are less than 100 each [33], mostly for pesticides and pharmaceuticals. With programs like REACH and TSCA reauthorization, the problem of throughput becomes obvious. It is also the key obstacle for frontloading toxicity testing in drug development, i.e., testing many more substances earlier to avoid late surprises of toxicological findings when the product is very advanced and has already consumed enormous development resources.

### 2.4. Economics

Animal numbers correlate roughly with costs [34], as we will argue again later. A rough calculation of future costs for chemical safety testing is illustrative: over the last decade in Europe, approximately 300 chemicals and pesticides (roughly representing the same mix of high and low production volume chemicals test requirements) were tested per year, costing about €600 million [14] – an average of €2 million per chemical. We now want to test 30–68,000 (REACH) to 86,000 (TSCA) chemicals in two decades, i.e., 1500–4300 per year (estimated 3000/a). We have to add costly evaluations [35] of an estimated several thousand nanoparticles (estimated 200/a) and increasing test demands for mixtures (estimated +10% test demand). Furthermore, we have to add new endpoints (endocrine disruption, developmental neurotoxicity, child asthma, immunotoxicity, etc.) (estimated +20% costs). This results in about €8 billion per year, a 15-fold increase compared to today with the addition of more than 4 million animals per year. Again, an unrealistic scenario, which can only be solved when new, cheaper approaches are found.

On the macroeconomic level, society needs to decide how many resources it wants to contribute to gain what are likely to be small increments in safety. We have more than doubled life expectancy since the introduction of chemicals into modern life. On the other hand, sorting out or delaying more efficient drugs, hygiene products, or health-improving functional foods also kills people. There must be a balance between precaution and progress, but this is difficult since the effect of precaution is not systematically assessed.

The economic consequences of the scientific limitations are already evident – and not just direct economical ones. We recently calculated [14] that safety assessment of substances costs about 0.03% of the turnover of the regulated products. But indirect costs are increasingly seen – longer time to market, wrong development decisions, product withdrawals, attrition rates in drug development, and precautionary approaches that eliminate far too many substances. It is frightening to realize that aspirin would likely not make it to the market today, since it fails too many routine tests [36]. Innovation, which is slowed down by the artificial delays in the introduction of new technologies, is often overlooked as a driving force of business and science. Many companies are hesitant to invest in the development of new technologies, which are difficult to protect by patents and can take a decade to gain acceptance before entering the market.

### 2.5. Predictive value of tests

Pharmacology and toxicology can be considered sister disciplines. They both address the desired or undesired effect of substances on humans. The appreciation of animal tests, however, is very different in both areas. A pharmacologist sees animal testing as a valuable but very limited tool to predict human effects: roughly 90% of drugs fail in human trials [37]. The main causes of failure in the clinic include safety problems (about 20%) and lack of effectiveness (about 40%) [38], both predicted by a series of animal models before entering the most costly part of drug development. The inability to predict these failures before human testing or early in clinical trials dramatically escalates costs.

In only very few other instances, we have human data to control the predictive capacity for toxicity test results. While experts estimate the proportion of reproductive toxicants at between 2% and 3%, two-species assessment in two generations appear to result in more than 60% positive results [18]. The estimated prevalence of 5–20% carcinogenic industrial chemicals corresponds to more than 50% of all chemicals testing positive in the cancer bioassay [39]. Similarly, in another study, 40% of the chemicals that irritated the skin of rabbits were found not to be irritants in the skin 'patch test' in humans [40].

Most of the methods we have were introduced for the safety evaluation of drugs under development. These are substances meant to be bioavailable and to have biological effects. Typically, there are also a number of structural variants from which to choose the lead compound in case a toxicological problem develops. At this stage in drug development, there is neither experience with human exposure nor a market value (other than the development costs). The question arises: Is a toxicology designed for pharmaceutical substances under development, used, to some extent, for new low production volume chemicals over the last decades, relevant for the assessment of existing high production volume chemicals? Not without, at minimum, an assessment of its suitability. We need to determine how cautious we want to be. The hassle involved in restricting the use of, or substituting for, substances with complex use scenarios may be considerable and should be undertaken only if well warranted. We also need to ask whether the methods are applicable to the types of substances. We might think that we have a lot of experience due to the Dangerous Substance Directive (which was in effect from 1981 for new chemicals until it was replaced by REACH), or the parallel Toxic Substance Control Act in the US. For many tests, despite being prescribed for three decades for new chemicals, we have, in fact, minimal experience with industrial chemicals, because they were not triggered for the relatively small production volumes required of new chemicals. We found 14 cancer bioassays and 46 two-generation studies in 28 years for about 4500 notifications in Europe. It is easy to accept major testing demands for high production volume chemicals if they are never applied. Now, 30 years later, it is obviously difficult to complain that they were entered in the first place.

### 2.6. Outdated science

It is most astonishing that the only area where we hinder new knowledge and methods from entering scientific discussions is where our safety is concerned. In regulatory toxicology, modern mechanistic toxicology comes as "mustard after the meal" to interpret what the canon of tests has delivered. We are using animal tests essentially the same way we have for 40–80 years. And we are using rats and mice not because these are the most predictive models for human beings, but because they are inexpensive, easy to breed and keep, and allow some throughput (though by far not enough, as discussed above). Once again, this recalls the image of the man looking for his key under the streetlight in the night, not because he lost it there but because it is well lit. In the same way, we do what is feasible – not what is necessary.

Scientific progress in the life sciences, often perceived as a biotechnology revolution, has been little embraced by regulatory practice. These represent a driver of change from neighboring disciplines and an enormous opportunity for the construction of a new toxicology. Scientific progress was also made in two decades of development and validation of alternative methods. Much of this now forms a basis for the construction of a new regulatory science [2].

### 2.7. New test needs

Regulatory toxicology is driven by scandals and perceived threats. New test requirements can thus be prompted by problematic products, such as the Draize rabbit eye test by the LashLure scandal [41] or reproductive toxicity testing brought about by the thalidomide (Contergan) disaster. But also assumed connections, such as endocrine disruption (reducing male fertility), increases in allergy/asthma, or neurobehavioral disorders in children create testing needs. New products also expand testing needs: biologicals (especially proteins) as drugs, cellular therapies, genetically modified foods, and nanoparticles are examples of products challenging the traditional regulatory approach. Toxicology will always need to adapt to novel products and substances, but it appears that the biotech revolution has strongly accelerated this trend.

## 3. The limitations of current tools

*"All models are wrong, some models are useful"* – this quote, from George E.P. Box, in a book chapter in *Robustness in the Strategy of Scientific Model Building* from 1979 [42], says it all. However, each tool has its own characteristics and thus its own limitations [43,44]. When describing the typical limitations (and advantages) of the major tools of regulatory testing, we must not forget that they do not hold true for all. This could well be expanded to any other approach from high-throughput testing, the use of lower organisms, or even human trials and epidemiology.

### 3.1. In vivo

For more extensive discussion, please see [43]. The advantages of animal models are obvious – they are living beings with hundreds of tissues and all physiological reactions and interactions are easily exposed. The basic technology is simple and not too expensive. Decades of experience and international harmonization aids interpretation and acceptance of data. However, limitations of the use of animal models are evident. The inevitable suffering and killing of animals, in particular, raises increasing concerns among the general public. Species differences become more and more evident, with concordances and correlations only around 60% for different species. We cannot rely on negative data (although we do) since the absence of evidence of an effect is not evidence of absence of the effect. Often, a hazard might be masked *in vivo*, for example, by protective metabolism not present in humans or a target structure/pathway of toxicity not present in the animal. Gene annotations show that genes essential in one species are not necessarily so for another. Animal models require the administration of relatively large amounts of test substances. The use of inbred strains does not reflect natural variances and the use of typically adolescent animals from one sex does not reflect age and sex differences. Data are difficult to interpret because of the complexity of interactions. The exposure scenarios are often unrealistic (maximum tolerated doses, for instance, and no co-exposures) [45]. The study design and statistical analyses are often poor (small group sizes "underpowered" with no correction for multiple testing). The predictive value of the tests (sensitivity, specificity) and prevalences (i.e., the proportion of toxic substances in the real world) to deduce predictive capacity (how much can you rely on a positive or negative finding) are not known. The tests are typically done only once [46] and often not published (especially negative findings) [47] or peer-reviewed, defeating the self-control mechanisms of science. The integration and interpretation of data often occurs in a case-by-case manner with limited transparency of the decision process.

### 3.2. In vitro

*In vitro* models originate directly from the current mainstream scientific approach in the life sciences and reflect our current mechanistic understanding of the health effects. Basic methodologies are widely established. Set-ups are small and use minimal amounts of test substances and allow low costs, high numbers of replicates, miniaturization, and automation. Novel technologies are quickly emerging, including imaging and diverse 'omic' technologies. The simplistic approach allows access to the test material and eases interpretation. Cell models for practically all tissues or laboratory animal species – and most importantly, humans – are available and can be compared. There are few ethical concerns, with the notable exceptions of human tissue donation and embryonic stem cells.

However, there are also clear limitations of *in vitro* models [48]: The broad use of *in vitro* tests in academic and industrial research in recent years has somewhat obscured the difficulties involved. Good Cell Culture Practices (GCCP) have been proposed [49] but are rarely applied. Guidance for Good Laboratory Practices to extend to *in vitro* studies for regulatory work have been developed based on GCCP, but again, its application in practice is not clear. There are some fundamental problems, including the artificial and non-physiological conditions in which the cells are maintained (not reflecting the body temperature of animals, the blood electrolyte concentrations of species, the extracellular matrix, or the extent of cell contacts, which is maximally 15% of normal in monolayer cultures). Cell densities are less than 1% of actual tissues, which impairs intracellular signaling. Most cell systems represent only one cell type (no cell–cell interactions), often monoclonal in origin and further degenerated during maintenance culture. Culture conditions are not homeostatic (sudden exchange of media, continuous depletion of nutrients, and accumulation of waste products) and oxygen supply is not sufficient (dissolved oxygen is typically consumed during the first hours allowing only diffusion-limited supply, resulting in anaerobic culture conditions, i.e., glycolysis seen as lactate accumulation by phenol red turning yellow). All growth conditions are usually optimized for rapid growth of cells and continuation to the next experiment – cell growth and differentiation, however, are just opposite programs (as we well know from cancer). We are driving cells into dedifferentiation and to select for mutations and subpopulations, which grow faster and do not waste time on differentiated cell functions. The cancer origin of many cells commonly used adds to this problem: It has been shown that cancer cells sometimes have tens of thousands of mutations, including losses of parts or entire chromosomes.

### 3.3. In silico

In toxicology but also in environmental health sciences in general, *in silico* tools [50] have a remarkably dichotomous reputation: while some colleagues uncritically embrace them, others are reluctant, skeptical, and avoidant. The advantages of *in silico* tools are certainly costs, standardization, minimal equipment needs, less necessary training, as well as duration of execution. They can also be easily integrated with other tools into integrated testing strategies (ITS, see below). In general, *in silico* tools come with the most advanced biometry (a weakness of many toxicological tests) and the clearest defined applicability domain.

Since all typical *in silico* tools in toxicology depend on the input from either *in vivo* or *in vitro* systems, they reflect their shortcomings. The limited public availability of such data, the lack of homogeneity of data sets, and the selection bias of data made available further impair every modeling. In consequence, essentially no model has stood external validation so far.

REACH is a strong promotor of *in silico* approaches. Notably, the treatment of *in silico* and *in vitro* methods in REACH is very different [51] – validation is not explicitly required and identification of non-toxic substances is much easier. *In silico* tools are still undergoing dramatic developments. The (Q)SAR tools, which derive correlations based on physico-chemical descriptors and which were in vogue at the time of drafting the REACH legislation, have generally not fulfilled expectations [51,52]. They are now increasingly substituted with mode-of-action-based systems, systems mining existing literature, and combinations with biological data in systems toxicology approaches. This reflects normal scientific development in a vibrant area, but also shows how dangerous it can be when legislation is tailored toward particular methodologies.

## 4. The emerging solutions

Interestingly, the two large-scale testing programs for chemicals, REACH, and the emerging TSCA reauthorization, have resulted in very different emerging solutions: In Europe, integrated testing strategies (ITS) have been introduced into the REACH testing guidelines for industry. In advance of TSCA revision, the US EPA commissioned the National Research Council to develop a vision for toxicity testing in the 21st century (Tox-21c). This 2007 report [3] represents the starting point of an intensive discussion of a renovation of the toxicological toolbox; it has created a remarkable departure from business-as-usual in the US. At the same time, the EPA ToxCast program (http://epa.gov/ncct/toxcast/) and the Tox21 partnership (http://ntp.niehs.nih.gov/go/28213) between different US agencies continue to promote the report's recommendations. Most remarkably, the EPA has already made the novel approach its official toxicity testing paradigm [10,11]. In contrast to the combination of largely established, integrated tools, the US approach is based mainly on new technologies, which are largely absent in the EU guidance. Tox-21c instead suggests basing a new toxicology primarily on pathways of toxicity (PoT). A mapping of such PoT shall enable the construction of a new test battery.

### 4.1. Integrated testing strategies

It is astonishing how long toxicology has survived with essentially one test per hazard, given the economical and societal importance of these safety decisions. Each test can reflect the human situation only in part – there will be also a grey zone, because of differences in various subpopulations, use scenarios, and cofactors. A test can only be optimized to be either sensitive (few things missed) or specific (few false accusations). However, while the former is in the interest of safety, typically leading to the use of very precautionary over-sensitive tests, there is also a societal desire for technical progress. When, for example, novel drug therapies do not make it to market because of precautionary tests, this too kills people who do not receive a possible treatment.

Only by combining sensitive and specific tests, we can optimize this balance. For this, however, we have to give up the illusion of "definitive tests." Furthermore, it is unrealistic that one test covers all PoT (also called "modes of action," or, most recently by OECD, "adverse outcome pathways"). This, too, argues for a combination of tests. Furthermore, test combinations can save costs with cheap screening tests or by minimizing animal use.

However, ITS are still in their infancy: we do not have strategies, knowledge of how to compose, validate, or adapt them. We also lack routine statistics for such complex decision trees. We might learn from clinical diagnostics, where similarly tests are combined for differential diagnosis. In many areas, decision theory has flourished, but in safety assessment of chemicals there are enormous opportunities.

### 4.2. New technologies

We have witnessed over the last 15 years a tremendous biotechnology and bioinformatics revolution. Small and large companies providing products from life science research have mushroomed. In the EU, this has had remarkably little impact on the novel concepts for toxicology, while the US Tox-21c movement is almost entirely based on them, be it diverse omics techniques, testing in lower species, image analysis systems, and high-throughput testing. This creates information-rich situations due to the high-throughput and high-content methods employed. These can only be handled because bioinformatics delivers the tools to mine and interpret these data agglomerates, often making use of the accumulated knowledge of mechanistic toxicology and biochemical/genetic pathways. Similar to "systems biology", "systems toxicology" is emerging.

It is beyond the scope of this article to give an overview or value of these new approaches. However, a word of caution seems to be warranted: we must not forget that the biological systems, which are under less development pressure (with the notable exception of stem cell models), limit the performance of the entire system. High-tech measurements do not compensate for lousy cells or make animal models more similar to humans.

### 4.3. PoT-based toxicology

The central idea of Tox-21c is that a finite number of distinct PoT exist; if we are able to map them, we can build a test battery to cover them. Different authors, however, appear to have somewhat different concepts of what a PoT actually is, e.g., a number of genes or metabolites or a signal transduction cascade transmitting the interaction of the toxicant, a series of events leading to damage on cell or organ or inter-organ level, the physiological pathways that are disturbed by high doses of toxicants, and so forth. It is also not clear, whether loops, branches, and variants of PoT need to be taken into consideration. This gets likely even more complicated when species differences are taken into account. The first step will be to get the terminology right. Then, we can start annotating PoT. Most probably, we need to simply start with some examples in order to understand the feasibility of the approach.

### 4.4. Probabilistic risk assessment

A key challenge for the introduction of new approaches is to abandon the black-and-white world of toxicology: *A substance is not a carcinogen or not* vs. *it has a certain probability to induce cancer in a given use scenario.* Our tests cannot determine carcinogens but estimate this probability with some uncertainty. This is not good news for regulators and the regulated community, since they want to close the books after a risk assessment. But that is what science can actually deliver. This will make life more difficult but will also make clear what a toxicological test actually does, i.e., it changes the pre-test probability of a hazard to a post-test probability. The stronger the possible change, the more valuable the test is within the testing strategy.

Pre-test probability of hazard refers either to the prevalence of a certain hazard in the respective chemical universe or the subselection after previous test steps. The same test has very different value when used on all chemicals or after prioritization, for example, by a screening test or structure/activity relationships.

### 4.5. Evidence-based toxicology (EBT)

Evidence-based medicine (EBM), as steered by the Cochrane Collaboration [53], is a quality assurance movement that aims to condense and provide to health care practitioners the best avail-

able evidence for a given treatment, diagnostic, or health care question. EBM has been broadly reviewed and the literature is extensive. In essence, EBM was born from the need to handle the flood of medical information and sort the available evidence objectively, which includes traditional approaches and new scientific developments of variable quality. More than half a million papers included in MedLine per year, of an estimated more than 2 million in medicine every year, address questions relevant to the life sciences and therapy, and there is no way individual physicians could look at all this information. MedLine (http://www.ncbi.nlm.nih.gov/pubmed/), the most popular resource, covers over 5000 journals with an estimated 8000 citations entered per week [54]. For example, entering the search term "toxicology" for the time period of 2003–2009 results in 28,500 article hits in PubMed, a database which does not even cover all relevant publications in the biomedical field.

Instead of expecting individuals to determine what is the best evidence for a specific question or approach at a given time, high-quality reviews available at a central repository should represent a primary resource of information. This requires agreed quality standards, so that the individual physician can rely on the information. And here the key difference between evidence-based and traditional ("narrative") becomes apparent: most reviews represent a story told by (knowledgeable) authors who present their personal views on their topic of interest in a more or less well disguised manner. They tend to select their own papers and those that fit the storyline of their review. The literature included is largely what has accumulated over time and shaped the opinion of the author(s). This is sometimes distinguished from EBM as "expert-based" or "eminence-based," to make clear that individual experts are responsible for the content. The systematic review, which is the main tool of EBM, proceeds differently: the sources (typically MedLine and other literature databases) and a search strategy, i.e., a decision about which papers shall be considered and which disregarded, are defined upfront. Before collecting the actual articles, the procedure for information analysis is defined. Ideally, these search and analysis strategies are peer reviewed to safeguard objective and efficient processes. The analysis of the collected evidence requires weighing the quality (the second main tool of EBM) of individual pieces of evidence and summarizing these as objectively as possible. The latter often involves the third major tool of EBM, meta-analysis. Meta-analysis describes statistical approaches to combine results from different studies. These studies will differ in key parameters. By either factor analysis or stratification of all data by one parameter after the other, the influential parameters can be identified. Then, where possible, an overall quantitative answer to the well-defined research question may be deduced.

Obviously, toxicology has the problem of information flooding and co-existence of traditional and modern methodologies, as well as various biases [47]. It is difficult to find and summarize the relevant information for any given major question, as has been nicely illustrated by Christina Ruden [55]. She showed the divergence in judgment and limitations of analysis for 29 cancer risk assessments carried out for trichloroethylene – 4 assessments concluded that the substance is carcinogenic, 6 said it is not, and 19 were equivocal. The main reason for this divergence was a selection bias in the materials considered, i.e., an average reference coverage of only 18%, an average citation coverage of most relevant studies of 80%, as well as an interpretation differences in most relevant studies (27%), and the lack of documented study/data quality assessment in 65%.

The similar problems in toxicology and clinical medicine, especially the similarities between making a diagnosis in medicine and deciding on whether a substance is hazardous [56], prompted us to think about whether EBM tools could be suitable for toxicology [57]. Often, the standardization and formalization of processes and committees disguises the nature of our decision making. Some people correctly speak of the "art of toxicology" (though it is more a craft) – this much more accurately reflects the intuitive components. Certainly, it is an applied science, in which compromise and pragmatic decisions are necessary, but we should be clear about where we have to take such shortcuts, otherwise we will soon forget the limitations of our decisions and make them gold standards, textbook knowledge, and the unquestioned basis for further decisions (read-across, QSAR, new use and exposure scenarios, reference for validation, etc.). Many aspects of the now highly respected test guidelines for methods or classifications and labeling of substances are based on committee consensus decisions and not necessarily science.

Already in 1993, Neugebauer and Holaday applied EBM approaches to *in vivo* and *in vitro* studies [58]. We, and others, have earlier suggested translating this approach to toxicology [58–61]. This translation should, in a transparent and objective manner, assess the performance of tools used in risk assessment, develop means of meta-analysis for data from different studies including quality scores for existing studies [62], and establish causation of health effects. At this stage, EBT lacks critical mass though an International Forum [62] and various symposia have been held.

EBT primarily lends itself to objective assessment of the strengths and weaknesses of current approaches, which is a necessary prerequisite for integration of or substitution by new approaches. It also has potential to serve as a quality assurance mechanism for new technologies when formal validation is not (yet) appropriate [63]. This can be the case when there is no point of reference for validation – for example, where there is no traditional (animal) test.

We often confuse weight-of-evidence with evidence-based approaches: The term "weight-of-evidence" is commonly used to describe a process of making a decision based on different pieces of information, each not definitive or even contradictory. In the absence of clear procedures, this is a highly subjective process. In many aspects, this is the contrary of an evidence-based approach. The term comes from the legal field, where it means the measure of credible proof on one side of a dispute when compared with the credible proof on the other, particularly the probative evidence considered by a judge or jury during a trial (Farlex Legal Dictionary). The weight of evidence is based on the believability or persuasiveness of evidence. Believability is certainly not an EBM criterion.

## 5. The challenges ahead

It must be clear that we need to do more than develop the scientific tools for a new approach. Many such tools exist, but solutions are partial, dispersed, and often used for other areas like drug development. It will need some central steering to put this puzzle together and construct a new regulatory approach. However, we also need a transition process. Since we have limited feedback from regulatory decisions, we have no outcome measures as found in clinical medicine. A substance that does not make it to market – rightly or wrongly – does not do any harm, though the decision to prohibit innovation might harm the company and society. A substance, which slips through the controls and harms people, is rather unlikely to be detected if producing only complex or long-term health effects. So we need a process of trust-building for the approach, as it was done with formal validation for its forerunners – the alternative methods. Formal validation, however, has become a relatively rigid and cumbersome process, which lasts 3+ years for the validation study itself (after optimization of the

test), 2+ years for reporting and peer-review, and 2+ years for regulatory acceptance and International harmonization.

### 5.1. Composition of ITS

The closest we come so far to the construction of ITS are tiered testing strategies, e.g., for skin irritation [64]. This means we pass one test (set) after the other and typically drop out if we have a positive finding. The tiers are normally based on economic reasoning, with the inexpensive steps of evaluation of existing data and *in silico* methods used first, followed by *in vitro* tests where available before proceeding to animal tests. This does not take advantage of the potential of ITS: more accurate predictions and fewer tests can be achieved if decisions on the need for a specific test is informed by previous test results and dynamically adapted [65]. Both decision theory and clinical differential diagnostics can be useful, but for toxicology this is still in its infancy. Promising forms of modeling, such as bootstrap resampling, can be applied to optimize testing strategies. Hoffmann et al. [66] provide an example for skin irritation.

Testing strategies represent the key opportunity to balance the shortcomings of all approaches. They have a key disadvantage for their users, i.e., they require decision points in-between consecutive steps. The set of tests (and thus costs and timing) cannot be determined upfront. But earlier, "no-go" criteria within such a testing strategy might actually reduce time and costs, given that the organization, execution, and analysis of a cancer bioassay, for example, often lasts 4–5 years.

### 5.2. Definition of adversity in PoT-based toxicology

The use of novel technologies often creates information-rich situations, which are difficult to interpret [67]. Currently, the biggest challenge in using such testing methods is determining how to deduce a result [68]. Information-rich approaches are, in principle, a form of multiple testing, which means that we multiply our noise, false-positives, and false-negative calls. If testing ten thousand genes (such as in a toxicogenomics approach), we must be very careful to distinguish what is noise and what is a specific signal (via PCR confirmation, for example). Still the number of replicate measurements (due to costs) is often small. It is extremely difficult to avoid falling into the trap of artifacts. The definition of adversity, i.e., what is indicative of a human adverse reaction from these technologies, is only emerging [69]. This is particularly difficult if decisions are based on signatures, that is, pattern changes in gene expression or abundance of proteins or metabolites. We do not really know, for example, how signatures change if several hazards are exerted by one substance, when two or more substances are combined, or when concentrations are lowered to no-effect levels. We also do not know how to take signatures of defense mechanisms, which might counterbalance the signature of harm, into account. Actually, it is possible that we might regularly include such defense reactions into the signatures as they are specifically induced by the hazard. The way forward is to deduce PoT from the signatures to give the signatures meaning. Then, we can translate PoT to other species or understand whether variants of signatures represent the same perturbation of physiology or the same PoT are being activated.

### 5.3. Quality assurance (QA) and validation

Complex ITS or complex information-rich technologies present an enormous challenge to quality control. Every manipulation step needs standardization and control. We quickly encounter the level of difficulty when reasoning about the validation of toxicogenomics [70]. Establishing QA for the new technologies will be enormously challenging. The more variables a system has, the more documentation and control are necessary. Some of the new technologies are incredibly complex and the interplay of parameters and variables is not at all understood.

For validation [71], the basic need to establish the relevance of tests is to obtain a "gold standard" point of reference. This is often the result of a reference test (the traditional approach). We should be clear, however, that this implies that we cannot get better than this reference method. The only alternative is to anchor our validation with other reference points, such as clinical data, composite knowledge on certain substances, or well-understood mechanisms. We have suggested earlier [72–74] that such points of reference are chosen, but this will require expert judgment and consensus processes (for example, see [75]).

The major challenge is that we are confronted with more and more partial models, i.e., models reflecting a certain mode of action of toxicity or target structure. As stand-alone models, they cannot substitute for an *in vivo* test. Typically, they are suggested as filters for sorting out positive substances. Most of the (industrial) chemicals, however, are not shown to be hazardous, even when we apply our current over-sensitive (precautionary) animal tests: 90% of chemicals are not acutely toxic, two-thirds are not skin sensitizing, 80% are not eye-irritating, 93% are not skin-irritating, 97% are not skin corrosive, etc. [57]. This means that we need methods to identify harmless substances or we will change very little. The solution suggested by Tox-21c is a comprehensive mapping of pathways of toxicity, whereby it might be possible to break down a particular hazard into the relevant pathways. If none of the pathways are influenced, we might be able to largely exclude potential hazard. Consequently, the partial models of a larger testing strategy would need to be validated against prototypic substances disturbing a given pathway, something we have called "mechanistic validation" [72,76].

### 5.4. Implementation

The need for a novel approach becomes more pressing every day. We can neither afford the attrition rate in drug development nor a precautionary elimination of high production volume chemicals in REACH or TSCA reauthorization [28], just as we cannot afford to ignore the multitude of old and new hazards to human health and the environment. Nanoparticles [77], genetically modified food and feed, and cell therapies are only a few of the latest additions to the catalogue. Additional health concerns, such as endocrine disruption, childhood asthma, and developmental neurotoxicity (autism, attentions deficit syndromes), and arteriosclerosis possibly associated with nanoparticles add to this. At this moment, the testing needs are growing faster than the solutions. If we want to turn this around, we need to aim for a new technology – failure is an option, not trying is not.

There is an understandable desire among global companies to harmonize approaches to risk assessment. They like to carry out a set of tests and "close the books". International harmonization of test guidelines by OECD for chemicals/pesticides and the (Veterinary) International Conference on Harmonization (ICH and VICH) have been very much appreciated and, by avoiding duplicate testing, probably saved more animals than any alternative method. But the very same mechanism can be an obstacle for the implementation of new methods [78]. Since not all countries or economically relevant areas adhere to the OECD and (V)ICH guideline process and follow their changes, it is appealing for companies to stay with the traditional methods until the last relevant economic area has converted. The example of the local lymph node assay (LLNA) for skin sensitization is instructive: although it has clear advantages over the guinea pig assays (reduction, refinement, semi-quantitative, preferred method in EU and OECD), it was utilized in less than

5% of new chemical notifications since its introduction [18]. There are very few mechanisms to force the implementation of alternatives – the horizontal legislation in Europe requiring that an alternative must be used if available loses all impact when companies can argue that they have to do the traditional test anyway for other economical areas. Prescribing the new tests in legislation (as, for example, the LLNA in the REACH legislation) is one way, but this anchors a particular methodology as the gold standard, discouraging further alternative developments. In general, an information requirement, such as skin sensitization hazard information, should be prescribed, allowing adaptation to technical developments. To force change, it might be worthwhile to insist on the execution of the alternative method despite the requirement for the traditional method elsewhere; this will also generate parallel testing data to build confidence into the new approach or discern differences. Furthermore, it will urge companies to push for change in their respective foreign regulatory arenas. Another option, which has rarely been used, is the elimination of the traditional method from the guidelines. The most pertinent example is certainly the classical $LD_{50}$ test for acute toxicity (OECD test guideline 401), which was abandoned with the introduction of tiered testing strategies – thus forcing change.

The best way forward, however, is toward International harmonization to optimize the post-validation process [79,80]. This eases and accelerates both acceptance and implementation of new methodologies. Remarkable progress has been made on various levels (EU test guideline regulation, OECD, ICH, and the creation of the International Collaboration of Cosmetic Regulators, or ICCR [27]). International collaboration of validation bodies is also extremely helpful – the author's call to create an International Council of Validation Bodies at the World Congress on Alternatives and Animal Use in the Life Sciences in Tokyo (2007) led within two years to the creation of ICATM, the International Collaboration on Alternative Test Methods (http://iccvam.niehs.nih.gov/about/icatm.htm). It is crucial that the four partners (US, EU, Japan, and Canada) integrate with the emerging new validation bodies (e.g., 2009 in Korea and possibly soon in Brazil [81,82]). International harmonization is as important as economic harmonization. The precautionary approach of drug safety assessment should not be used for all regulated products; rather, new developments in other sectors should be applied. Food is one sector still largely outside of the 3Rs or Tox-21c developments [83]. It is also unfortunate that the novel plant protection product legislation in Europe does not incorporate the REACH approaches.

## 6. Perspectives

In the development of the life sciences in recent years, *in vitro* approaches have contributed largely to the biotech revolution. Regulatory toxicology has embraced them only in part, especially in the field of genotoxicity and, more recently, for hazard identification in topical toxicity. Mechanistic models play an enormous role in supporting regulatory decisions. Still, the core of regulatory approaches has not changed despite many decades of use (40–80 years). One reason for the slow adaptation to advances in science and technology is that internationally harmonized guidelines, once agreed upon, are difficult to change. There are other drivers (e.g., psychological and economical) that make questioning the status quo inconvenient. The review literature on the limitations of basic toxicological tools is astonishingly scarce, with the notable exception of the cancer bioassay (of enormous interest because of public concerns and animal studies which frequently cost more than $1 million per substance). In contrast, *in vitro* tests undergo the most extensive evaluation of any models in the life sciences; the validation process easily exceeds $500,000 per test and often requires decades. The process – although only begun for the most promising candidate tests – results in two-thirds of the candidates being established as not (yet) valid. Most of these die a silent death and failures are not necessarily published. More general reviews of the limitations of *in vitro* approaches are also rare.

Even if we are able to create technological solutions for a new type of toxicology, there is still a long road ahead [68,2]. What we need is a pathway definition, a pathway annotation scheme, and a central repository of the annotated pathways. Methods to identify pathways and their targets need to be identified and evaluated. The big challenge will be to use orthogonal methods; for example, combining pathways identified as leading to clinical manifestations by gene polymorphisms with results of toxins identified in omics approaches. Gene silencing, species comparisons, and the like will help identify them further, as will our knowledge of physiological and biochemical interactions of genes, their proteins, and their metabolites. And these are only the first steps down the road, followed by definition of test strategies, quality assurance, validation, international harmonization, objective analysis of current practices by evidence-based toxicology, etc. Tox-21c is moving the discussion in this direction, but it has not been sufficiently acknowledged that Tox-21c is the answer to some of the major challenges posed by REACH. The fact that Tox-21c was prompted by the anticipation of the reauthorization of the Toxic Substance Control Act (TSCA), which is likely to become the US equivalent to REACH, should have made this more obvious. The call for action is getting louder [2,6,9,12,84]. We have already formed an informal implementation group for Tox-21c, organized by CAAT, to promote the various activities (the fourth meeting is planned as a satellite to the SOT conference in March 2011). However, we will need to transform the group into a consortium to take action. The €50 million project initiated by the European cosmetic industry for novel approaches to systemic toxicity has the chance to become a nucleus for such developments.

It is not even clear, though, whether the new approach will be more predictive or more cost-effective than the old. Estimating the efficacy of the current risk assessment is already an enormous problem – the relative absence of scandals might be a delusion. We lack essentially all control mechanisms of the normal scientific process, such as repetition of experiments and public availability of results. And the long delay of exposure and health effects makes it almost impossible to retroactively trace chronic effects. Where we have some indications (e.g., from cross-species studies, human trials or repeated efforts), the results are often disappointing. Between species, complex endpoints usually correlate by only about 60%, and human trials illustrate both missed substances and overlabeling. But the risk assessment process integrating such data also has its shortcomings. The quality of individual tests also warrants some doubt, as well as the use of data produced for risk assessment. It will not, however, be easy to assess the possible gain of safety or its impairment by new approaches on this ground.

New technologies have no fewer limitations than old ones: The shortcomings of *in vitro* approaches [48,49] have been discussed earlier. We often forget that the golden rule for computational models [50] is garbage in, garbage out, and this is true for *in vitro* or *in vivo* models as well. The most sophisticated analysis procedure does not change this.

Another enormous challenge will be to translate such hazard findings to a risk paradigm, i.e., to combine them with exposure considerations [6]. The classic strategy requires a dose–response relationship. Here, only few new approaches are emerging, such as physiology-based (pharmaco-)kinetic (PBPK) modeling or human microdosing. Also frequently overlooked is that we are dealing not only with hazards to humans, but also to our environment and veterinary animals. Expansion to these areas will avoid a disconnect. Despite the challenges, toxicology is a $3–4

billion industry [14] with likely growth (as shown for REACH [20]) – providing incentive to pursue state of the art technologies.

The transatlantic divide – the focus on alternative methods (3Rs) in Europe and the drive to new technologies (Tox-21c) in the US—is a challenge and an opportunity [2]. There is certain reluctance in Europe to embrace the latest technologies for the development of alternative approaches. This is well warranted, considering lessons learned about the difficulties of standardization, quality control, and validation of alternative methods. One lesson is that it takes great effort to make models robust, reproducible, and relevant. Certainly, the most sophisticated measurements cannot compensate for poorly conceived cellular models – a problem often overlooked in technology-driven developments. Tox-21c can take lessons from the 3Rs – in essence, the first decade of validation of cellular tests resulted only in a better validation process.

At the same time, classical 3Rs methods slowly approach a ceiling, where the "easy" topical and acute toxicities have been successfully tackled. Without a shift in methodology and increased effort, we will not be able to expand these successes to more demanding complex endpoints. Tox-21c has clearly identified the promise of new technologies and the need for large-scale efforts. Ultimately, this will require the comprehensive mapping of pathways of toxicity, sometimes referred to as the Human Toxicology Project [2,84]. This is the key proposal of the Tox-21c report, which even provides an estimate of about $100 million per year over one to two decades. The obvious association with the Human Genome Project is correct with regard to the similar financial dimensions and the need for international collaboration. The basic idea is to base this new toxicology on identification of PoT (still not clearly defined) and an analysis of which test reflects which pathways. In addition, a type of retro-PBPK must be developed. "Retro-PBPK" means that we change the classical pharmacological direction of PBPK, i.e., from calculating a tissue concentration from a dose of the drug applied to calculating the dose required to achieve a tissue concentration found relevant in the *in vitro* test (thus linking dose and hazard).

We should keep in mind, however, that a key goal of any risk assessment is the identification of the typically much more abundant and much more useful non-toxic substances. The hope, here, is that a comprehensive mapping of pathways shows that a substance does not trigger relevant concerns. This requires that our list of pathways is sufficiently complete and covered by the test battery. A Human Toxicology Project could be far more than a catalogue of cell/chemical interactions – it could boost our understanding of how to manipulate cells and how to avoid the disturbance of critical pathways. This would impact not only toxicology, but identify targets for drug development and provide research tools to manipulate cellular pathways. As such, it would bring 21st century science into broad use for preserving our health and environment.

## References


[1] W.S. Russel WMS, R.L. Burch, The principles of humane experimental technique, Methuen, London, 1959. <http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc>.

[2] T. Hartung, Lessons learned from alternative methods and their validation for a new toxicology in the 21st century, J. Toxicol. Environ. Health B 13 (2010) 277–290.

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

[4] B.J. Blaauboer, M.E. Andersen, The need for a new toxicity testing and risk analysis paradigm to implement REACH or any other large scale testing initiative, Arch. Toxicol. 81 (2007) 385–387.

[5] S. Gibb, A review of the National Research Council report toxicity testing in the 21st century: a vision and a strategy, Reprod. Toxicol. 25 (2008) 136–138.

[6] M.E. Andersen, D. Krewski, Toxicity testing in the 21st century: bringing the vision to life, Toxicol. Sci. 107 (2009) 324–330.

[7] R.J. Kavlock, The future of toxicity testing—the NRC vision and the EPA's ToxCast program national center for computational toxicology, Neurotoxicol. Teratol. 31 (2009) 237.

[8] R.J. Kavlock, C.P. Austin, R.R. Tice, Commentary – toxicity testing in the 21st century: implications for human health risk assessment, Risk Analysis 29 (2009) 485–487.

[9] F.S. Collins, G.M. Gray, J.R. Bucher, Toxicology: transforming environmental health protection, Science 319 (2008) 906–907.

[10] EPA, 2009. <http://www.epa.gov/osa/spc/toxicitytesting/>.

[11] M. Firestone, R. Kavlock, H. Zenick, M. Kramer, The U.S. EPA Working Group on the future of toxicity testing, The U.S. Environmental Protection Agency strategic plan for evaluating the toxicity of chemicals, J. Tox. Environ. Health B 13 (2010) 139–162.

[12] T. Hartung, Towards a new toxicology – evolution or revolution?, ATLA – Altern Lab. Anim. 36 (2008) 635–639.

[13] European Commission, Fifth Report from the Commission to the Council and the European Parliament on the Statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union, 2007. COM/2007/675 final. <http://ec.europa.eu/environment/chemicals/lab_animals/reports_en.htm>.

[14] A.A. Bottini, T. Hartung, Food for thought... on economics of animal testing, ALTEX 26 (2009) 3–16.

[15] M. Leist, S. Kadereit, S. Schildknecht, Food for thought ... on the real success of 3R approaches, ALTEX 25 (2008) 17–24.

[16] T. Hartung, M. Leist, Food for thought ... on the evolution of toxicology and phasing out of animal testing, ALTEX 25 (2008) 91–96.

[17] M. Leist, T. Hartung, P. Nicotera, The dawning of a new age of toxicology, ALTEX 25 (2008) 103–114.

[18] T. Hartung, Toxicology for the twenty-first century, Nature 460 (2009) 208–212.

[19] C. Rovida, T. Hartung, Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals, ALTEX 26 (2009) 187–208.

[20] T. Hartung, C. Rovida, Chemical regulators have overreached, Nature 460 (2009) 1080–1081.

[21] T.S. Kuhn, The Structure of Scientific Revolutions, second ed., University of Chicago Press, Chicago, 1970.

[22] P. Skarabnek, J. McCormick, Follies and Fallacies in Medicine, third ed., Tarragon Press, Whithorn, 1998.

[23] EUROHORC (2009). European Directive 86/609 on the protection of animals used in scientific procedures: summary of the opinions of the members of EUROHORCs on the main areas of concern on the revised draft of the Directive. 6 April 2009. <http://www.eurohorcs.org/SiteCollectionDocuments/EU_animals_Directive_revision_0904.pdf>.

[24] L. Festinger, A Theory of Cognitive Dissonance, Stanford University Press, Stanford, CA, 1957.

[25] T. Hartung, Comparative analysis of the revised Directive 2010/63/EU for the protection of laboratory animals with its predecessor 86/609/EEC – a t4 report, ALTEX 27 (2010) 285–303.

[26] T. Devolder, K. Reid, V. Rogiers, S. Webb, D. Wilkins, A review of national public funding programmes in European countries, ALTEX 25 (2008) 233–242.

[27] T. Hartung, Food for thought ... on alternative methods for cosmetics safety testing, ALTEX 25 (2008) 147–162.

[28] T. Hartung, Food for thought... on alternative methods for chemical safety testing, ALTEX 27 (2010) 3–14.

[29] T. Hartung, C. Rovida, That which must not, cannot be... a reply to the EChA and EDF responses to the REACH analysis of animal use and costs, ALTEX 26 (2009) 301–305.

[30] ECHA (European Chemical Agency), New study inaccurate on the number of test animals for REACH, 2009 Press release posted Helsinki, 28 August 2009. <http://echa.europa.eu/doc/press/pr_09_11_animal_testing_20090828.pdf>.

[31] S. Bremer, C. Pellizzer, S. Hoffmann, T. Seidle, T. Hartung, The development of new concepts for assessing reproductive toxicity applicable to large scale toxicological programs, Curr. Pharm. Des. 13 (2007) 3047–3058.

[32] H. Greim, M. Arand, H. Autrup, H.M. Bolt, J. Bridges, E. Dybing, R. Glomot, V. Foa, R. Schulte-Hermann, Toxicological comments to the discussion about REACH, Arch. Toxicol. 80 (2006) 121–124.

[33] M. Fleischer, Testing costs and testing capacity according to the REACH requirements – results of a survey of independent and corporate GLP laboratories in the EU and Switzerland, J. Busin. Chem. 4 (2007) 96–114.

[34] A.A. Bottini, T. Hartung, The economics of animal testing, ALTEX 27 (2007) 67–77.

[35] J.-Y. Choi, G. Ramachandrran, M. Kandlikar, The impact of toxicity testing costs on nanomaterial regulation, Environ. Sci. Technol. 43 (2009) 3030–3034.

[36] T. Hartung, Per aspirin ad astra..., ATLA – Altern. Lab. Anim. 37 (Suppl. 2) (2009) 45–47.

[37] J.A. DiMasi, H.G. Grabowski, The cost of biopharmaceutical R&D: is biotech different?, Manag Decis. Econ. 28 (2007) 469–479.

[38] I. Kola, J. Landis, Can the pharmaceutical industry reduce attrition rates?, Nat Rev. Drug Discov. 3 (2004) 711–715.

[39] B.N. Ames, L. Swirsky Gold, Paracelsus to parascience: the environmental cancer distraction, Mut. Res. 447 (2000) 3–13.

[40] D.A. Basketter, M. York, J.P. McFadden, M.K. Robinson, Determination of skin irritation potential in the human 4-h patch test, Contact Dermat. 51 (2004) 1–4.

[41] T. Hartung, L. Bruner, R. Curren, C. Eskes, A.M. Goldberg, P. McNamee, L. Scott, V. Zuang, First alternative method validated by a retrospective weight-of-

evidence approach to replace the Draize eye test for the identification of non-irritant substances for a defined applicability domain, ALTEX 27 (2010) 43–51.
[42] G.E.P. Box, Robustness in the strategy of scientific model building, in: R.L. Launer, G.N. Wilkinson (Eds.), Robustness in Statistics, Academic Press, New York, 1979, pp. 201–236.
[43] T. Hartung, Food for thought ... on animal tests, ALTEX 25 (2008) 3–9.
[44] T. Hartung, G. Daston, Are in vitro tests suitable for regulatory use? Tox. Sci. 111 (2009) 233–237.
[45] M.C.M. Rietjens, G.M. Alink, Future of toxicology – low-dose toxicology and risk–benefit analysis, Chem. Res. Toxicol. 19 (2006) 977–981.
[46] I.F.H. Purchase, Fraud, errors and gamesmanship in experimental toxicology, Toxicology 202 (2004) 1–20.
[47] B. Wandall, S.O. Hansson, C. Ruden, Bias in toxicology, Arch. Toxicol. 81 (2007) 605–617.
[48] T. Hartung, Food for thought ... on cell culture, ALTEX 24 (2007) 143–147.
[49] S. Coecke, M. Balls, G. Bowe, J. Davis, G. Gstraunthaler, T. Hartung, R. Hay, O.-W. Merten, A. Price, L. Schechtman, G. Stacey, W. Stokes, Guidance on good cell culture practice, ATLA – Altern. Lab. Anim. 33 (2005) 261–287.
[50] T. Hartung, S. Hoffmann, Food for thought on.... in silico methods in toxicology, ALTEX 26 (2009) 155–166.
[51] S.R. Johnson, The trouble with QSAR (or how I learned to stop worrying and embrace fallacy, J. Chem. Inf. Model. 48 (2008) 25–26.
[52] G.M. Maggiora, On outliers and activity cliffs – why QSAR often disappoints, J. Chem. Inf. Model. 46 (2006) 1535.
[53] K. Dickersin, E. Manheimer, The Cochrane Collaboration: evaluation of health care and services using systematic reviews of the results of randomized controlled trials, Clin. Obest. Gynecol. 41 (1998) 315–331.
[54] N.L. Wilczynski, R.B. Haynes for the Hedges Team, Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey, BMC Med. 2 (2004) 23–27.
[55] C. Rudén, The use of mechanistic data and the handling of scientific uncertainty in carcinogen risk assessments: the trichloroethylene example, Regulat. Toxicol. Pharmacol. 35 (2002) 80–94.
[56] S. Hoffmann, T. Hartung, Diagnosis: Toxic! – trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations, Tox. Sci. 85 (2005) 422–428.
[57] S. Hoffmann, T. Hartung, Towards an evidence-based toxicology, Human Exp. Toxicol. 25 (2006) 497–513.
[58] E.A. Neugebauer, J.W. Holaday, Handbook of Mediators in Septic Shock, first ed., CRC-Press, Boca Raton, 1993.
[59] P.S. Guzelian, M.S. Victoroff, N.C. Halmes, R.C. James, C.P. Guzelian, Evidence-based toxicology: a comprehensive framework for causation, Human Exp. Toxicol. 24 (2005) 161–201.
[60] T. Hartung, Food for thought... on evidence-based toxicology, ALTEX 26 (2009) 75–82.
[61] C. Griesinger, S. Hoffmann, A. Kinsner-Ovaskainen, S. Coecke, T. Hartung, in: Proceedings of the First International Forum Towards Evidence-Based Toxicology, Conference Centre Spazio Villa Erba, Como, Italy, 15–18 October 2007; Human Exp. Toxicol., Evidence-Based Toxicology (EBT) 28 (2009) 81–163 (special issue).
[62] K. Schneider, M. Schwarz, I. Burkholder, A. Kopp-Schneider, L. Edler, A. Kinsner-Ovaskainen, T. Hartung, S. Hoffmann, "ToxRTool", a new tool to assess the reliability of toxicological data, Toxicol. Lett. 189 (2009) 138–144.
[63] T. Hartung, Evidence based-toxicology – the toolbox of validation for the 21st century? ALTEX 27 (2010) 241–251.
[64] M. Macfarlane, P. Jones, C. Goebel, E. Dufour, J. Rowland, D. Arakie, M. Costabel-Farkas, N.J. Hewittg, J. Hibatallah, A. Kirst, P. McNamee, F. Schellauf, J. Scheel, A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: skin irritation, Regul. Tox. Pharmacol. 54 (2009) 188–196.
[65] J. Jaworska, S. Gabbert, T. Aldenberg, Towards optimization of chemical testing under REACH: a Bayesian network approach to Integrated Testing Strategies, Regul. Tox. Pharmacol. 57 (2010) 157–167.
[66] S. Hoffmann, A. Gallegos Saliner, G. Patlewicz, C. Eskes, V. Zuang, A.P. Worth, A feasibility study developing an integrated testing strategy assessing skin irritation potential of chemicals, Tox. Lett. 180 (2008) 9–20.
[67] L.L. Smith, Key challenges for toxicologists in the 21st century, Trends Pharmacol. Sci. 22 (2001) 281–285.
[68] T. Hartung, A toxicology for the 21st century: mapping the road ahead, Tox. Sci. 109 (2009) 18–23.
[69] K. Boekelheide, S.N. Campion, Toxicity in the 21st century: using the new toxicity testing paradigm to create a taxonomy of adverse effects, Tox. Sci. 114 (2010) 20–24.
[70] R. Corvi, H.-J. Ahr, S. Albertini, D.H. Blakey, L. Clerici, S. Coecke, G.R. Douglas, L. Gribaldo, J.P. Groten, B. Haase, K. Hamernik, T. Hartung, T. Inoue, I. Indans, D. Maurici, G. Orphanides, D. Rembges, S.-A. Sansone, J.R. Snape, E. Toda, W. Tong, J.H. Van Delft, B. Weis, L.M. Schechtman, Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use, Environ. Health Persp. 114 (2006) 420–429.
[71] T. Hartung, S. Bremer, S. Casati, S. Coecke, R. Corvi, S. Fortaner, L. Gribaldo, M. Halder, S. Hoffmann, A. Janusch-Roi, P. Prieto, E. Sabbioni, L. Scott, A. Worth, V. Zuang, A modular approach to the ECVAM principles on test validity, ATLA – Altern. Lab. Anim. 32 (2004) 467–472.
[72] T. Hartung, Food for thought ... on validation, ALTEX 24 (2007) 67–72.
[73] S. Hoffmann, L. Edler, I. Gardner, L. Gribaldo, T. Hartung, C. Klein, M. Liebsch, S. Sauerland, L. Schechtman, A. Stammati, E. Nikolaidis, Points of reference in validation – the report and recommendations of ECVAM workshop, ATLA – Altern. Lab. Anim. 36 (2008) 343–352.
[74] H.-J. Ahr, N. Alepee, S. Breier, C. Brekelmans, I. Cotgreave, L. Gribaldo, G. Dal Negro, O. De Silva, T. Hartung, A. Lacerda, B. Leblanc, C. Lecerf, J.P. Linge, S. Luhimies, I. Manou, G. Maxwell, K.K. Müller, W. Pape, K. Redhead, K.R. Schröder, D. Sladowski, K. van der Jagt, P. Vanparys, Barriers to validation. A report by European Partnership for Alternative Approaches to animal testing (EPAA) working group 5, ATLA 36 (2008) 459–464.
[75] M. Vinken, T. Doktorova, H. Ellinger-Ziegelbauer, H.J. Ahr, E. Lock, P. Carmichael, E. Roggen, J. van Delft, J. Kleinjans, J. Castell, R. Bort, T. Donato, M. Ryan, R. Corvi, H. Keun, T. Ebbels, T. Athersuch, S.A. Sansone, P. Rocca-Serra, R. Stierum, P. Jennings, W. Pfaller, H. Gmuender, T. Vanhaecke, V. Rogiers, The carcinoGENOMICS project: critical selection of model compounds for the development of omics-based in vitro carcinogenicity screening assays, Mutat. Res. 659 (2008) 202–210.
[76] S. Coecke, A.M. Goldberg, S. Allen, L. Buzanska, G. Calamandrei, K. Crofton, L. Hareng, T. Hartung, H. Knaut, P. Honegger, M. Jacobs, P. Lein, A. Li, W. Mundy, D. Owen, S. Schneider, E. Silbergeld, T. Reum, T. Trnovec, F. Monnet-Tschudi, A. Bal-Price, Incorporating in vitro alternative methods for developmental neurotoxicity into international hazard and risk assessment strategies, Environ. Health Persp. 115 (2007) 924–931.
[77] T. Hartung, Food for thought... on alternative methods for nanoparticle safety testing, ALTEX 27 (2010) 87–95.
[78] A.A. Bottini, P. Amcoff, T. Hartung, Food for thought... on globalization of alternative methods, ALTEX 24 (2007) 255–261.
[79] M.-J.W.A. Schiffelers, B.J. Blaauboer, J. Martje Fentener van Vlissingen, J. Kuil, R. Remie, J.W.G.M. Thuring, M.A. Vaal, C.F.M. Hendriksen, Factors stimulating or obstructing the implementation of the 3Rs in the regulatory process, ALTEX 24 (2007) 273–278.
[80] A.A. Bottini, N. Alepee, O. De Silva, T. Hartung, C. Hendriksen, J. Kuil, P. Pazos, B. Philips, C. Rhein, M.-J. Schiffelers, W. Stokes, A. Theobald, J.-M. Vidal, H. Van de Sandt, B. Blaauboer, Optimization of the post-validation process. The report and recommendations of ECVAM workshop 67, Altern. Lab. Anim. 36 (2008) 353–366.
[81] C. Eskes, V. De Moura Sa-Rocha, J. Nunes, O. Presgrave, D. de Carvalho, P. Masson, E. Rivera, S. Coecke, J. Kreysa, T. Hartung, Proposal for a Brazilian centre on alternative test methods, ALTEX 26 (2009) 265–268.
[82] O. Presgrave, C. Eskes, R. Presgrave, E. Alves, J.C. Freitas, C. Caldeira, I. Gimines, R. Silva, S. Nogueira, J. Nunes, E. Rivera, V. Sá-Rocha, S. Coecke, T. Hartung, Proposal to establish a Brazilian Center for Validation of Alternative Methods (BraCVAM), ALTEX 27 (2010) 159–161 (special issue).
[83] T. Hartung, H. Koeter, Food for thought ... on alternative methods for food safety testing, ALTEX 25 (2008) 259–264.
[84] T. Seidle, M.L. Stephens, Bringing toxicology into the 21st century: a global call to action, Toxicol. In Vitro 23 (2009) 1576–1579.