Food for Thought ... on the Real Success of 3R Approches

Marcel Leist, Suzanne Kadereit and Stefan Schildknecht
Doerenkamp-Zbinden Chair for alternative in vitro methods, Faculty of Natural Sciences and Mathematics, University of Konstanz, Germany

Introduction

Can the value of a scientific discipline be gauged? Where does the discipline stand? Is such a consideration important at all? Some may have doubts about the usefulness of asking such questions. Some will see the research into alternative methods as a categorical imperative. Others again will have a more differentiated view, a group that will comprise policy makers, those that provide funding and infrastructure for research, those that are responsible for our safety and that of our environment, and, last but not least, those that are about to choose their future field of scientific work. All these person groups will at some point use cost-benefit considerations and value balances. This type of thinking also applies largely to the general population of tax payers and voters. For this reason, some thoughts on the real value of 3R approaches (replace, reduce, refine, as defined by Russell and Burch, 1959, and the declaration of Bologna, 1999 (3R, 2004)) appear to be justified and necessary. On a qualitative level, the value of alternative methods is underscored by the fact that the EU funds an entire research institution (ECVAM) dedicated solely to the evaluation of alternative methods, and that major new EU legislation, such as REACH, has a strong focus on the use of alternative methods (REACH, 2006). Moreover, industry and the European Commission work together in a partnership for alternative approaches (EPAA, 2007), and universities are starting to establish departments dedicated to 3R research (Leist, 2006; Wendel, 2002). More quantitative approaches to describe the success of the new field make use of the statistics of animal use in the EU or its individual member states, or they count the number of OECD test guidelines that rely on alternative assays for safety evaluations. On this basis, progress of 3R is sometimes described as being relatively slow. In our opinion, such strategies to gauge the success of alternative approaches largely underestimate the real success of this emerging research field. Therefore, we will highlight in the following a number of conceptual errors that contribute to the underestimation of the value of 3R, and that are frequently encountered in public discussions.

Conceptual error I: Focus only on animal use for safety evaluations

The relatively standardised set of experiments in the area of toxicology is a particularly good target for alternative methods. In addition, the strong focus of 3R research on safety evaluations is justified by the particularly stressful experiments in this domain of animal experimentation. However, the main driver for research in this area is the particularly strong support of industry and governments – to a large part for economic reasons. Let's look at the larger picture. The overall use of experimental animals in the European Union was around 11 million animals (EU, 2005) in 2002 and about 12 million in 2005 (EU, 2007). Of these, only about 10% were used for toxicological studies in 2002 (Fig. 1), and this percentage dropped even further to around 8% in 2005. Therefore, it appears as too narrow an approach to evaluate the success of alternative methods only on the basis of substitution of OECD guidelines for toxicity testing (Gruber and Hartung, 2004). The problems of this approach become even more apparent when one takes into account that OECD test guidelines only exist for certain subdomains of safety testing (e.g. for safety testing of chemicals, but not for safety testing of drugs). This means that counting the number of accepted or validated tests in this area alone narrows down the overall success of 3R research to a small, single digit percentage of all experimental animals used (in total about 200,000 animals out of 12 million).

To obtain a better idea of the real success of 3R approaches one needs to expand one's view to other domains requiring animal testing (Gruber and Hartung, 2004). For instance, the number of animals used in education has dropped by 50 % (from 3.2 % of all animals in 2002 to 1.6% in 2005) due to strong efforts in this domain (see for example Dewhurst, 2006; Gruber and Dewhurst, 2004). A large area of animal use (15% of all animals in the EU) is the quality assurance and production of medicines. Here, new *in vitro* tests for pyrogenicity and the introduction of ELISA technology for batch control of vaccines represent great success stories (Hendriksen, 2006; Montag et al., 2007; Hartung, 2001; Rosskopf-Streicher et al., 2004; Hoffmann et al., 2005). In the largest domain of animal use, the statistical category "study of disease", consuming > 50 % of all animals (> six million/year), one should mention the ban of ascites mice and substitution by *in vitro* monoclonal antibody production methods (Kuhlmann et al., 1989), and the many *in vitro* systems used for instance for the study of borreliosis (Kröber and Guerin, 2007), angiogenesis (Bahramsoltani et al., 2006), genetic damage (Akyüz and Wiesmüller, 2003; Kreja et al., 2003), Parkinson's disease (Lotharius et al., 2005), etc. Moreover, many animal assays of hormonal activity have been substituted by more modern *in vitro* methods. Another noteworthy replacement assay is the *in vitro* colony forming assay, which is used to predict myelosuppression and to examine its mechanisms (Pessina et al., 2005;

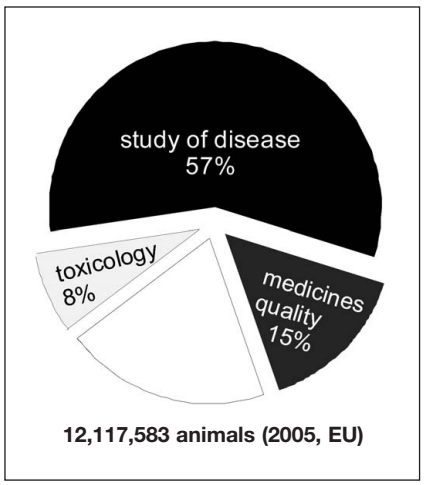

Fig. 1: Animal use in the European Union (EU) in 2005

ECVAM DB-ALM, 2006). Another large field of animal testing consuming hundred thousands of mice is consumer protection from contaminated shellfish. Here, one should mention as a good example Germany's and New Zealand's efforts to replace the mouse test for shell fish toxins by chemical analytical methods (Biosecurity New Zealand, 2007). Similarly, the substitution of the fish bioassay for waste water quality control by a fish embryo assay merits mentioning (Nagel, 2004).

Conceptual error II: Naïve use of statistics

The EU issues animal statistics every third year, and in addition most European countries issue annual statistics. Both are frequently used to judge the success of animal protection efforts, and, indirectly, of alternative methods. The use of such data is, however, complicated by the fact that the statistical rules are different in the member countries and that the statistical basis keeps changing. For instance, between 2002 and 2005 new member states joined the EU. Consequently, the increased numbers of animals used in 2005 do not indicate a lack of progress of alternative approaches but rather reflect a change of the statistical basis. With respect to the judgement of the success of 3R one also needs to take care to avoid systematic errors due to altered definitions of experimental animals (e.g. apparently increased animal numbers because of inclusion of organ removals from dead animals to animal experiments or counting embryos as animals).

What is more problematic than just the technical problems described above, is the conceptual error of using animal statistics to define the success of alternative methods. Let's assume a constant number of EU member countries and clear statistical rules for all. Would then constant numbers of animal experiments indicate that alternative methods have not been successful in a given period? No! Scientific research is expanding, and the number of scientists and publications is exploding. For instance, the research expenditure of the drug industry has risen 8-fold within 25 years (DiMasi et al., 2003). The number of publications in any biomedical field has often increased tenfold during that period. As an example, consider the figures for Alzheimer's disease and Parkinson's disease (Fig. 2). If animal consumption has nevertheless remained constant during that period, this should be regarded as a success and a large reduction in animal use relative to research intensity. Similar considerations should apply for the comparison of regions and countries. We would like to put forward the argument that really successful countries reduce the use of animals in relation to their research output, though not necessarily in absolute terms. In this context one may also consider the issue of globalisation (Bottini et al., 2007). We need to be careful and watchful with regard to outsourcing of animal experiments. Performing animal studies in non-EU countries would yield a cosmetic improvement of our animal statistics, but would not be beneficial for animals or an indicator of the success of 3R approaches!

Conceptual error III: Underestimation of 3R by measuring publication frequencies

It appears from the above that measuring the success of 3R methods requires a more differentiated approach than looking at animal statistics. One approach commonly used in science is to look at the number of relevant publications. This will certainly indicate a positive trend for 3R research. However, the value of alternative methods is likely to be underestimated by this approach. Let's for example take a closer look at drug discovery.

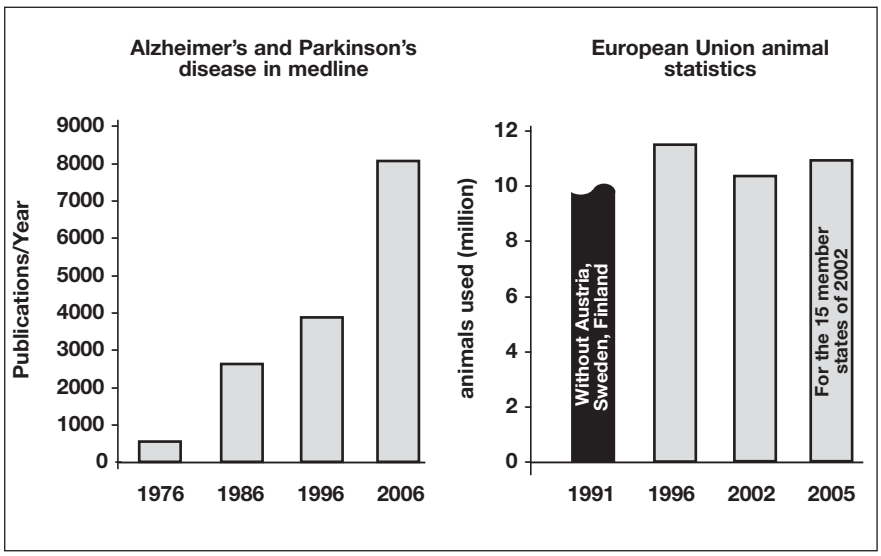

Fig. 2: Constant animal numbers vs. growing science
The number of publications in nearly all fields of science is steeply growing (for example the publications on Alzheimer's disease and Parkinson's disease, upper panel). The number of animals used in the EU has remained relatively constant over the last 15 years (lower panel; the 2005 bar contains only the animal number for the 15 member states that were part of the EU in 2002).

More than 99% of newly synthesised compounds are excluded at early stages, and little of the data obtained from such compounds will ever be published. This is mostly explained by disinterest of the companies involved. Moreover, there is a general under-representation bias in the literature concerning publication of negative data. A lot of these unpublished results are derived from *in vitro* methods, and the extensive use of such methods will thus never become known to the public (Fig. 3). In addition, simple *in vitro* safety screens, such as the human ether-a-gogo related receptor (HERG) assay or the Ames assay, eliminate compounds from drug discovery and development that will generally never be published (Fig. 3). Once compounds advance further, a positive publication bias for *in vivo* studies further contributes to the underestimation of the use of alternative methods. It is still commonplace that a single animal experiment with negative data can be published. The present status quo makes such a publication on the basis of alternative methods unthinkable. In the latter case, one takes for granted that *in vitro* data presented in a publication have been obtained at least three times in different experiments, and additionally varied in parameters such as concentration and time.

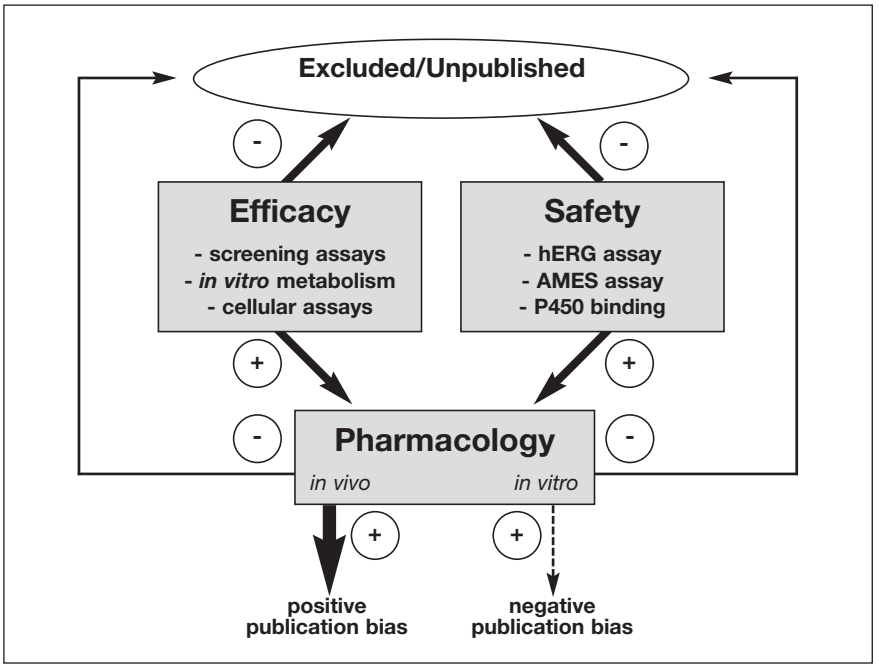

Fig. 3: Publications do not reflect the use of alternative assays in drug discovery
In early phases of drug discovery, many *in vitro* assays are used (Efficacy box and Safety box). Unsuitable compounds (minus sign) are excluded and usually not published. Suitable compounds (plus sign) are tested pharmacologically and, if found to be unsuitable, excluded. For compounds that are further promoted, *in vivo* pharmacological data have a higher chance of being published than data form alternative assays.

Conceptual error IV: Assumption of 1:1 substitutions

Classical toxicology developed a system of safety testing based on the opportunities and limitations of the animal as the model system. This form of testing shaped and determined the current set of rules for safety evaluations, which may be called the "animal game". Such rules include the classification of hazard domains (e.g. mutagenicity, corrosion, reproductive toxicity), but also the way doses are selected and extrapolated. Presently, moving outside this set of rules and way of thinking will result in "failure" or being "caught cheating". In order to keep to these rules, and in an attempt to follow the same classifications, 3R research has often attempted to replace certain animal experiments one-for-one (1:1) with an alternative method. For instance, the *in vivo* phototoxicity test was replaced with the fibroblast phototoxicity assay (OECD Test Guideline 432, 2007).

However, if one looks at hazard domains such as reproductive toxicity testing or evaluation of the sensitisation potential, most will quickly agree that a 1:1 substitution of the currently used assays by an *in vitro* assay is highly unlikely. Nevertheless, a lot of progress has been made in the development of 3R methods for the two domains just mentioned. However, these assays measure individual steps of a biological process. For instance, in the process of sensitisation, the binding of hapten to protein or the activation of dendritic cells is used as the readout; and in the process of developmental toxicity, binding to steroid receptors, toxicity to spermatocytes or disturbed differentiation of embryonic stem cells are analysed separately. Such individual endpoints and readouts will need to be combined to integrated testing strategies, which may eventually cover the entire biological process (Fig. 4). None of the individual assays will be comparable to the original animal experiment, but each may have a high value within a test strategy combining different *in vitro* methods or being composed of *in vitro* methods and some reduced and refined animal experiments (Combes, 2007). Thus, looking at 1:1 substitutions only leads to an underestimation of the progress of 3R methods. What makes matters worse is that this way of thinking not only forces the rules of the "animal game" onto 3R methods, but also contributes to preventing their development and implementation. As long as animal experiments are used as the gold standard for the each alternative method, such methods will always have the limitations inherent to the animal experiment (Fig. 4), and entirely different and innovative approaches have no chance of passing the validation process (Hartung, 2007). Ironically, the situation of underestimation is even worse for areas where the animal gold standards are less established. For example, the testing of biologics or nanoparticles are relatively recent domains, and animal testing is less standardised here than for classical

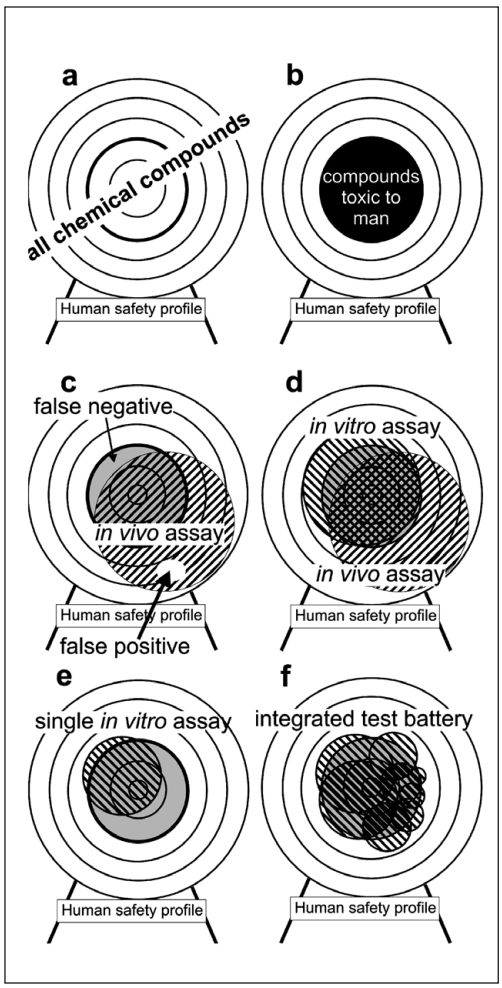

Fig. 4: Overlap of assay results and reality
a) All chemical compounds coming into contact with man are represented as an archery target.
b) The compounds toxic to man (grey) are represented as those in the middle of the target. Good assays require a high overlap with that area.
c) An example of a typical *in vivo* safety assay is shown. The area (hatched) covers most of the toxic compounds, but there are some false negatives (upper left area) and many false positives (lower right area).
d) A theoretical example of an *in vivo* and *in vitro* assay (for the same safety domain; e.g. pyrogenicity) is shown. Here, the *in vitro* assay has less false positives and less false negatives than the *in vivo* assay. Nevertheless, there is a poor overlap between both assays. If the *in vivo* assay was regarded as gold standard, the in vitro assay would fail in validation.
e) Shows a theoretical example of an alternative assay as part of an integrated test strategy (e.g. spermatotoxicity as part of developmental toxicity). Taken alone it has poor safety characteristics (many false negatives).
f) Shows the same assay as in e) as part of an integrated test battery with optimal safety characteristics (no false negatives, few false positives)

small molecules. Here, alternative methods can and do already now valuable information for the overall safety evaluation, but, as there is no clear animal gold standard to be substituted, the success of 3R methods is hard to measure and therefore underestimated by many classical statistical methods.

Conceptual error V: Monodimensional focus on replacement

Measuring 3R success in terms of replaced (1R) animal experiments or OECD test guidelines tends to underestimate the success of the full integrated 3R approach in terms of reduction of animal suffering (Fig. 5). Although there is no doubt that the ultimate goal of the approach is replacement, it appears to us as a conceptual error to neglect the successes of the other two domains, which have the potential to reduce suffering in a huge number of animals. Refinement approaches do not necessarily alter the number of animals used, but much less suffering is inflicted. As "non-replacement approaches" are sometimes forgotten, we would like to mention three examples. The most frequently cited example for refinement is the fully validated and regulatory-accepted (OECD Test Guideline 429, 2002) local lymph node assay, which replaces the Bühler guinea pig maximisation test for sensitisation potential of chemicals. A prominent example for the reduction approach is the group of new assays for the testing of acute toxicity (OECD Test Guidelines TG420, 423, 425), which, since 20. December 2002, replace the old LD_{50} guideline (TG401) and reduce the number of animals needed by >60-70%. Less well known are many approaches to more intelligent testing strategies, which have a huge potential of animal reduction. An illustrative example is the test strategy for acute toxicity to aquatic organisms (daphnia, algae, fish), where about 60% of fish are spared by the new sequential test approach with fish being used last (ESAC, 2006; Jeram, 2005). Reduction approaches are unspectacular in terms of publications (because they ironically eliminate the reason for publication), but they are highly effective, and more importantly, fast in their implementation and effect. One cannot value their effect in big programmes like REACH high enough (e.g. >1 million rats saved just by an altered approach to acute toxicity testing). In the large scale testing of already marketed chemicals in Europe (according to REACH), the number of animals used will depend highly on the extent of the use of read-across approaches and the intelligent use of information, and millions of animals can be rescued by intelligent test strategies. The following example from the field of skin corrosion testing illustrates the power of the approach: the development and validation of the replacement method CORROSITEX (Corrositex, 2007) lasted 10 years. This method mostly detects compounds with extreme pH. Therefore, an alternative approach would be to apply a test strategy that automatically (without *in vivo* or *in vitro* testing) classifies compounds with extreme pH as corrosive. Such an "intelligent" test strategy (as part of the OECD test guideline) prevents extreme suffering immediately, i.e. ten years of development time and additional distress are saved.

Conceptual error VI: Monodimensional views on value domains

Like each science discipline, 3R research defines its overall value from various dimensions comprising "quality of science", "applicability" and "ethics" (Fig. 6). In the particularly multidisciplinary 3R field, different groups are strongly focused on one of the dimensions and frequently neglect the others, with the effect of an underestimation of the overall success of the discipline, and

a failure of 3R research to reach its full potential. The scientific dimension focuses on the relevance and coherence of methods, but has also an important role in identifying and addressing new challenges and constantly inventing new types of solutions. Not to be forgotten here, is the inherent pleasure of good science in itself and the fascination for this extremely interdisciplinary field. Awareness of this point may help to attract more and more established researchers from other disciplines as well as students looking for an interesting career to the field. The application dimension addresses factors like cost, unmet need, assay performance, and definition of standards. Stronger awareness of the value of this domain would lead to better infrastructure for such work, which in many countries does not exist at all and in most others is only weakly developed. An approach leading in the right direction is the development of a European reference laboratory (COmmunity RefeREnce Laboratory for Alternative Testing, CORRELATE, 2007) at ECVAM, but here also scope and especially funding appear minute in relation to the huge task and value potential. A third dimension comprises the ethical issues of the field, like the questions of dignity of animals, the balancing of pain and potential benefits from animal experiments, questions relating to the proper value of animals independent of their use to humans, the acceptable risk for humans in relation to economical factors, the relative differences between different animal species like primates vs. companion animals vs. rodents, etc. Strong additional value can be gained from a detailed consideration of these issues instead of a dogmatic and simplified approach (see chapter below)

Conceptual error VII: numbers of experimental animals correlate with animal suffering

This last conceptual error discussed here can lead to underestimation OR overestimation of the value of 3R methods. Moreover, this chapter, together with the chapters below, may form the basis for further discussion in this or another forum.

In many countries, and also in the EU, the animal statistics do not give information on animal suffering. The animal experiments included span an incredibly broad range of pain and distress, which is not being accounted for at all. A further grey zone is the breeding of genetically modified animals, which in most cases is not registered in animal statistics (as opposed to the generation of such animals). Whether the breeding of genetically modified animals constitutes a particular stress is still not given sufficient consideration, and judgment of that matter sometimes is beyond the competence of those who have created the animals (Mertens and Rulicke, 2007; Sauer et al., 2006).

If one tries to dig below the surface of statistical summaries, one experiences how hard, or often impossible, it is to extract much information from the numbers. For instance, in 1999 Greece did not use a single fish. In 2002, over half a million were used, and this number nearly doubled until 2005, the number now being higher than all fish used by the other 24 EU members together. There is no simple way to find an explanation for that. As another example, consider an apparently simple question: "How can we improve the situation of companion animals?" - i.e. how can we find methods to reduce the use of e.g. 24,000 dogs per year as experimental animals. To approach this problem, we need to find an answer to the question "Where are dogs used as experimental animals?" There is no way to answer this question from European or national statistics. We can see that 500 dogs are used for pesticide testing, 6,000 in biomedical research and about 13,000 for drug safety testing, but no more detail than that. This situation has a large impact on the question "Is a particular assay successful in improving the situation of dogs?" The examples illustrate that the value of alternative methods is in part so hard to judge because animal statistics are so poor and non-transparent.

In order to understand another statistical shortcoming, let's look into another situation, the "Draize eye irritation test", which so urgently requires alternatives (with the notable exception of France, where the HET-CAM assay is fully accepted). A number of assays (e.g. isolated bovine, chicken or rabbit eyes) can be used as prior filter assays, and in case of positive findings, data will be accepted by EU and member state authorities. Accord-

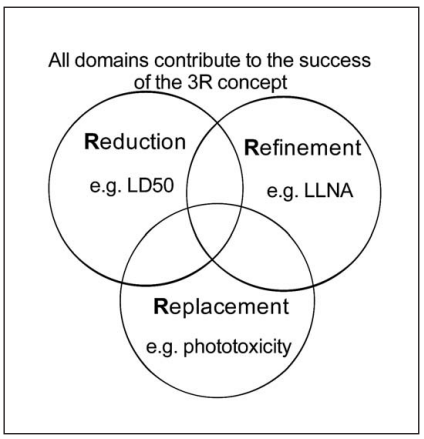

Fig. 5: All 3R domains reduce suffering and pain
For example, the acute lethal dose (LD_{50}) assay is a reduction assay, the local lymph node assay (LLNA) is an accepted refinement method and phototoxicity is assessed by an *in vitro* assay using fibroblasts; another relevant example for this latter domain would be "skin corrosion" where the *in vivo* test guidline has been replaced by *in vitro* testing.

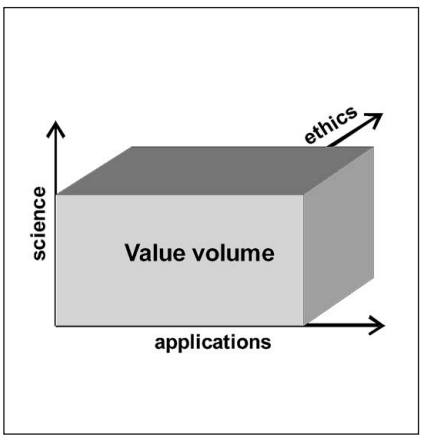

Fig. 6: The overall value of the 3R approach is defined by the product of different dimensions

ing to the past 25 years' statistics, about 7% of chemicals have been classified as eye irritants. This would mean that alternative assays can reduce the number of Draize tests by about 7%. Is this impressive or rather negligible? We need not give an answer, as the question itself appears to us to be wrong. We would rather know how much suffering has been reduced than getting information on the numbers of Draize tests performed. After the introduction of alternative tests, still a high number of animal eye irritation tests must be performed, but now mostly with compounds that are innocuous. Severe irritants are filtered out by *in vitro* testing before they are used in an often painful animal test. This demonstrates clearly that a large success domain of alternative assays is at the moment their filtering capability, keeping the most noxious and pain-inducing compounds away from animals. This is, however, not reflected in animal use statistics.

Last, we want to briefly touch on an issue that is often neglected but should not be forgotten: underestimating the effect of alternative assays on research throughput. Some areas of drug discovery are still limited in size because animal experiments present a serious bottleneck. A good *in vitro* system may increase the number of companies interested in the field, and the number of compounds screened in each company. In the end, this may require a much larger number of animals in pharmacology, safety and quality evaluations than ever before. But in parallel more alternative methods will be developed. So, we do not know at the moment how the ethical value of such alternative methods should be judged.

How else may the value of 3R approaches be judged?

If current statistics do not help us, how else can one obtain an appropriate estimate of the value of 3R methods? It appears important to us to find an alternative for the negative definition via animal statistics, and to rather give the field a positive basis. Following this line of thought, one may for instance ask: "How big is the output of alternative methods?", "How much did 3R methods and approaches

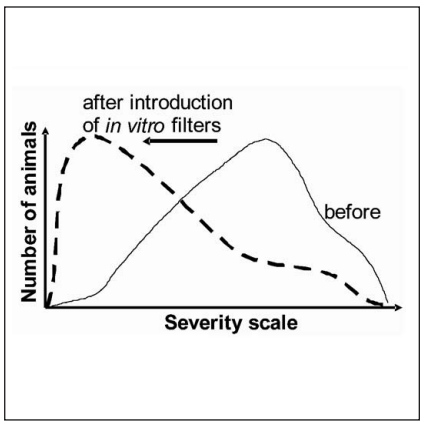

Fig. 7: Alternative assays as filters can shift the severity distribution of animal experiments
The solid line shows a hypothetical severity distribution of animal experiments before introduction of alternative methods to prefilter compounds used on animals. The dashed line shows the altered distribution after introduction of assays which are accepted as positive filters, i.e. compounds positive in these assay (the most harmful compounds) will not be tested on animals.

contribute to the overall output (research, safety evaluation, disease mechanisms)?" or, "Would the end result have been achieved without *in vitro* methods, and how well?" One may also look at the increasing number of people working in the area and finding career opportunities as well as scientific challenges. When doing so, it is essential to look at all application domains, and to demonstrate the broad variety of the field as for example the Swiss 3R Foundation or the Doerenkamp-Zbinden Foundation are doing (Maier und Wick, 2007; DZF, 2007).

One may also judge 3R on a medical or scientific background, and ask, "How much better have *in vitro* models become over time in predicting human safety or human pharmacology?" In the same vein, one may test how much better our mechanistic understanding of important processes in toxicology and medical research has become. Although this aspect appears particularly important, it is important to point out that not all *in vitro* research is 3R research, and *vice versa*. Another non-negligible value of 3R comes from the time and money that has been saved by alternative methods. We have seen, also in many other fields that economic considerations cannot be completely uncoupled from ethical, ecological or moral considerations, and actually can be a valuable driving force. On the other hand, the ethical dimension itself is not only a driver of the field, but is also suitable as an alternative value basis: newer methods of evaluation certainly need to include the extent to which distress and suffering were reduced, not only in statistical averages, but also in many individual case stories.

Ways forward

A broader value basis is an ideal platform for a large variety of approaches to how 3R research can further improve its perception from the outside as well as from within the field. Notably, altering the perception is not only a cosmetic effort, but is important for bringing the field forward conceptually and technically. Such progress will then in turn have a major impact on overlapping fields such as toxicology, stem cell research, pharmacology and disease biology. Ways forward may be grouped in action packages addressing different issues.

Package one would for instance comprise efforts towards the improvement of animal statistics and their use. Statistics and databases should be more transparent, more traceable, more open to access and more suitable for the easy use of the data, and they should certainly contain measures of stress and suffering, together with a rationale for the animal experiments. Approaches based on simple animal counting, as often used for the comparison of different REACH scenarios by all parties involved, appear cynical and distract from the real problems.

Package two may address publication biases and lack of important information from industry. Incentives could be given for case studies and publications from industry. A good example for the usefulness of such publications is a study at Organon on the impact of 3R methods (Verbost et al., 2007). "Publication" is defined here in a very wide sense, also including the feeding of data into broadly-accessible databases, allowing and improving *in vivo-in vitro* comparisons and giving information on the availability and practical

application of alternative methods. This process may be considerably enhanced by the creation of larger national or international centres dedicated to the creation of infrastructure for development of alternative methods. Large organisations like EPAA have already started to work on this package.

Package three would further extend into this direction by definitely focusing on a number of centres that would be responsible for reference compounds and reference compound databases. This effort may appear trivial, but it shows how underdeveloped the infrastructure of the field still is, and where immediate action should be taken. Imagine you are looking for a list of tool compounds, for example to validate a model of developmental neurotoxicity, and you wonder why you simply cannot find this in the literature. The explanation is simple. Good, validated lists of such compounds are not available, and when it comes to the compounds themselves, the problem is even larger. A lot could be achieved here with relatively little effort, and some of these attempts have already been initiated, for instance by ECVAM.

Package four contains actions of a different nature and is focused on the unity and maintaining the core strengths of the field. As the research field grows, it will be important to keep the different value domains and R domains in good contact and improve the interaction with one another, something which is amongst the great merits e.g. of the Linz congresses on alternatives to animal testing, organised by the Centre for Alternative and Complementary Methods to Animal Testing (ZET). A huge challenge is also the delimitation towards other disciplines. On one hand, a certain self control will be essential to keep the definitions of alternative research and pure mechanistic or *in vitro* research from being confused. On the other hand, one has to take great care not to commit the mistake of defining the field too narrowly and of excluding neighbouring disciplines. Such behaviour bears the risk of impoverishing the field as experienced already e.g. by toxicologists (Lotti and Nicotera, 2002).

Package five comprises all measures pushing for a fair comparison of alternative methods and classical animal experiments. Especially in the area of toxicology, this means similar validation requirements for the animal experiments as for alternative methods and moving away from animal experiments as gold standards. One should ask for a stronger focus on relevance-based and mechanistically-characterised animal models that are validated under stringent criteria, resembling those applied in evidence-based medicine. A new movement in this direction calls itself "evidence-based toxicology" (EBT) (Guzelian et al., 2005; Hoffmann and Hartung, 2006). This type of thinking would be a way forward to give alternative methods their proper place and value.

Package six takes this thought consequently to its end - to an end that would mean the end of dominance of the "animal game". In a landmark document, "Toxicity testing in the 21st century", the National Research Council of the USA (NRC, 2007) defined a vision in which the "3R game" plays a major role. The dominance of the sets of rules of animal and 3R game are being reversed. Instead of looking at animals as a black box, and then trying to find out what happened in cases of toxicity, it is argued for a bottom-up approach of mechanistic understanding. Mechanistic research with alternative methods can define the essential pathways that are common to many forms of toxicity. Toxic compounds would then initially be characterised by the pathways they trigger and only at later stages of integration would animal experiments be used to complement this information and close remaining gaps. The strong interdisciplinarity and imbedding of the 3R field into biomedical research is an ideal basis for this approach (Lotti and Nicotera, 2002). In this new form of mechanism-based toxicology 3R research develops to achieve its full value, for animals, for science and for man.

References

3R Foundation (2004). http://3r-training.tierversuch.ch/en/module_3r/3r-bologna_declaration/the_3rs.html

Akyüz, N., Wiesmüller, L. (2003). Proof of principle: detection of genotoxicity by a fluorescence-based recombination test in mammalian cells. *ALTEX 20(2),* 77-84.

Bahramsoltani, M., Kaessmeyer, S., Plendl, J. (2006). A novel quantitative in vitro model of angiogenesis. *ALTEX 23 Suppl.*, 265-269.

Biosecurity New Zealand (2007). http://www.biosecurity.govt.nz/publications/biosecurity-magazine/issue-75/shellfish

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

Combes, R. (2007). Developing, validating and using test batteries and tiered (hierarchical) testing schemes. *ATLA 35,* 375-378.

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

Corrositex (2007). http://corrositex.mbresearchlabs.com/

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

Dewhurst, D. (2006). Computer-based alternatives in higher education--past, present and future. *ALTEX 23(3)*, 197-201.

DiMasi, J. A., Hansen, R. W., Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *J. Health Econ. 22(2)*, 151-185.

Doerenkamp-Zbinden Foundation (DZF) (2007). http://www.doerenkamp.ch/en/default.html?id=5

EPAA (2006). European Partnership to Promote Alternative Approaches to Animal Testing http://ec.europa.eu/enterprise/epaa/conf_2006_presentationvdgraaf_unilever.pdf

ESAC (2006). http://ecvam.jrc.it/publication/ESAC24_statement_UTC_step_down_approach_1.pdf

EU (2005). Fourth Report on the Statistics on the Number of Animals used for Experimental and other Scientific Purposes in the Member States of the European Union. http://ec.europa.eu/environment/chemicals/lab_animals/pdf/com_2005_7_en.pdf

EU (2007). Fifth Report on the Statistics on the Number of Animals used for Experimental and other Scientific Purposes in the Member States of the European Union. http://ec.europa.eu/environment/chemicals/lab_animals/pdf/5th_stat_rep_lab_animals_en.pdf

Gruber, F. P., Dewhurst, D. G. (2004). Alter-

natives to animal experimentation in biomedical education. *ALTEX 21 Suppl 1*, 33-48.

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

Guzelian, P. S., Victoroff, M. S., Halmes, N. C. et al. (2005). Evidence-based toxicology: a comprehensive framework for causation. *Hum. Exp. Toxicol. 24(4)*, 161-201.

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

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

Hendriksen, C. F. (2006). Towards eliminating the use of animals for regulatory required vaccine quality control. *ALTEX 23(3)*, 187-190.

Hoffmann, S., Hartung, T. (2006). Toward an evidence-based toxicology. *Hum. Exp. Toxicol. 25(9)*, 497-513.

Hoffmann, S., Peterbauer, A., Schindler, S. et al. (2005). International validation of novel pyrogen tests based on human monocytoid cells. *J. Immunol. Methods 298*, 161-173.

Jeram, S., Sintes, J. M., Halder, M. et al. (2005). A strategy to reduce the use of fish in acute ecotoxicity testing of new chemical substances notified in the European Union. *Regul. Toxicol. Pharmacol. 42(2)*, 218-224.

Kreja, L., Schochat, T., Finking, G. (2003). Development of an in vitro model to study oxidative DNA damage in human coronary artery endothelial cells. *ALTEX 20(2)*, 71-76.

Kröber, T. and Guerin, P. (2007). The Tick Blood Meal: From a Living Animal or from a Silicone Membrane? *ALTEX (Special issue)*, 39-41.

Kuhlmann, I., Kurth, W. und Ruhdel, I. (1989). [Monoclonal antibodies: in vivo- and in vitro-production in laboratory scale with consideration of the legal aspects of animal protection]. *ALTEX 6(2)*, 12-26.

Leist, M. (2006). What can a chair on alternatives to animal experimentation effectuate? *ALTEX 23(3)*, 211-213.

Lotharius, J., Falsig, J., van Beek, J. et al. (2005). Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed lineage kinase pathway. *J. Neurosci. 25*, 6329-6342.

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

Maier, P. und Wick, H. (2007). Editorial. *ALTEX 24 (Special issue)*, 1.

Mertens, C. and Rulicke, T. (2007). Welfare assessment and phenotypic characterization of transgenic mice. *ALTEX 24 (Special issue)*, 46-48.

Montag, T., Spreitzer, I., Löschner, B. et al. (2007). Safety testing of cell-based medicinal products: opportunities for the monocytes activation test for pyrogens. *ALTEX 24(2)*, 81-89.

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

NRC (2007). Committee on Toxicity Testing and Assessment of Environmental Agents, National Research Council: Toxicity Testing in the 21st Century: A Vision and a Strategy (196). The national academies press; http://www.nap.edu/catalog.php?record_id=11970

OECD (2002). OECD Guidelines for the Testing of Chemicals Test No. 429: Skin Sensitisation: Local Lymph Node Assay. http://www.ingentaconnect.com/content/oecd/1607310x/2001/00000001/00000004/9742901e

OECD (2007). OECD Guidelines for the Testing of Chemicals Test No. 432: In Vitro 3T3 NRU Phototoxicity Test. http://oberon.sourceoecd.org/vl=1642044/cl=21/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s33/p1

Pessina, A., Malerba, I., Gribaldo, L. (2005). Hematotoxicity testing by cell clonogenic assay in drug development and preclinical trials. *Curr. Pharm. Des. 11(8),* 1055-1065.

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

Rosskopf-Streicher, U., Volkers, P., Noeske, K., Werner, E. (2004). Quality assurance of C. perfringens epsilon toxoid vaccines--ELISA versus mouse neutralisation test. *ALTEX 21 Suppl. 3*, 65-69.

Russell, W. M. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique* (328). London: Methuen.

Sauer, U. G., Kolar, R., Rusche, B. (2006). The use of transgenic animals in biomedical research in Germany. Part 2: Ethical evaluation of the use of transgenic animals in biomedical research and perspectives for the changeover in research to research animal-free methods. *ALTEX 23(1)*, 3-16.

Verbost, P. M., van der Valk, J., Hendriksen, C. F. (2007). Effects of the introduction of in vitro assays on the use of experimental animals in pharmacological research. *ATLA 35(2)*, 223-228.

Wendel, A. (2002). Do we need a "Chair of alternative methods", and where? *ALTEX 19(2)*, 64-68.

Acknowledgement

We gratefully acknowledge the valuable ideas, information and food for thought received from Thomas Hartung and Franz P. Gruber during countless good discussions.

Correspondence to

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