

Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH)

W. Lilienblum · W. Dekant · H. Foth · T. Gebel ·
J. G. Hengstler · R. Kahl · P.-J. Kramer ·
H. Schweinfurth · K.-M. Wollin

Received: 14 December 2007 / Accepted: 7 January 2008
© Springer-Verlag 2008

Abstract During the last two decades, substantial efforts have been made towards the development and international acceptance of alternative methods to safety studies using laboratory animals. In the EU, challenging timelines for phasing out of many standard tests using laboratory animals were established in the seventh Amending Directive 2003/15/EC to Cosmetics Directive 76/768/EEC. In continuation of this policy, the new European Chemicals Legislation (REACH) favours alternative methods to conventional in vivo testing, if validated and appropriate. Even alternative methods in the status of prevalidation or validation, but without scientific or regulatory acceptance may be used under certain conditions. Considerable progress in the establishment of alternative methods has been made in some fields, in particular with respect to methods predicting local toxic effects and genotoxicity. In more complex

important fields of safety and risk assessment such as systemic single and repeated dose toxicity, toxicokinetics, sensitisation, reproductive toxicity and carcinogenicity, it is expected that the development and validation of in silico methods, testing batteries (in vitro and in silico) and tiered testing systems will have to overcome many scientific and regulatory obstacles which makes it extremely difficult to predict the outcome and the time needed. The main reasons are the complexity and limited knowledge of the biological processes involved on one hand and the long time frame until validation and regulatory acceptance of an alternative method on the other. New approaches in safety testing and evaluation using “Integrated Testing Strategies” (ITS) (including combinations of existing data, the use of chemical categories/grouping, in vitro tests and QSAR) that have not been validated or not gained wide acceptance in the

W. Lilienblum (✉)
Dr. Lilienblum Consulting Toxicology LiCoTox,
30966 Hemmingen/Han, Germany
e-mail: werner.lilienblum@t-online.de

W. Dekant
Institute of Toxicology, University of Würzburg,
Versbacher Straße 9, 97078 Würzburg, Germany

H. Foth
Institute of Environmental Toxicology,
University of Halle, 06097 Halle/Saale, Germany

T. Gebel
Federal Institute for Occupational Safety and Health,
Friedrich-Henkel-Weg 1-25, 44149 Dortmund, Germany

J. G. Hengstler
Leibniz Research Centre for Working Environment and Human
Factors (IfADo), University of Dortmund, Ardeystrasse 67,
44139 Dortmund, Germany

R. Kahl
Institute of Toxicology, Heinrich-Heine-University,
PO Box 101007, 40001 Düsseldorf, Germany

P.-J. Kramer
Institute of Toxicology, Merck KGaA,
64271 Darmstadt, Germany

H. Schweinfurth
Nonclinical Drug Safety, Bayer Schering Pharma AG,
13342 Berlin, Germany

K.-M. Wollin
Lower Saxony Governmental Institute of Public Health,
Roesebeckstraße 4-6, 30449 Hannover, Germany

scientific community and by regulatory authorities will need a thorough justification of their appropriateness for a given purpose. This requires the availability of knowledge and experience of experts in toxicology. The challenging deadlines for phasing out of *in vivo* tests in the Cosmetics Amending Directive 2003/15/EC appear unrealistic. Likewise, we expect that the application of validated alternative methods will only have a small or moderate impact on the reduction of *in vivo* tests under the regimen of REACH, provided that at least the same level of protection of human health as in the past is envisaged. As a consequence, under safety aspects, it appears wise to consider established *in vivo* tests to be indispensable as basic tools for hazard and risk assessment with respect to systemic single and repeated dose toxicity, sensitisation, carcinogenicity and reproductive toxicity, especially regarding quantitative aspects of risk assessment such as NOAELs, LOAELs and health-related limit values derived from them. Based on the overall evaluation in this review, the authors are of the opinion that in the short- and mid-term, the strategy of the development of alternative methods should be more directed towards the refinement or reduction of *in vivo* tests. The lessons learnt during these efforts will provide a substantial contribution towards the replacement initiatives in the long-term.

Keywords Risk assessment · Chemicals · REACH · Alternative method · *In vitro* · *In vivo* · *In silico* · Validation · Integrated testing strategy

Introduction

Since the implementation of the Animal Welfare Guideline 86/609/EC in 1986, it is the declared policy of the EU institutions to support the development and use of alternative methods, i.e. “any method that can be used to replace, reduce or refine the use of animal experiments in biomedical research, testing or education”. The “alternatives” concept is attributed to Russell and Burch (1959), who defined three types of alternatives, the replacement, reduction or refinement of animal tests, the so called RRR (or 3R) Principles (for details, see for instance Balls 2002; Worth and Balls 2002a). In the frame of safety testing and safety assessment, the definition of alternative methods includes “testing methods” (e.g. *in vitro*, *ex vivo* or reduced/refined methods *in vivo*) as well as “non-testing methods” such as the use of expert systems. In particular, for the purposes of REACH, a consortium of the industry, EU institutions and Member State bodies under the lead of CEFIC¹ have com-

pleted conventional and alternative methods for the testing of chemicals and developed non-testing and testing strategies for minimizing the volume of testing and thereby reducing costs and the use of animals (ECB 2005, 2007, <http://ecb.jrc.it/reach/rip>). Under REACH, non-testing methods also include the adequate use of existing data of a substance or other considerations such as the chemical category or chemical analogue (read-across) approach that may contribute to reduce or avoid testing *in vivo* (OECD 2007a, b).² Partly the definitions for testing and non-testing methods differ. For instance, although (quantitative) structure–activity relationships [(Q)SARs] may be considered as testing methods, they may also be considered as non-testing methods, such as in the Draft Technical Guidance Document (TGD) which has been developed in the frame of REACH (ECB 2005, 2007). Another distinction can be made between reduced/refined animal tests and non-animal methods. The main focus of this paper will be on non-animal testing methods, since from our view it is of utmost importance that these methods are realistically assessed in respect to their use in hazard identification and risk assessment of chemicals in terms of human health.

Alternative methods to animal testing: an important perspective in current regulatory toxicology

Recent regulatory developments in the EU

In the EU, the promotion of alternative methods is increasingly favoured at the expense of conventional animal testing. Similar developments of the promotion of alternative methods can be observed in other industrialised regions such as USA and Japan. They are co-ordinated by OECD at the international level (e. g. OECD 1990, 1996, 2005a). As a consequence of the EU Directive 86/609/EEC, in 1991 the European Centre for the Validation of Alternative Methods (ECVAM) was founded which has become a unit of the Joint Research Centre of the EU Commission in Ispra, Italy, since 1992. Meanwhile, the EU has invested more than 200 million dollars into the development of alternative methods by funding respective research (Zuang and Hartung 2005) apart from the funding at the national level by the Member States themselves. On the international level, a close cooperation between ECVAM and other similar institutions exists such as the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxi-

¹ European Chemical Industry Council

² Another term for chemical category approach is *grouping* which is the favoured term in the frame of REACH (Art. 13, Annex XI; see also ECB 2005, 2007).

Table 1 Time schedule of the REACH Regulation relevant for the implementation of alternative test methods in safety testing

1 June 2007	Regulation enters into force
1 June 2008	European Chemicals Agency officially starts working
1 December 2008	Deadline for the pre-registration of phase-in substances (Art. 28)
Deadlines for the registration of phase-in substances (Art. 23):	
1 December 2010	Classified as carcinogenic, mutagenic or toxic to reproduction, category 1 or 2, 1,000 tons or more manufactured or imported per year;
1 June 2013	100 tons or more manufactured or imported per year;
1 June 2018	1 ton or more manufactured or imported per year,
Draft decisions of the Agency on additional testing of substances (Art. 40(3) and 43(2)):	
1 December 2012	Classified as carcinogenic, mutagenic or toxic to reproduction, category 1 or 2, 1,000 tons or more manufactured or imported per year;
1 December 2016	100 tons or more manufactured or imported per year;
1 December 2018	1 ton or more manufactured or imported per year.

ological Methods (NICEATM) but increasingly also with the OECD (Abbot 2005; IHCP 2004, 2005; Zuang and Hartung 2005; OECD 2005a).

Following the directive 86/609/EEC in 1986, (EU 1986), the EU institutions adopted in 2003 the seventh Amendment Directive 2003/15/EC on cosmetics, which aims at a step-wise phasing out of experimental animal safety tests of cosmetics by 2008 and 2013, respectively (EU 2003). Similarly to the Amendment Directive 2003/15/EC on cosmetics, also the REACH Regulation on chemicals adopted by the European Council and the European Parliament in December 2006 aims at avoiding animal testing and taking preference of alternative methods to animal testing as far as possible (EU 2006). The aim of Registration, Evaluation and Authorisation of Chemicals (REACH) is to systematically evaluate the risks, to human health and environment, of more than 30,000 chemical substances that are produced or imported in the EU in volumes of 1 ton or more per year. Bands of data requirements are 1, 10, 100 and 1,000 tons per year. Article 25(1) of REACH Regulation reads: “In order to avoid unnecessary animal testing, testing on vertebrate animals for the purpose of this Regulation shall be undertaken only as a last resort.” Furthermore, Annexes VIII–XI of the Regulation contain exceptions from the volume of testing under certain circumstances and in particular from animal testing which are termed derogation or waiving (for a discussion, see for instance ECB 2005, 2007; Greim et al. 2006). The time schedule of REACH as far as it is relevant for the purposes of this paper is shown in Table 1, while the toxicological endpoints for which data are required under REACH are listed in Table 2. In current toxicological testing, most of these endpoints are addressed by animal experiments carried out according to OECD guidelines.

There is much activity to make the indeterminate provisions of European legislation concerning animal experimentation more operational. Recently, by the implementation of three large Integrated Projects, ReProTect, A-Cute-Tox and Sens-it-iv, a new dimension of tailored

development of alternative methods *in vitro*, *in silico* and combinations thereof has been envisaged. These projects that involve more than 90 institutions and a funding of about \$40 million aim at making available batteries of alternative methods including the respective test strategies within 5 years each (Zuang and Hartung 2005). The current progress of each of these projects can be followed at the respective web sites (<http://www.reprotect.eu>; <http://www.acutetox.org>; <http://www.sens-it-iv.eu>). For testing of industrial chemicals under REACH, proposals for tiered approaches have early been made (Combes et al. 2003; Hartung et al. 2003), these issues have also comprehensively been treated in the so called REACH Implementation Projects (RIP 3.3-1 and 3.3-2) conducted by the European Chemicals Bureau (ECB) on behalf of the European Commission (<http://ecb.jrc.it>). It is prescribed by REACH [Articles 13(1), 25(1), Annex XI] to exhaust all means of collecting existing data and other means of predicting toxic effects, in particular chemical analogue and (broader) chemical category approaches including data gap filling techniques such as read-across,³ trend analysis, (Q)SARs and computational expert systems before animal testing is undertaken as a last resort (for details see ECB 2005, 2007; see also OECD, <http://cs3-hq.oecd.org/scripts/hpv>). In this respect, *in vitro* tests among alternative methods for predicting toxic effects have been developed most extensively.

Validation of alternative methods

By definition, the validation of an alternative (non-animal) method is the process by which the relevance and reliability of the method are established for a particular purpose (Balls et al. 1990, 1995). In the context of safety testing of chemicals, relevance refers to the scientific basis of the test system and to the predictive capacity of an associated

³ With regard to the definitions and use of various data gap filling techniques the reader is referred to the Draft TGD (ECB 2007).

Table 2 Toxicological standard data requirements under REACH (waiving conditions not cited)

Toxicological endpoint	Production volume
8.1 Skin irritation, skin corrosion	≥1 t/a
8.2 Eye irritation	≥1 t/a
8.3 Dermal sensitisation	≥1 t/a
8.4 Mutagenicity	≥1 t/a
8.5.1 Acute oral toxicity	≥1 t/a
8.5.2 Acute inhalative toxicity	≥10 t/a (case by case) ^a
8.5.3 Acute dermal toxicity	≥10 t/a (case by case) ^a
8.6.1 Repeated dose toxicity (28 days)	≥10 t/a
8.6.2 Repeated dose toxicity (90 days)	≥100 t/a
	≥10 t/a (case by case) ^b
8.7.1 Reproductive toxicity screening	≥10 t/a
8.7.2 Developmental toxicity	≥100 t/a
8.7.3 Two-generation reproductive toxicity study	≥100 t/a (case by case) ^b
	≥1,000 t/a
8.8 Toxikokinetics	≥10 t/a (assessment on the basis of relevant information available)
8.9.1 Carcinogenicity study	≥1,000 t/a (case by case) ^b

^a Depending on the assumed route of exposure to humans

^b For details see text

prediction model (PM), whereas reliability refers to the reproducibility of test results. Whereas a test system provides a means of generating experimental physicochemical or biological *in vitro* data for a chemical of interest, a prediction model (PM) is defined as an unambiguous algorithm (e. g., formula, rule or set of rules) for converting the data into predictions of a pharmacological or toxicological endpoint in animals or humans. The principles according to which alternative methods should be validated have been agreed at an international level, although the actual process by which the validation process is performed varies between different validation authorities (Balls et al. 1990; ICCVAM 1997; OECD 1990, 1996, 2005a, 2007a; Worth and Balls 2002a, b).

In practice, the validation process of an alternative method, e. g. an *in vitro* test, consists of several stages: (1) the development and/or improvement of the test method in one laboratory, (2) optimisation of the method by contributions of other laboratories, (3) prevalidation,⁴ (4) formal validation,⁵ (5) scientific acceptance by a competent committee,⁶ and (6) regulatory acceptance (by EU or OECD). Experience with validated methods in Europe indicates that

⁴ A prevalidation study is a small-scale interlaboratory study that is designed to obtain a preliminary assessment of the reliability and relevance of the study.

⁵ A formal validation study is an interlaboratory study that is designed to obtain a more definite assessment of reliability and relevance.

⁶ In the EU, the ECVAM Scientific Advisory Committee (ESAC) is responsible for scientific acceptance. It consists of experts from all Member States, relevant industrial associations, academic toxicology institutions, the animal welfare movement, as well as certain Commission services.

each of these stages takes a time scale of at least 1–3 years. Therefore, the validation process of a new alternative method altogether takes at least 6 years until scientific acceptance is achieved. In certain cases the validation process of a method can be accelerated if a validated “similar” method already exists (so called “catch-up validation”) (Worth and Balls 2001; 2002a, b). Recently, a “modular approach” has been proposed that is expected to accelerate the validation process (Hartung et al. 2004). The basis for the development of a prevalidation or validation study is mostly existing *in vivo* data (animal or human data) to which the results of the new non-animal method will be compared. However, the *in vivo* data is often heterogeneous with regard to the relevance or reliability of the respective studies because modern quality criteria as to the conductance and documentation of studies are often not fulfilled, in particular when considering data from older studies. Hence, it has been claimed that existing *in vivo* data are not a “gold standard” in terms of relevance or reliability. Another problem closely related to the heterogeneous quality of existing data is the selection of a set of representative substances for the objectives of a (pre) validation study as to the number, chemical class, use of the substances and the assumed spectrum of human exposures to the substances. The less suited substances are available for an intended (pre) validation study, the more critical is the issue of representativity of the study. Without doubt, the selection procedure and the criteria for the inclusion or the exclusion of substances will have an impact on the outcome of the (pre) validation study.

Among non-animal tests, up to now the validation as well as the scientific and/or regulatory acceptance of alter-

native tests is limited to in vitro tests (including ex vivo/in vitro tests such as the WEC test for developmental toxicity). According to OECD (2005a) the principles of validation as described in the OECD Technical Guidance Document were written for biology-based tests but they may be applicable to other alternative methods, too. For instance, two recent examples of validated rule-based non-testing approaches for the prediction of local toxic effects on skin and eyes are described below. Furthermore, an OECD Guidance Document for the validation of (Q)SARs has recently been published (OECD 2006c, 2007a). Besides (Q)SARs, other alternative methods such as computerised expert systems, mathematical and statistical approaches such as toxicokinetic models and pattern-based systems (so-called “omics”) are increasingly developed. However, up to now none of them has been validated.

Validation studies of more complex alternative methods or methodologies such as testing batteries and tiered testing approaches may often be not appropriate, necessary or even possible. In such circumstances a Weight of Evidence (WoE) validation assessment may be more appropriate than a validation study. ICCVAM has used the methodology of WoE validation assessment in the past for the validation of alternative testing methods such as the skin penetration test, the local lymph node assay (LLNA), the Up and Down procedure for acute oral toxicity and in vitro tests for endocrine disruption (further information is available at <http://iccvam.niehs.nih.gov>). An ECVAM Workshop has dealt with the issues of validation versus WoE validation assessment in 2004. Some prerequisites and formal rules when performing a WoE validation assessment have been described in the workshop report (Balls et al. 2006). However, up to now neither a validation nor a WoE validation assessment has been conducted with testing batteries or tiered testing approaches yet.

Current status and developments of in vitro and other non-animal methods

Several in vitro tests have been established as Testing Guidelines of the OECD, e.g. tests for genotoxicity or mutagenicity. Also, other in vitro tests such as for the determination of local toxicity to skin or eyes have been established as Testing Guidelines of the OECD for regulatory use and have partly replaced the respective in vivo tests. The in vitro methods available for regulatory use are listed in Table 3 and compared to the corresponding in vivo methods accepted by the OECD (Bernauer et al. 2005; ECB 2005, 2007; Genschow et al. 2004; Liebsch and Spielmann 2002; Piersma et al. 2004; Spielmann et al. 2004). If the in vitro methods are accepted by the OECD (footnote f in Table 3), they will be accepted by the national regulatory authorities of OECD member countries. In contrast, other

in vitro tests are only accepted at the national level by some of the Member States of the EU (footnote g in Table 3). Reviews on current status and developments are given, e.g. by Bernauer et al. (2005), ECB (2005, 2007), Eskes and Zuang (2005) or can be found in the ECVAM database (<http://ecvam-dbalm.jrc.cec.eu.int/>) or in the ZEBET⁷ database (<http://www.bfr.bund.de/cd/1591>). A good source are also the ECVAM workshop reports on specific issues of toxicological concern that are available at the ECVAM website (<http://ecvam.jrc.it>).

At the end of 2003, an ad hoc group of experts on behalf of the European Commission reviewed the state of the art of safety assessment by animal and non-animal tests for 11 different human health effects of concern in the frame of the seventh amendment (Directive 2003/15/EC) of the Cosmetics Directive 76/768/EEC, aiming at establishing time tables for the implementation of marketing and testing bans including deadlines for the phasing out of the various animal tests. The expert subgroups provided inventories of the most advanced or most promising alternative methods in each of the toxicological areas, identified problems to be solved and gaps to be filled, assessed timelines for the replacement of animal tests by alternative methods and also gave recommendations for future activities (for details see Eskes and Zuang 2005). The state of the art, the development of alternative methods and the estimated timelines required for scientific acceptance or full replacement of animal tests in different toxicological areas as assessed by these experts are summarised in Table 4. It should be stressed that the experts made their assessments of timelines required under the provision of optimal conditions related to financial and human resources and optimal co-operation and organisational support. In terms of reality, some delay may be predicted. Although the assessments were made for testing and regulatory evaluation of cosmetics, they hold also true for industrial chemicals. For the following human health-related effects, the ad hoc experts group did not foresee full replacement of animal testing before the cut-off dates provided by the seventh Amendment to the Cosmetics Directive:⁸

- Acute toxicity
- Skin sensitisation
- Genotoxicity and mutagenicity
- Subacute and subchronic toxicity
- Toxicokinetics and metabolism
- Carcinogenicity
- Reproductive and developmental toxicity

⁷ Zentrale Erfassungs- und Bewertungsstelle für den Ersatz von Tierversuchen (Centre for Documentation and Evaluation of Alternatives to Animal Experiments, Berlin)

⁸ *Phototoxicity* is of minor importance for chemicals under REACH and is therefore not considered in the following.

Table 3 Overview of toxicological in vivo and in vitro testing methods accepted for regulatory purposes

Toxicological end point	In vivo methods (no. of OECD test guideline in brackets)	In vitro methods (no. of OECD test guideline in brackets, if applicable)
1. Acute Toxicity	Acute Toxicity Oral [401, 420 (fixed dose procedure), 423 (acute toxic class method), 425 (up-and-down procedure)] ^f , dermal (402, 434 ^a) ^f or inhalation [403, 433 ^b (fixed concentration procedure), 436 ^c (acute toxic class (ATC) method)] ^f	
2. Dermal resorption	Skin absorption (427) ^f	Skin absorption (428) ^f
3. Phototoxicity		3T3 NRU Phototoxicity test (432) ^f
4. Skin irritation and corrosion	Acute dermal irritation/corrosion (404) ^f	Transcutaneous electrical resistance test (TER) (430) ^f Skin corrosion: human skin model test (431) ^f Membrane barrier test method for skin corrosion (435) ^f
5. Eye irritation and corrosion	Acute eye irritation/corrosion (405) ^f	Embryonated chicken egg (HET-CAM) ^g Isolated bovine cornea (BCOP) ^g Isolated chicken eye (CEET) ^g Isolated rabbit eye (IRE) ^g
6. Sensitisation	Skin Sensitisation (406) ^f Local lymph node assay (LLNA) (429) ^f	
7. Hepatotoxicity	Repeated dose toxicity study in rodents oral (28 day) (407) ^f , dermal (21/28-day) (410) ^f or inhalation (28-day or 14-day study) (412) ^f	
8. Nephrotoxicity		
9. Hematotoxicity		
10. Cardiotoxicity	Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test (422) ^f Repeated dose 90-day toxicity study oral in rodents (408) ^f or non-rodents (409) ^f , dermal (411) ^f or inhalation (413) ^f Chronic toxicity studies (452) ^f	
11. Endocrine effects	Two-generation reproduction toxicity study (416) ^f Repeated dose 28-day oral toxicity study in rodents; updated with parameters for endocrine effects (May 2007 version) (407 ^d) ^f	
12. Neurotoxicity	Delayed neurotoxicity of organophosphorus substances following acute exposure (418) ^f Delayed neurotoxicity of organophosphorus substances, 28-day repeated dose study (419) ^f Neurotoxicity study in rodents (424) ^f	
13. Immunotoxicity		
14. Prenatal development	Prenatal developmental toxicity study (414) ^f	Whole embryo culture (WEC) ^g Micromass Test (MM) ^g Embryonic stem cell test (EST) ^g
Pre- and postnatal development	Two-generation reproduction toxicity study (416) ^f Developmental neurotoxicity study (426) ^f	
15. Fertility	Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test (422) ^f Reproduction/developmental toxicity screening test (421) ^f Repeated dose 28-day toxicity study (407, 410, 412) ^f Repeated dose 90-day toxicity study (408, 409, 411, 413) ^f One-generation reproduction toxicity study (415) ^f Two-generation reproduction toxicity study (416) ^f	

Table 3 continued

Toxicological end point	In vivo methods (no. of OECD test guideline in brackets)	In vitro methods (no. of OECD test guideline in brackets, if applicable)
16. Genotoxicity, mutagenicity	Mammalian erythrocyte micronucleus test (474) ^f	Bacterial reverse mutation test (471) ^f
	Mammalian bone marrow chromosomal aberration test (475) ^f	<i>Saccharomyces cerevisiae</i> , gene mutation assay (480) ^f
	Sex-linked recessive lethal test in <i>Drosophila melanogaster</i> (477) ^f	<i>Saccharomyces cerevisiae</i> , mitotic recombination assay (481) ^f
	Rodent dominant lethal test (478) ^f	Mammalian chromosome aberration test (473) ^f
	Mammalian spermatogonial chromosome aberration test (483) ^f	Mammalian cell gene mutation test (476) ^f
	Mouse spot test (484) ^f	Micronucleus test (487 ^e) ^f
	Mouse heritable translocation assay (485) ^f	Sister chromatid exchange assay in mammalian cells (479) ^f
16. Carcinogenicity	Unscheduled DNA synthesis (UDS) test with mammalian liver cells (486) ^f	DNA damage and repair, UDS in mammalian cells (482) ^f
	Carcinogenicity studies (451) ^f	Cell transformation assays (SHE, Balb/c 3T3, C3H10T) ^g
	Combined chronic toxicity/carcinogenicity Studies (453) ^f	

^a Draft Test Guideline 434 Acute Dermal toxicity-Fixed Dose Procedure, May 2004 Version (http://www.oecd.org/searchResult/0,3400,en_2649_34377_1_1_1_1_1,00.html)

^b Draft Test Guideline 433: acute inhalation toxicity—fixed concentration procedure, June 2004 Version (http://www.oecd.org/searchResult/0,3400,en_2649_34377_1_1_1_1_1,00.html)

^c Draft New Test Guideline 436: acute inhalation toxicity—acute toxic class (ATC) Method, December 2004 Version (http://www.oecd.org/searchResult/0,3400,en_2649_34377_1_1_1_1_1,00.html)

^d Draft Updated Test Guideline 407: repeated dose 28-day oral toxicity study in rodents; Updated with parameters for endocrine effects, May 2007 version (http://www.oecd.org/document/55/0,3343,en_2649_34377_2349687_1_1_1_1,00.html)

^e Draft Proposal for a New Guideline 487: in vitro micronucleus test, December 2006 Version (http://www.oecd.org/document/55/0,3343,en_2649_34377_2349687_1_1_1_1,00.html)

^f OECD Guidelines for the testing of chemicals (<http://oberon.sourceoecd.org/vl=5003995/cl=13/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/vIn4/contp1-1.htm>);

^g Tests listed only in Annex V of Directive 67/548/EEC or accepted by regulatory authorities of some EU Member States

The adequacy of alternative methods for the prediction of toxicological endpoints deserves some commenting and evaluation as follows.

Toxicokinetics and metabolism

This topic is discussed in the first place because it is of importance in most of the other toxicological areas discussed below and is regarded as a crucial issue and a “bottleneck” hampering the predictivity of toxicity by alternative tests. Metabolism often plays a key role in the origin of toxic effects and in inter- and intra-species differences. However, the drug metabolising capacity of most in vitro systems is either often different and thus inappropriate in relation to the human situation or low, variable or continuously reduced under culture conditions. Frequently used metabolising systems for in vitro toxicity tests and their advantages and disadvantages are compiled in Table 5. Co-culture of indicator cells with drug metabolising competent cells or addition of S9-mix is often not pos-

sible or may induce severe problems in data interpretation. Although there has been considerable progress in the development of expert systems and other in silico techniques, computerised systems are at present not capable to predict metabolism in a sufficiently exact manner. Hence, the inclusion of metabolism in predicting toxicity by means of in vitro tests and/or computerised systems is far from being satisfactory (Coecke et al. 2006a). Toxicological testing of the main metabolites of the compound under consideration would be desirable but is not possible in most cases because the analytical methods for the characterization of the metabolic profile of industrial chemicals are not available as a rule and synthesis of such metabolites would be an expensive effort. One promising way to better incorporate metabolism in in vitro tests could be the development of genetically engineered cell lines capable of phase 1 and 2 metabolism preferably by enzymes of human origin. The use of human enzyme sources and several possible applications are reviewed by Pelkonen et al. (2005).

Table 4 Current developments of alternative methods according to an expert group on behalf of the EU Commission: Status at the end of 2003 (according to Eskes and Zuang 2005)

Health effect or test area	Current status of test development	Estimated time to achieve	
		Scientific acceptance by ECVAM Scientific Advisory Committee	Full replacement of animal tests ^a
Acute toxicity	Two in vitro tests under validation, Several tests under prevalidation, test batteries and tests strategies required	≥2 years ≥4 years	>10 years
Skin irritation	Three in vitro tests under validation, One test optimised	1–2 years	4–5 years for hazard identification, 6 years for risk assessment
Eye irritation	Thirteen in vitro tests, six tests most promising	4 years for most promising tests	>6 years
Skin sensitisation	Optimised in silico models, in vitro tests under R&D	5–6 years for in silico methods, 5–12 years for in vitro tests	≥12 years
Subacute and subchronic toxicity	In vitro models for five target organs under R&D, test batteries and test strategies required	>10 years	Not estimated ^c
Genotoxicity and mutagenicity	Seven in vitro tests accepted at the EU and OECD level ^b , several tests under R&D or optimised	1–6 years for optimised tests, >6 years for tests under R&D	>12 years
Toxicokinetics and metabolism	In vitro tests under R&D, partly optimised, In silico methods under R&D, test batteries and test strategies required	3–8 years >12 years including excretion	≥10 years
Carcinogenicity	Three optimised tests accepted at the EU level ^b One test under R&D, test batteries and test strategies required	5 years for optimised tests, >10 years for test under R&D	Not estimated ^c
Reproductive and developmental toxicity	Three validated tests accepted at the EU level ^b , several tests under (pre) validation or R&D, test batteries and test strategies required	0 years, 6 and >10 years for tests under (pre) validation and R&D, respectively	Not estimated ^c

^a Regulatory acceptance at the EU level

^b See Table 3

^c The experts assessed the time required for full replacement of animal tests to be probably much longer than 10 years

Table 5 Frequently used metabolizing systems for in vitro toxicity tests

In vitro system	Advantages	Differences compared to the in vivo situation
Liver S-9 mix	Easily available High sensitivity due to a highly induced phase I metabolism	Strongly reduced phase II metabolism High fraction of false positive results for S-9 from induced rats
Microsomes	Easily available High activities of cytochrome P450 enzymes	Includes only a subset of all relevant drug metabolizing enzymes
Heterologous expression of enzymes	Possibility for identification of metabolites generated by specific enzymes	Only a limited number of enzymes can be expressed (more than two becomes technically demanding)
Primary hepatocytes	In vivo like metabolic activities for limited periods of time	Alterations of several metabolising activities depending on the culture conditions Co-culture with target cells does not guarantee in vivo like pharmacokinetics
Liver slices	In vivo like metabolic activities for limited periods of time	Applicable only for short incubation periods due to necrosis

Apart from the metabolism, the requirements for inclusion of absorption, distribution and excretion need to be discussed. For instance, the extent of absorption and penetration of a substance through skin together with its inher-

ent toxicity is crucial for the risk assessment of dermal exposure, occupational hygiene measures and may trigger the necessity of setting Biological Exposure Limits. Skin absorption and penetration of a substance can be measured

and expressed as dermal flux. For skin absorption and penetration, an *in vitro* test has been accepted at the OECD level [Testing Guideline (TG) 428], although it was not formally validated in a prospective validation study (Table 3). In some respects it can be regarded as an alternative to the animal test; however, metabolism in the skin is not covered by the test and repeated dose testing is limited to 1–2 days. It should also be pointed out that experimentally determined dermal fluxes of industrial chemicals might largely differ between *in vivo* data and experimental systems based on *in vivo* human or animal skin. In addition, data may also largely vary between individual humans *in vivo* and between different skin areas when applying the test substance to the same human individual (European Commission 2006; IPCS 2006; Williams 2004, 2006). Thus, all the models including alternative *in vitro* and human *in vivo* models available for predicting skin absorption and penetration have considerable limitations and need further research and development. For intestinal absorption, the presence of active transporters (e.g. ABC proteins) in the enterocytes is crucial but is not covered by most tests currently used. In this respect, the models of Caco-2 cell monolayers and other *in vitro* membrane barrier models need to be improved and standardised (Le Ferrec et al. 2001; Prieto et al. 2004). *In vitro* tests predicting the elimination of a substance via the kidney are not described in the literature.

ADME properties of a substance are dependent on the physicochemical properties of the substance and hence this fact offers an opportunity for useful ADME prediction by computation. Although *in silico* ADME prediction systems may have great potential, they all require further improvement before they can be considered acceptable for specific applications. Moreover, it is important to be aware of the fact that for the prediction of absorption *in silico* systems must incorporate the role of metabolism as well as the role of active transporters. Bergström (2005) concluded in a review on the prediction of drug absorption by *in silico* models that larger volumes of consistent high quality experimental data compiled in the databases are required as “prerequisites for a quantitative rather than qualitative prediction of absorption”. Glomerular filtration as well as tubular re-absorption and secretion can be predicted from the physicochemical properties of the compound and its plasma protein (albumin) binding. If active secretion or re-absorption and saturation kinetics are also involved these processes are less predictable. Also, the role of active transport via specific mechanisms is important for elimination processes and is difficult to predict. Since there are various reasons why *in silico* ADME models may fail, van den Waterbeemd (2005) proposed further development of robust *in silico* tools in combination with high-throughput *in vitro* screening that will lead to an “in combo” approach towards the estimation of ADME properties of drugs and

other substances. Such development will require many years for validation, scientific and regulatory acceptance.

Discourse: toxicokinetics in man—a useful alternative methodology

Methods for studying toxicokinetics in humans at low doses are alternative methods *par excellence* with an inherent high degree of relevance. However, the methodology to investigate the toxicokinetic properties of chemical substances in humans is limited for various reasons. The knowledge on the fate of chemical substances in humans often originates from accidental or suicidal uptake of high doses or from biomonitoring or biological effect monitoring studies of chemicals in exposed workers or exposed groups of the general population. Studies on pharmacokinetics/toxicokinetics in humans are performed to provide data support in the extrapolation process of animal toxicity data to the human situation, in the validation of physiologically-based pharmacokinetic models developed based on *in vitro* data, and to obtain basic information of ADME for new pharmaceuticals.

In the past, e.g. in case of heavy metals or other, e.g. organic chemicals, radioactive isotopes or organic compounds containing radiolabel have been applied in humans because the analytical methods were often not sufficiently sensitive to determine low concentrations of the element or compound in question. Apart from ethical considerations, the high costs for the synthesis of radio-labelled substances and problems of contamination are further reasons why the published literature does not contain many examples of the intentional experimental intake of low doses of radio-labelled toxic chemicals by humans in order to determine a substance’s toxicokinetics, apart from drugs. One prominent example is the determination of the biological half-life of the toxic compound 2,3,7,8-TCDD in some volunteers, after intake of a very small dose of the radiolabelled compound (Poiger and Schlatter 1986).

The recent developments in analytical chemistry permit conclusive identification and quantitation of xenobiotic metabolites present in low concentrations in human body fluids such as urine and blood. The use of unlabelled compounds requires the establishment of analytical methods to quantitate parent compounds and all relevant metabolites, which is often expensive and time-consuming. Metabolite identification usually has to be performed based on experiments in animals with labelled compounds. Studies on the ADME of chemicals in humans have been used to characterize the toxicokinetics of a variety of environmental contaminants, occupational chemicals or chemicals present in food, and are recognized as important contributions in the extrapolation process in the risk assessment (Amberg et al. 2001; Bernauer et al. 1996; Ernstgard et al. 2005; Fennell et al. 2005; Filser et al. 1992; Gunnare et al. 2006; IPCS 2006; Pähler et al. 1999; Schauer et al. 2006; Völkel et al. 2002).

Studies on ADME of chemicals in humans require review by ethical review boards as outlined in the Declaration of Helsinki and follow-on documents. They also require detailed information of potential risk to the participants and informed consent. In addition, a justification of dose administered should be provided. For non-genotoxic agents, dose selection may be guided by occupational exposure limits or tolerable daily intakes. If such endpoints are not defined, dose selection may be based on known endpoints in animal toxicity studies applying an adequate margin of safety (MOS).

Skin irritation and corrosion

Validated alternative methods accepted for regulatory use determining skin corrosion exist that have displaced the respective animal tests (see Table 3). The validation of two methods for the testing of skin irritation, EPISKIN[®] and EpiDerm[®], has been finalised and scientifically accepted by ESAC in Spring 2007. The EPISKIN[®] method is foreseen to completely replace the regulatory Draize skin irritation test whereas the EpiDerm[®] model is seen as a constituent of a testing strategy. Other aspects such as reversibility of irritation and dose-response characteristics, important components for risk assessment, have yet to be addressed. Further information is available at the ECVAM website (<http://ecvam.jrc.it/index.htm> section news, events and meetings).

By use of a rule-based non-testing approach and using data from the EU New Chemicals Database for chemicals notified in the EU, a decision support system for skin corrosion/irritation potential in terms of C&L has been developed and validated consisting of two predictive tools: (1) physicochemical exclusion rules (cut-off values) to identify chemicals with no skin irritation/corrosion potential and (2) inclusion rules to identify chemicals with skin irritation/corrosion potential by use of structural alerts of chemicals. Furthermore, a Skin Irritation Corrosion Rule Estimation Tool (SICRET) has been developed which is a user-friendly tool that enables non-QSAR experts to identify chemicals with or without skin corrosion or irritation potential based on physicochemical properties or structural alerts (Walker et al. 2005). The validation study for predicting skin corrosion/irritation comprised 1,833 substances and the predictivity was >95% (Gerner et al. 2000, 2004; Walker et al. 2004). In an additional external validation study using 201 notified chemicals not identical with the training set, the validation study was confirmed (Hulzebos et al. 2005).

Eye irritation and corrosion

Four tests, although not formally validated, have been accepted by regulatory authorities of some EU member states for the identification and classification of severe eye

irritants (Table 3). In Spring 2007, two of these tests, the Bovine Corneal Opacity and Permeability (BCOP) and the Isolated Chicken Eye (ICE) test methods have been validated and scientifically accepted by ESAC. Up to now it has not been possible to identify a single method or a combination of *in vitro* methods that could replace the animal test for mild eye irritants. Also, major problems such as reversibility or persistence of irritation as well as differentiation of slight effects and irritation potential relevant for human health have not adequately been addressed yet. This may be due to the fact that the eyeball has complex functional structures and tissues with different vulnerabilities to chemicals and partly also to inherent limitations of the animal tests or low quality of animal test data which render the development of non-animal tests more difficult. In analogy to the decision support system for skin corrosion/irritation potential in terms of C&L, a similar rule-based non-testing approach based on physicochemical properties and structural alerts was successfully carried out to predict (a) chemicals having no eye irritation/corrosion potential and (b) chemicals possessing eye irritation/corrosion potential, respectively (Gerner et al. 2005). The advantage of these approaches is that data from standardised *in vivo* data (OECD Testing Guidelines) are available for notified new chemicals (different from most existing chemicals). A limitation is the fact that the data in the EU New Chemicals Database are confidential and therefore not available to the scientific community. Another limitation should be seen in that new chemicals are often developed for a limited range of applications (e. g. dyes, polymer components) and therefore do not comprise the whole range of chemicals humans may be exposed to. Thus, these rule-based non-testing approaches and their validation should be extended to chemical groups and structures that are not included in the EU New Chemicals Database.

Skin sensitisation

Following a retrospective analysis of published data, the Local Lymph Node Assay, the *in vivo* alternative method for testing of skin sensitisation (LLNA, see Table 3), has recently been modified in terms of reduction of animal use. ESAC in April 2007 accepted that, within a tiered testing strategy in the context of REACH, a reduced version of the LLNA (rLLNA) using only the equivalent of the high dose group from the full LLNA can be used as a screening test to distinguish between sensitisers and non-sensitisers. However, ESAC pointed out that some limitations compared to the full test method have to be taken into account (for further information see <http://ecvam.jrc.it/index.htm>, section news, events and meetings).

In contrast to the development of *in vivo* alternative test methods, cellular systems capable of distinguishing

between sensitising and non-sensitising substances are available only at the research level and will probably require more than 10 years for scientific and regulatory acceptance. Since the complex mechanisms of sensitisation are not well understood, it is difficult to develop tests that are based on the relevant biological events associated with skin sensitisation. One suggested approach is the test of chemicals in cultures of human blood-derived dendritic cells since the differentiated dendritic cells of the skin (Langerhans cells) play a key role in the recognition of haptens and the complex process of the immune response by T lymphocytes. Another approach is the measurement of cytokine expression in keratinocytes, in co-culture with dendritic cells or in reconstituted epidermis. Current status and problems of this approach are reviewed, e.g. by Casati et al. (2005). Since single in vitro tests will not be able to cover all aspects of the complex mechanisms of skin sensitisation, a test battery should be developed. Many substances act as prohaptens requiring metabolic activation to exert their sensitising properties. Hence, metabolic competence of cellular systems is an important component that should be considered in the development of a test battery. Furthermore, important aims for the future are to distinguish between contact allergens and skin irritants and to distinguish between contact allergens of varying sensitising potency. A recent project, Sens-it-iv started at the end of 2005 with a view to complete the development of animal-free test strategies for skin and lung sensitisation (<http://www.sens-it-iv.eu>).

A recent publication recommends the use of (Q)SAR to predict skin sensitisation via expert systems taking into account functional groups of chemicals and information on mode or mechanism of action (Roberts et al. 2007). Taking this into account, a (Q)SAR model was developed including LLNA test data for 258 substances, guinea pig maximisation test data for 360 substances, and data from 244 substances from a list of contact allergens (Schlede et al. 2003), respectively (Patlewicz et al. 2007). To validate this model, in an initial effort, data were generated for 40 new chemicals in the LLNA and then compared with predictions made by the model. The results showed an overall concordance of 83% between experimental and predicted values. Further studies and improvements will show whether this approach will gain relevance in practice.

Acute systemic toxicity

The objectives of testing acute systemic toxicity are to detect (1) the type(s) of acute toxic effects of a chemical and (2) to determine severe acute toxic effects or lethality in a quantitative manner and to allow for classification and labelling for acute toxicity. Basal cytotoxicity is an essential component of acute toxicity. It appears from a number

of studies showing positive correlations between in vitro cytotoxicity and in vivo acute toxicity that in vitro test methods may have a potential to predict quantitative aspects of acute toxicity (Botham 2004). ICCVAM/NICEATM and ECVAM completed in 2005 a joint validation study to characterize the usefulness and limitations of in vitro cytotoxicity assays as predictors of starting doses for rodent acute oral toxicity test methods (ICCVAM/NICEATM 2001a, b, 2006; Stokes et al. 2002; Zuang and Hartung 2005). In this study, 72 reference substances with known human toxicity and/or human exposure have been tested for cytotoxicity in BALB/c 3T3 mouse fibroblasts (3T3) and normal human epidermal keratinocytes (NHK). The Neutral Red Uptake (NRU) test in both the cell systems correctly predicted the hazard category, according to the Globally Harmonized System of Classification and Labelling (GHS), of only about 30% of the reference substances. The data obtained were used to estimate starting doses for rodent acute oral testing, based on linear regressions developed from the Registry of Cytotoxicity (RC) database (Halle 2003).

ICCVAM's recommendations for use of the in vitro NRU test methods are as follows (ICCVAM/NICEATM 2006): The 3T3 and NHK NRU test methods alone are not sufficiently accurate to predict acute oral rodent toxicity for regulatory hazard classification. They may be used in a weight-of-evidence approach to determine the starting dose for current in vivo acute oral toxicity protocols, i.e., Fixed Dose Procedure (OECD guideline 420), Acute Toxic Class Method (OECD guideline 423), Up and Down Procedure (OECD guideline 425). The starting doses for substances with certain toxic mechanisms that are not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic or cardiotoxic) will likely be underestimated by these in vitro basal cytotoxicity test methods. Therefore, the results from basal cytotoxicity testing with such substances may not be appropriate for estimating starting doses.

The development of alternative methods for predicting other important components of acute toxicity such as type, onset, duration and reversibility of the toxic as well as toxicokinetics and metabolism is only at the research level. Thus, an integrated project, *ACuteTox*, dealing with these issues has been started in January 2005 with 35 research groups from 13 European countries and an overall budget of 16 million € (duration 5 years, <http://www.acute-tox.org>). Besides the development of a database with in vitro and in vivo data on acute toxicity for 100 to 140 chemicals, the selection and refinement of in vitro tests is planned. The approach is to add new in vitro tests for toxicokinetics and organ specificity to a basal set of cytotoxicity assays in combination with computer modelling. These components have to be integrated into a validated in vitro test battery or testing scheme combined with a prediction

model for data extrapolation aiming at hazard identification of acute toxicity to humans. Further aspects and problems of this approach have been discussed by Gennari et al. (2004).

We consider the progress, which has been made in replacement of acute toxicity testing, is regrettably small compared to the large efforts made. However, one should keep in mind that prediction of acute toxicity has to deal with similar difficulties as prediction of other endpoints which are generally regarded as “more complex”. Predicting acute toxicity is more than just predicting the “simple endpoint death”. It has to take into account the kinetics of the substance, mode of action and its possible interaction with all tissues as well as the extensive capability of the organism to compensate for disturbance of homeostasis. In accordance with these considerations, development of replacement methods for acute toxicity testing cannot proceed much faster than for, e.g. repeated dose toxicity testing. Finally, one should not forget that refinement of acute toxicity testing (resulting in the above-mentioned OECD Guidelines) has already resulted in a substantial reduction of animals used for such studies and the pain involved. We are not convinced that animal-free methods will satisfy future regulatory requirements concerning acute systemic toxicity, but consider further refinement of animal tests by improved *in vitro* testing possible. However, we suspect that the number of animals that can thus be saved may be small.

Subacute and (sub) chronic toxicity

In Annexes VIII–XI of the REACH Regulation, repeated dose toxicity studies are required for substances with volumes of 10 tons up to more than 1,000 tons per year. The minimum requirement for testing of chemicals is the test for subacute toxicity with 28 days of exposure (OECD Testing Guidelines 407, 410 and 412 or the respective guidelines B.7, B.8 and B.9 in Annex V of the Guideline 67/548/EEC). These Testing Guidelines require a minimum of five animals per dose group and gender. In the frame of REACH, the 28-day short-term toxicity test has a great importance since this test is foreseen to be applied to approximately 15,000 chemicals that are produced in amounts of 10 tons per year or more. According to Annex VIII, No. 8.6.1 additional studies such as the 90-day subchronic study may be required, in particular when toxicity of particular concern (e.g. serious/severe effects) is indicated or in case of an effect for which the available evidence is inadequate for toxicological and/or risk characterisation. It should be noted that compared to the regulatory requirements of drug and biocide toxicity testing, the 28-day toxicity test for chemicals is a compromise with respect to duration and minimum animal number and

hence with respect to costs and animal welfare since for the registration of drugs and biocides, a 90-day toxicity study is *a priori* required.

The intention of the 28-day toxicity study is (1) to detect the target tissue(s) of the chemical to be tested, potential cumulative effects and the reversibility of the toxicity and (2) to determine a dose without effect in order to provide a basis for risk assessment. While no elaborated non-animal concept does at present exist for the latter objective one can in principle design a set of cell cultures from different tissues which might indicate the target organ(s). However, such data do not represent the information which comes from a conventional rodent 28-day toxicity study. OECD guidelines demand data on body weight development, clinical chemistry parameters, organ weights and gross necroscopy and histopathological examination of all tissues. Cell culture models lack a number of conditions present *in vivo* which determine the outcome of these measurements. Both the homeostatic capacity of the organism and the possible potentiation of toxicity induced by cell–cell and tissue interaction cannot be imitated in *in vitro* test systems. For instance, a cell culture model will in general not allow for an exacerbation of injury caused by the immigration of inflammatory cells to the site of the primary lesion. To mimic this, complex co-culture models would have to be constructed for every target cell culture. Reversibility of injury is another important aspect which critically influences the results of an *in vivo* toxicity study. However, cell culture models are rarely capable to also mimic of tissue regeneration. Thus, while a variety of *in vitro* tests for some of the main targets of repeated dose toxicity testing have been developed at the R&D level or have been improved (liver, kidney, lung, central nervous system, haematopoietic system) none, single or in combination, is at present capable of adequately predicting all possible repeated dose toxic effects on the target organ they represent. Even more basically, there are inherent problems of culturing primary cells which have not yet been solved satisfactorily such as the limited life time of most of the cells or tissues in culture, loss of important cellular functions due to the culture conditions (a feature also typical for immortalised cell lines) and the limited spectrum of responses to toxicity by cells or tissues in culture compared to the whole organ *in vivo*. Despite considerable progress of *in vitro* research (e.g. co-cultures, sandwich or other three-dimensional cultures) it is only partly understood which factors are needed to allow cellular systems *in vitro* to maintain their functions for more than a few hours, days or weeks. This is a matter of basic research in the coming years.

In a recent review, Prieto et al. (2006) have discussed these issues and have reached the general conclusion that conventional *in vivo* tests for subacute or subchronic toxicity cannot be replaced at present or in the near future by *in*

vitro or other alternative methods. In particular they came to the following conclusions: “The major limitations to the use of in vitro models for the assessment of toxicity after repeated dosing are: (1) the lack of suitable in vitro systems to mimic all the possible interactions which may result in vivo, (2) the limited possibilities of using cell culture systems to account for kinetics and biotransformation, (3) the difficulty to derive from in vitro systems values such as the NOAELs, which are traditionally used as the starting point for risk assessment.” Some of their proposals deserve to be discussed in more detail because these approaches, while not claiming to replace them, could contribute to reduce the number of in vivo tests. First, they consider using organotypic in vitro models as a “filter” to exclude chemicals with specific unwanted or unacceptable effects from further unnecessary in vivo testing. They further propose to develop a strategy to predict repeated dose (28 days) toxicity by short-term low-dose exposure in vitro. For establishing such a strategy, a set of relevant end points in vivo and corresponding (presumably) early biological indicators in vitro that are sufficiently sensitive and relevant would be needed. In particular, the anticipations into such indicators are directed to the rapid development of pattern-based technologies such as toxicogenomics (see below). However, the identification and establishment of relevant end points and corresponding sensitive biological indicators in vitro for the prediction of toxicity in vivo is only at the research level. To avoid the necessity to develop and validate hundreds of biological indicators for reactive or adaptive responses at the cellular and tissue level more effort in basic research for better understanding pathogenetic processes leading to chronic toxicity is a prerequisite for test development. The existing tests should be optimised and further tests may be developed based on a mechanistic understanding of the underlying pathogenetic processes, which, however, is far from well developed.

Another approach discussed by Prieto et al. (2006) concerns the determination of a maximal concentration without any effect in vitro and the correlation of such a NOEC in vitro to a NOEC or NOEL in vivo. They propose to predict the actual target organ concentration in vivo under the expected exposure conditions by use of biokinetic models, to compare it to the NOEC in vitro and thus to establish a margin of safety (MOS). This approach may be useful for substances with very low toxicity since in vivo testing is not required when the corresponding NOEL in vivo (which is calculated from the in vitro NOEC) is higher than the limit dose for in vivo testing. Also, in vivo testing would not be required when human exposure is low and the derived MOS is large. In addition, this approach may have the advantage that the establishment of biomarkers for the determination of critical effects in vitro and for the derivation of a NOAEL in vivo is not required.

However, one of the main unsolved problems of such an approach is the requirement of valid biokinetic models to arrive at a target organ concentration in vivo. It is even questionable whether the nominal concentration in vitro exactly determines the concentration of the substance at the site of action, which is a prerequisite to assess the concentration of the substance at the site of action in the target organ in vivo. Although there is much effort to use PBPK models and biokinetic modelling for calculating in vivo kinetics on the basis of in vitro data, predicting substance concentrations at the site of action (in vitro) and in vitro–in vivo scaling it must be stated that the models including QSAR methods are presently not sufficiently developed to extrapolate from in vitro to in vivo situations and that their validation is not expected in the next future.

Recently, several groups have reported on the use of whole genome transcriptional profiling in short term animal tests to identify substance specific alterations in mRNA expression patterns (review: Hengstler et al. 2006). It has been reported that gene array analysis allows the differentiation between selected compounds associated with different subtypes of hepatotoxicity. Specific gene clusters for microvesicular lipidosis, hepatocellular necrosis, inflammation, hepatitis, bile duct hyperplasia and fibrosis have so far been identified (Huang et al. 2004; Waring et al. 2001). These examples illustrate the potential of such novel techniques to predict specific endpoints of toxicity after short-term in vivo exposure of laboratory animals and thus to contribute to a refinement and reduction of animal testing. However, further research is required for establishing, standardising and validating these methods as prediction tools of specific toxic mechanisms or effects. Study conditions such as optimal species, strain, required minimum of study duration and other open questions should be thoroughly investigated. In particular the pool of substances investigated by these methods should be extended including substances exerting weak or borderline effects or acting by atypical mechanisms. Toxicogenomics and other pattern-based methods such as proteomics and metabolomics might successfully be incorporated into integrated testing strategies. However, we anticipate that the development and validation of computerised methods, testing batteries and tiered testing schemes for predicting repeated dose toxicity have to overcome many scientific and regulatory obstacles, which make it extremely difficult to predict the outcome and the time needed.

Neurotoxicity

The REACH Regulation substantiates in Annexes VIII–X that neurotoxicity studies shall be proposed by the registrant or may be required by the Agency in case of indications of neurotoxic effects for which the available evidence

is inadequate for toxicological and/or risk characterisation. In such cases, it may be more appropriate to perform specific toxicological studies that are designed to investigate these effects.

The functional specifics of the central and peripheral nervous system result in highly specific mechanisms of action of neurotoxic substances. Thus, *in vivo* test guidelines currently used to identify a substance-related neurotoxic potential are based on a number of endpoints of neurobehaviour, neuropathology, neurophysiology, and neurobiochemistry. The value of alternative testing for neurotoxicity in the context of regulatory needs has been discussed by Coecke et al. (2006b). One way is to study neurotoxicity in specific cell types of the brain and to derive generalised mechanisms of action of the toxicants. Additionally, toxicokinetic models including the blood–brain barrier (BBB) are to be developed. The development of batteries of *in vitro* tests with adequate endpoints as well as the integration of *in vitro* toxicity data with biokinetic modelling (Forsby and Blaauboer 2007) could be promising in order to cover the complexity of the nervous system. ECVAM's Workshop 49 focused on *in vitro* models for studying the BBB (Prieto et al. 2004). It was concluded that a number of *in vitro* models of the BBB are available but minimal requirements for these models need to be better defined, particularly in relation to availability and ease of culture, the functional expression of transporter mechanisms, the possibility of studying polarity, restrictive paracellular transport, and closeness of morphology to that of *in vivo* systems. An important aspect was the consideration of the inclusion of biokinetic modelling and the BBB in integrated testing strategies.

Different *in vitro* models have been examined in an ECVAM prevalidation study with the aim to identify a model for the BBB based on the use of continuous cell lines derived from the BBB, and to investigate the specificity of this model (Garberg et al. 2005). The generally rather low correlation between *in vitro* and *in vivo* data obtained indicate that other factors than permeability influence the distribution of compounds to the brain *in vivo*. Protein binding, blood flow, metabolic stability and lipophilicity, as well as the affinity for different transporters expressed in the BBB, seem to be such factors. It was concluded that the installation of a battery of *in vitro* tests is likely to be necessary in order to improve *in vitro*–*in vivo* correlations and to make it possible to perform predictions of *in vivo* brain distributions from *in vitro* data.

Under the sponsorship of the EC, Gartlon et al. (2006) evaluated an *in vitro* testing strategy for predicting *in vivo* neurotoxicity. The sensitivity of differentiated PC12 cells and primary cerebellum granule cells (CGC) were compared to that of undifferentiated PC12 cells. Fifteen cytotoxicants and neurotoxicants selected for testing covered a

range of mechanisms and potencies *in vivo*. Neurotoxicants could not be clearly distinguished from cytotoxicants despite significantly different cell system responses using the endpoints cell viability/activity, ATP depletion, mitochondrial membrane potential (MMP) depolarisation, ROS production, and cytoskeleton modifications. It was concluded that further work is required to determine suitable combinations of cell systems and endpoints capable of distinguishing neurotoxicants from cytotoxicants.

Most recently, incorporating *in vitro* methods for developmental neurotoxicity (DNT) testing into international hazard and risk assessment strategies has been discussed within a workshop hosted by ECVAM (Coecke et al. 2007), and within the first workshop of the TestSmart DNT programme (Lein et al. 2007). They focused on: (1) the models available to capture the critical DNT mechanisms and processes, (2) the creation of a high quality open database to catalogue DNT data, and (3) policy and strategy aspects to assess the integration of alternative methods in regulatory decision-making. A first step would be to refine current *in vivo* strategies by integrating information derived from *in vitro* models and using non-mammalian species (e.g. zebrafish, *C. elegans*, medaka) in alternative test strategies. Because at present *in vivo* based DNT testing cannot be replaced by *in vitro* approaches, incorporation of *in vitro* testing as a part of an intelligent testing strategy could at least refine and reduce animal usage.

We conclude that the development of *in vitro* neurotoxicity tests is at present only at the research level. Up to now, no *in vitro* models for neurotoxicity testing have been validated or have been accepted for regulatory purposes. The scientific and regulatory validation of alternative neurotoxicity models remains challenging.

Genotoxicity and mutagenicity

Amongst all toxicological endpoints, chemical mutagenesis and genotoxicity, have made the most intensive use of various *in vitro* approaches, over the last decades. As one of the consequences, several *in vitro* tests indicating genotoxic or mutagenic effects are currently accepted at the OECD level (Table 3). However, these tests still have relevant limitations, such as an insufficient metabolic capacity (Hewitt et al. 2007a; Gebhardt et al. 2003; Hengstler et al. 2000). Moreover, *in vitro* genotoxicity testing in doses concomitant with relevant cytotoxicity may lead to test oversensitivity compared to the *in vivo* situation (Kirkland et al. 2007). Further characteristics like karyotypic instability and deficiencies in, e.g. p53 and DNA repair mechanisms of the commonly used rodent cell lines (e.g. V79, CHO) are additional factors potentially leading to results not representative for the *in vivo* situation. Misleading *in vitro* results may also be particularly likely to be obtained in case of a

target organ specific mode of action (Hengstler et al. 2003). As a general consequence, a high portion of mainly false positive but also false negative test results may arise. An expert group on behalf of the European Commission considered the current situation as insufficient and recommended the development of a completely new-tiered testing strategy, i. e. the development of a tier of *in vitro* target organ/system models. The time required for the development and validation of such models or testing blocks is assumed to be 8–10 years (Eskes and Zuang 2005).

Hazard classification under REACH specifically relates to germ cell mutagenicity, i.e. genetic damage, which is passed over to the next generation. In case of the negative outcome of tests conducted *in vitro*, indicating no genotoxic or mutagenic effects, no further *in vivo* testing is required. Only if the *in vitro* tests are positive *in vivo* testing has to be considered. In case there is a positive *in vivo* mutagenicity/genotoxicity test in somatic cells, the substance under study is classified as being suspected to cause germ cell mutagenicity. In such case, further studies to assess the potential of germ cell mutagenicity can be considered. The established test strategy is to perform further *in vivo* tests, e.g. the mouse heritable translocation assay (OECD Test Guideline 485). There is no respective non-animal testing approach available at present, like, e.g. PBPK models, to predict whether the somatic cell mutagen/genotoxicant reaches the germ cells *in vivo*.

In summary, despite long-term intensive scientific efforts in many working groups around the world, it is still not possible to sufficiently predict *in vivo* somatic cell mutagenicity by *in vitro* studies. Consequently, REACH uses positive *in vitro* study results only as a trigger for *in vivo* studies. On the other hand, negative *in vitro* test results abrogate the requirement to perform further *in vivo* tests. With respect to the prediction of germ cell mutagenicity by alternative methods, conventional *in vitro* tests are used since there are currently no established non-animal approaches available that are specific to mutagenicity in germ cells.

Carcinogenicity

Chemical carcinogenesis is a complex and multifactorial phenomenon of long-term toxicity that is mechanistically by far not understood although some progress has been made during the last decades. Some important puzzling features are: the multistage character of the process, changes of cellular regulation, organ or species specificity of many carcinogenic substances as well as mechanisms that antagonize the multistep process of carcinogenesis (Hengstler et al. 2003; Bolt et al. 2004). Metabolic inactivation, DNA repair, cell cycle arrest, apoptosis, oncogene induced senescence, and control by the immune system represent

well-documented mechanisms that influence the shape of dose-response relationships in animal carcinogenicity studies (Hengstler et al. 2003; Trost et al. 2005; Spangenberg et al. 2006; Hengstler et al. 2006). Several properties of carcinogenic substances influence the outcome: the degree of genotoxicity in the cells of origin of cancer, cytotoxicity at the site of carcinogenic action and promoter activity such as inhibition of apoptosis or hormonal activity. Due to the complexity of carcinogenesis, prediction of carcinogenic properties of a given substance is difficult. Presently, no single *in vitro* test or test combination is considered to be sufficient.

According to REACH, classification of a high production volume compound in the band of >1,000 tons per year as a mutagen category 3 can give rise to a full carcinogenicity study while mutagens in categories 1 and 2 are usually regarded as carcinogenic (default presumption) without further *in vivo* testing. A carcinogenicity study may be proposed by the registrant or may be required by the agency if the substance has a widespread use or if there is evidence of frequent or long-term human exposure and the substance is classified as mutagen category 3 or if there is evidence from the repeated dose study(ies) that the substance is able to induce hyperplasia and/or pre-neoplastic lesions (Annex X, 8.9.1).

However, a considerable number of non-genotoxic substances are also carcinogenic. Several *in vitro* transformation assays have been established for their detection (Sakai et al. 2002). The largest database is available for transformation assays such as the Syrian hamster embryo cell assay (SHE), the C3H10T1/2 and the BALB/c3T3 assay, although it should be considered that these tests have not yet been finally approved by the OECD (OECD 2001, 2006d; see Table 3). All assays rely on the formation of transformed cell foci due to the influence of the test substance. Sensitivity (specificity) of the SHE assay, the C3H10T1/2 assay and the BALB/c3T3 assay was determined to be 84% (81%), 71% (67%) and 62% (62%), respectively (OECD 2001). A limitation of these methods is the relatively long time period of 4–8 weeks until focus formation can be observed as an indication of transformation. In order to reduce the time required for *in vitro* transformation assays, Bhas42 cells have been established by transfection of BALB/c 3T3 cells with v-Ha-ras (Sasaki et al. 1988). Transformed foci can be induced in these cells within a period of 2 weeks. Inter-laboratory studies have been performed reporting that the Bhas42 cell assay may represent a sensitive screening method for the identification of transforming chemicals. So far, however, much lower numbers of substances have been tested by this assay compared to the SHE, C3H10T1/2 or BALB/c3T3 assays (Ohmori et al. 2004, 2005; Umeda 2006). These data demonstrate that cell transformation assays may represent valu-

able tools for the identification of non-genotoxic carcinogens. Besides non-genotoxic carcinogens, also, many genotoxic carcinogens are positive in transformation assays. However, the identification of genotoxic carcinogens is not the major scope of transformation assays, since genotoxic carcinogens can be identified by conventional and easy to handle mutagenicity assays with high sensitivity. On the other hand, the capacity of transformation assays to identify many non-genotoxic carcinogens does not mean that *in vivo* carcinogenicity studies can be replaced by *in vitro* cell transformation assays. Examples of Ames-test negative rodent carcinogens with negative results in cell transformation assays are TCDD, ethinyl estradiol and methapyrilene (OECD 2001). These examples illustrate that quite relevant carcinogens can be tested false negative in cell transformation assays. In addition, false negative results in transformation assays may be due to inadequate metabolic activation. Taken together, although genotoxic or mutagenic substances can be identified by *in vitro* tests a complex long-term adverse effect such as carcinogenicity cannot sufficiently be predicted at present by non-animal tests. Still, the issue of potency remains unresolved, and potency may be a major driver in risk assessment.

Pattern based technologies such as toxicogenomics have been proposed for the prediction of the outcome of long-term carcinogenicity studies and other endpoints in long-term studies by short-term animal tests (Corvi et al. 2006; OECD 2005b). This is based on the expectation that patterns of gene expression deregulation specific for carcinogens can be identified. Ellinger-Ziegelbauer et al. (2005, 2008) published gene expression data from rats exposed to four non-genotoxic hepatocarcinogens (methapyrilene, diethylstilbestrol, Wy-14643 and piperonylbutoxide) and four genotoxic carcinogens (2-nitrofluorene, dimethylnitrosamine, NNK and aflatoxin B₁) for up to 14 days and identified substance specific alterations in gene expression patterns. For instance, the genotoxic carcinogens induced predominantly genes involved in DNA damage response, apoptosis and survival signalling. The non-genotoxic substances predominantly deregulated genes related to signal transduction pathways in cell cycle progression and response to oxidative DNA damage. Usually, a single gene or pathway will be insufficient to assign a specific mechanism of carcinogenicity. However, specific patterns of pathway-associated genes did allow a correct assignment of the tested substances to the groups of “genotoxic” or “non-genotoxic” rat carcinogens (Ellinger-Ziegelbauer et al. 2005). In recent years characteristic RNA expression signatures have been described for numerous toxic and/or carcinogenic chemicals, including formaldehyde (Sul et al. 2007), peroxisome proliferators (Tamura et al. 2006), 2-acetylaminofluorene, 2-nitropropane, 2-nitro-*p*-phenylenediamine, 2,4-diaminotoluene (Nakayama et al. 2006), *N*-ethyl-*N*-nitrosourea (Okamura et al. 2004), ochra-

toxin A (Arbillaga et al. 2007), aristolochic acid (Stemmer et al. 2006), and flumequine (Kashida et al. 2006). However, prospective and comprehensive validation studies aiming at the identification of carcinogens by predefined rules based on RNA expression signatures are not yet available. The latter would be a precondition to replace carcinogenicity studies by *in vivo* studies with shorter exposure periods and smaller numbers of animals, where test substances will be classified by alterations of RNA expression patterns in the respective target organs.

REACH legislation assumes mutagenic chemicals (M1 and M2) to be carcinogenic without demanding proof from a conventional carcinogenicity study or from alternative methods. Concerning non-genotoxic carcinogens the situation is problematic. Only for chemicals in the band of $\geq 1,000$ tons per year, the Agency can propose carcinogenicity studies for non-genotoxic chemicals in case there is evidence for extensive exposure of man or for the occurrence of preneoplastic changes in animal experiments. For chemicals in the bands $< 1,000$ tons per year, no assays for identification of non-genotoxic carcinogens are required according to REACH. This leads to a problematic situation because non-genotoxic carcinogens may be missed. Of course it is not practicable to perform 2-year carcinogenicity studies for all chemicals ≥ 10 to $< 1,000$ tons per year. On the other hand there are no validated *in vitro* tests for the identification of non-genotoxic carcinogens. As mentioned above, *in vitro* transformation assays will identify some but not all non-genotoxic carcinogens. This scenario illustrates that additional research is required. It should be clarified to which percentage non-genotoxic carcinogens may be missed by subchronic toxicity studies in case of chemicals ≥ 100 to $< 1,000$ tons or by subacute toxicity studies in case of chemicals ≥ 10 to < 100 tons per year. We encourage the development of *in vivo* studies with shorter exposure periods and smaller numbers of animals, where test substances will be classified by pattern-based methods in the respective target organs. In addition, pattern-based *in vitro* assays may play an increasing role if these assays will be successfully validated.

Reproductive and developmental toxicity

According to Annex VIII of REACH, the standard requirement is a screening test on reproductive toxicity (OECD TGs 421 or 422) at the 10 tons level per year. A developmental toxicity study (OECD TG 414; EU Method B.31 in Annex V of Directive 67/548/EEC) is mandatory for all compounds at the 100 tons level and higher (Annex IX). Starting from 100 tons per year, the two-generation test (OECD TG 416; EU Method B.35) is required if there are indications of potential reproductive toxicity from a repeated dose toxicity study. In the absence of indications of

potential reproductive toxicity, the two-generation study is only required for substances that are manufactured or imported in quantities of 1,000 tons or more per year. The current status of reproductive testing and assessment has been described and discussed in a recent Draft Guidance Document of the OECD (2007c). Some of the current tests require large numbers of animals. In particular, the two-generation test may require 2,600 animals (including the offspring generated in these studies). Thus, by far the largest proportion of animals in safety testing as required by REACH will be needed for testing of reproductive and developmental toxicity (Pedersen et al. 2003; van der Jagt et al. 2004; Höfer et al. 2004). Hence, the efforts in developing alternative tests are in particular directed to this area. The complexity of the reproductive cycle in mammals might be the main reason why success is limited up to now. Three *in vitro* models, the embryonic stem cell test (EST), the micromass test (MM) and the rat postimplantation whole embryo culture test (WEC) have been formally validated and recommended as screening tests for developmental toxicity testing (see Table 3). These tests can, however, cover only certain aspects of developmental toxicity. Generally, all three assays focus on the endpoint embryotoxicity, and each of the tests employs a limited window of mammalian reproduction. The EST represents specific cellular differentiation pathways (mainly cardiomyocyte differentiation), the MM test focuses on the differentiation of limb bud cells into cartilage-producing chondrocytes and the WEC assay constitutes a complex developmental system with a very narrow period of embryo-fetal development (Spielmann et al. 2006). The predictive performance of the above-mentioned validated assays for the compound classes tested was considered good and mainly suited for distinguishing between strong and non/weak embryotoxicants (Genschow et al. 2004). However, there is uncertainty on the broad applicability of these tests for a wide diversity of chemical compounds, since relevant classes of industrial chemicals were not included, only a limited number of toxicological mechanisms were represented and the inclusion of biotransformation for detection of indirect acting compounds was deferred to further validation studies. Hence it was recommended to improve the EST by supplementing it with a suitable *in vitro* metabolization system (Pellizzer et al. 2005; Coecke et al. 2006a). Since embryonic stem cells possess the capacity to differentiate *in vitro* into a variety of cell lineages, additional tissue-specific endpoints which are considered to be relevant for the test compound can be used (zur Nieden et al. 2004; Seiler et al. 2006). However, such approaches will require further validation. In conclusion, three validated *in vitro* assays for the determination of an embryotoxic potential of a test compound have become available. However, up to now the application of these tests for regulatory purposes is limited as no corresponding guideline is in place. At the

time being they mainly represent valuable tools for the clarification of mechanistic aspects. Furthermore they are used in some laboratories of the pharmaceutical industry for screening of drug candidates for possible embryotoxicity (Whitlow et al. 2007). Accordingly, the outcome of these assays might be used to select industrial chemicals for which follow-up *in vivo* testing is still considered necessary. It is conceivable that a compound belonging to a well-known chemical class and producing a result in such assays which conforms to the expectations may not need further *in vivo* embryotoxicity assessment.

An integrated research project, ReProTect, started in July 2004 aimed at building a predictive test strategy in the field of reproductive toxicology including endocrine disruption. This project is aiming at generating information which is not obtainable from the three validated *in vitro* tests (for reviews see Hareng et al. 2005; Bremer et al. 2005). It will last 5 years and has been divided in several research areas, namely male and female fertility, implantation and prenatal development. Newly developed and existing *in vitro* tests addressing different aspects of the reproductive cycle are planned to be integrated into an alternative testing strategy. A fourth research area, “cross-cutting technologies”, is aimed to develop concepts for the validation of innovative test systems, namely murine and human embryonic stem cells or reporter gene-based tests using genetically engineered cells. Furthermore, it is envisaged to implement novel technologies such as sensor technologies, (Q)SARs and pattern-based methods (e.g. toxicogenomics, proteomics and metabonomics) and to evaluate their use for toxicological safety testing.

Several tests have been established to evaluate male/female fertility, implantation, endometrial function and placental toxicity. For early and late development of the embryo/fetus, two of the above-mentioned existing tests (EST, WEC, see Table 3) are to be optimised and co-culture systems to be developed aimed at detecting teratogens acting via their metabolites. Further assays are directed to the identification of endocrine disruption (e.g. via androgenic or estrogenic activity). Most of the tests are at the R&D level, some of the advanced tests are in a prevalidation phase. One test for evaluating the ability of chemicals to bind to the estrogen receptor is envisaged for validation at the level of the OECD (for a more detailed description of the *in vitro* tests and their toxicological background see Hareng et al. 2005; for the particular conditions and testing strategy using embryonic stem cells see Bremer and Hartung 2004).

It is recognised that the project is focused on sensitive endpoints and critical events and stages of the mammalian reproduction cycle but even a complete set of 24 (or more) validated *in vitro* tests will not be capable of covering all relevant aspects of toxicity on fertility and development as for example postnatal development is not assessed yet.

Therefore, combined testing strategies including *in vitro* methods, (Q)SARs and other alternative methods and tools are planned to be developed. For instance, the *in vitro* tests are planned to be complemented by means of (Q)SARs focussing on the blood-testis barrier as well as the blood-placental barrier (Hewitt et al. 2007b). By use of combinations of these *in silico* approaches and physiologically based pharmacokinetic (PBPK) modelling (Verwei et al. 2006) it is expected to improve the predictive value of *in vitro* reproductive toxicity testing in risk assessment. Although single tests may achieve validation status within some years, it is not possible to suggest a time limit for achieving an appreciable reduction or even the full replacement of animal tests by *in vitro* tests, non-testing methods and combinations thereof. In addition, effects on some life stages such as postnatal development either induced prenatally or postnatally cannot be assessed *in vitro* or by non-testing methods but only *in vivo*, if any.

In terms of refinement and reduction, recently several reviews have been published evaluating the testing strategy and the results from *in vivo* studies conducted according to current testing guidelines on reproductive toxicity, in particular the two-generation studies, which are costly and time-intensive and require relatively large numbers of animals. Mangelsdorf et al. (2003) have reviewed previous reports and concluded that testes histopathology was the most sensitive endpoint to detect adverse effects on male reproduction. Indicators such as reproductive organ weights and sperm parameters also showed higher sensitivity than conventional fertility parameters (e.g. number of implantations per female). Testes weight and histopathology evaluations are required in the rat oral 28-day study (OECD Test Guideline 407); and testis, epididymis, uterus, and ovary weight and histopathology evaluations are required in the rat subchronic study (OECD Test Guidelines 408, 411, and 413). It was shown that the lesions in male reproductive organs can be detected in most cases after 2 weeks of treatment (Sakai et al. 2000) suggesting that the duration of repeated dosing in a subacute or subchronic study could be sufficient in most cases to detect effects on testes weight and histopathology by a substance exerting reproductive toxicity to males.

Based on similar considerations the “enhanced” 28-day repeated dose toxicity study in rats (according to OECD Guideline 407) has been shown to detect compounds with a moderate to strong potential to affect the gonads or the thyroid through disturbance of the endocrine regulation (Gelbke et al. 2007; OECD 2006a, b). If a thorough histopathologic examination of the gonads is conducted within such a study, compounds affecting major aspects of male or female fertility will be identified. Similarly, the absence of relevant findings in repeat-dose toxicity studies and in a prenatal developmental toxicity study together with a negative sex

hormone receptor assay was considered as an argument to lower the priority of *in vivo* fertility studies (Dent 2007). Accordingly, one should consider whether a combination of the above-mentioned 28-day test with an *in vitro* embryotoxicity screen, will be adequate for the characterization of reproductive toxicity for most chemicals, once the problem of lacking metabolising capacity will have been solved,

Using a different approach, Vermeire et al. (2007) have recently constructed a database consisting of rat subchronic (90 days), rat reproductive (two-generations) and rat and rabbit developmental toxicity studies for substances classified as toxic to fertility or development. These authors aimed at comparing the results of these different study types and focused on the added value to hazard and risk assessment of the two-generation reproductive toxicity study when a subchronic study was available. The impact of the second generation on risk assessment within the two-generation reproductive toxicity study and the added value to risk assessment of the rabbit developmental study was analysed when a rat developmental study was available. As a first outcome, Janer et al. (2007) have published a retrospective analysis of 176 multi-generation studies predominantly conducted with pesticides but including also about 40 industrial chemicals. The collection comprised 58 studies with substances that are considered as toxic to reproduction according to EU or other international regulations. In 3 out of the 176 studies, reproductive toxicity was identified in the second generation but not in the first generation. However, for different reasons, none of these substances was classified as toxic on reproduction. Among the studies with substances classified as toxic on reproduction, six studies indicated lower NOAELs in F_1 animals compared to P_0 adults (five substances belong to the group of phthalates) and two studies indicated lower NOAELs in the F_2 offspring compared to F_1 animals. Of course, the limited number of studies available is not considered being representative for industrial chemicals in general. Nonetheless, the analysis raises concern about the ‘added value’ of the second generation with respect to the efficiency and/or the quality of the study design of the respective guideline and the testing strategy of the complete set of repeated dose studies.

From these and other reports it has been concluded that the current testing strategy for reproductive toxicity (but also systemic repeated dose toxicity) has to be re-considered. For instance, Cooper et al. (2006) have proposed a “Life Stages Tiered Testing Approach” for agrochemicals in which Tier 1 consists of a base set of studies that provides data on ADME, systemic toxicity in adults and in other life stages from development to adolescence, gestation and elderly including consideration of neurological, immunological and endocrine endpoints. A centrepiece of the testing paradigm is an extended one-generation study.

Certain triggers may induce a Tier 2 testing including, e.g. a two-generation study. Compared to the complete current testing scheme on reproductive toxicity, which according to Cooper et al. (2006) requires about 5,300 animals including a two-generation study, the proposed testing scheme requires about half the number of animals when conducting Tier 1 and 2 including a two-generation study. Although the validation of the proposed testing scheme may take many years, the authors demonstrate that there is a great potential in refinement and reduction of *in vivo* studies on reproductive toxicity.

Some notes on test batteries, tiered testing approaches and integrated testing strategies (ITS)

According to OECD (2005a), a test battery or tiered testing approach is a series of tests usually performed at the same time or in close sequence. Each test in the battery is selected so as to complement the other tests (e.g., to identify a different component of a multi-factorial effect) or to confirm another test. A battery of tests may also be arranged in a hierarchical (tiered) approach to testing. In a tiered approach, the tests are used sequentially; the tests selected in each succeeding level are determined by the results in the previous level of testing. At present this is a common approach for, e.g. ecotoxicity testing.

Comprehensive guidance on the validation of test batteries and tiered testing approaches has not been developed yet. However, a few general considerations are given in a Guidance Document (OECD 2005a). Individual tests within a battery of tests (or in a tiered testing approach) should be validated using the validation principles described in the Guidance Document, taking into consideration their restricted roles in the test battery/tiered testing approach. Justification for the acceptance of a test battery should be primarily on the basis of its overall performance for its intended purpose. The performance of a test battery or a tiered testing approach, i.e., its predictive capacity and ability to replace or reduce or refine the use of animals, may be evaluated by simulating possible outcomes *in vivo* using existing data.

Recent additional considerations that may help to avoid unnecessary testing are: (1) the assessment of human exposure and the application of negligible or no human exposure, e. g. according to the concept of Toxicological Thresholds of Concern (TTC, Kroes et al. 2004; ECB 2007) and (2) the use of non-testing methods for gaining data on toxicological endpoints such as grouping or the chemical category approach for chemicals similar in their structure or possessing similar properties including data gap filling techniques such as read-across, trend analysis, (Q)SARs and computational expert systems (ECB 2005, 2007). It has been proposed to develop and apply Integrated

(“Intelligent”) Testing Strategies (ITS) combining all the different information sources available. An Integrated Testing Strategy is by a definition by Blaauboer (2002) “any approach to the evaluation of toxicity that is based on the use of two or more of the following: physicochemical data, *in vitro* data, human data (for example, epidemiological data, clinical case reports on intoxication), animal data (where unavoidable) and computational methods such as (Q)SAR and biokinetic models” (see also Combes et al. 2003; Hartung et al. 2003; Combes and Balls, 2005; ECB 2005, 2007).

The RIP consortium recently published a general decision-making framework (GDMF) making use of endpoint-specific ITS. They recognised that the testing and data requirements for different endpoints are too different to be treated within a general ITS scheme. Instead they developed individual ITS schemes for the different endpoints within the GDMF. Treating different kinds of data for each endpoint in different steps, the procedure comes either to a classification or no classification in one of the steps or in case of inconclusive data to a decision point termed “weight of evidence (WoE) analysis”, the function of which is to decide whether the sum of inconclusive data is sufficient for classification or not. In the latter case, further (supplementary) testing *in vitro* or *in vivo* is required.

Although the practicality of the endpoint-specific ITS schemes was in general demonstrated, the RIP consortium identified some issues raising concern. Some of these issues are, for instance:

- Existing data were generated in a heterogeneous way in the past. In particular the quality of toxicological data and their evaluation is often a matter of concern and disagreement among experts despite the fact that there exist some guidance [e.g., quality criteria for checking the relevance, reliability and adequacy of the data described by Klimisch et al. (1997)]. The experience of adequate data evaluation is restricted to a low number of experts compared to the great amount of existing data available.
- The REACH Regulation (Annex XI) does not in detail specify the prerequisites for using non-test information as replacement for testing, but does only contain some general statements about this. The current status and some prerequisites for the grouping of the substances, use of the analogue or category approach and data-filling techniques are described in the Draft TGD (ECB 2007). Although some experience in the frame of the High Production Volume Programme (HPV) of the OECD exists (OECD 2007b), scientific valid criteria or guidance are lacking and need to be developed in particular when defining applicability domains.
- The scientifically adequate use of the weight of evidence analysis depends on (1) the correct identification of

inconclusive data and (2) the adequate evaluation of the sum of these data both of which is clearly a matter of expertise and experience. Thus, the question on the number of experienced experts in relation to the amount of work arises and has to be solved.

For risk assessment, alternative methods are at present suited only to a limited extent. In particular, the development of sets or combinations of complementary *in vitro* tests, other prediction tools such as *in silico* toxicology including (Q)SAR, toxicokinetic modelling and combinations thereof is only in a starting phase (e. g., Anderson 2003; Combes et al. 2003; IHCP 2005; Zuang and Hartung 2005; Eskes and Zuang 2005; ECB 2005, 2007; OECD 2004, 2005a, 2007a; Gubbels-van Hal et al. 2005; Blauboer and Anderson 2007). Dose-response characteristics are a central issue in risk assessment but it is at present unclear whether PBPK models will ever allow for the reliable extrapolation of dose-response data from *in vitro* experiments to the *in vivo* situation.

In conclusion, although there may be a great potential for using non-testing methods such as existing data, chemical analogue and chemical category approach/grouping or *in silico* methods in supplementing, reducing or replacing *in vivo* testing, these methods themselves or as parts of tiered or integrated testing schemes are at the present time not validated and need therefore to be applied and evaluated case by case. This requires toxicological knowledge, expertise and experience and it should be noted that this also holds true for data admission and evaluation in relation to the avoiding of unnecessary animal testing in general [Article 25 (1)] and the provisions for waiving according to Annex XI of REACH in particular.

Discussion: expectations and reality

Primacy of health protection versus animal welfare and costs

In the public, among protagonists of animal welfare and scientists, there is much concern about the anticipated high numbers of laboratory animals required for safety studies under REACH. However, it must be stated that the main reason for this apparently spectacular high number of animals needed has been congestion in producing toxicological data for chemicals, which has been occurring for two decades or more due to an insufficient regulation of chemicals in the EU. Furthermore, the perception and understanding of the role of *in vitro* tests and other alternative methods in the process of hazard identification and risk assessment of chemical substances is extremely different between the scientific community, stakeholders such as animal welfare

groups, politicians and the public. On one hand, the anticipations of animal welfare groups and other stakeholders concerning the potential of alternative methods for hazard identification and risk assessment are often very high. On the other hand, in the scientific community there is for the most part only some cautious optimism regarding this issue. Arguments from the industry concerning possible cost reduction by avoiding *in vivo* tests support the position of animal welfare groups although this is not necessarily intended. The political discussion appears to primarily care for animal welfare and cost reduction and to supplant considerations on the adequacy of dispensing animal experimentation for the protection of human health and environment within the time frame of REACH. We are concerned about this development. Similarly, the Scientific Committee on Toxicity, Ecotoxicity and the Environment of the EU Commission (CSTEE 2004) has recently addressed this bias: “Toxicological testing aims to predict possible adverse effects in humans when exposed to chemicals. Currently it is extensively based on animal testing to identify hazards and the dose-response relationships of chemicals. Ethical concerns have been raised by the use of laboratory animals. However, independent of ethical concerns, the primary objective of the risk assessment of chemical exposures is the protection of human health, wildlife and ecosystems”.

The primacy of health protection requires thorough examination of novel toxicological methods. This does not mean that animal experimentation according to OECD guidelines is thought to be an everlasting gold standard, which can never be overtaken by any novel approach of hazard identification and risk assessment. Rather, the advantages and limitations of *in vivo* testing and alternative methods have to be compared impartially with respect to the possible contribution of each of the methods to the detection or elucidation of toxic effects of a substance. One additional argument that has to be taken into account with regard to risk assessment is that use of all *in vitro* testing in the regulatory context requires even more extrapolation and “unsafety factors” when introduced to protect human health. Recently, we have discussed the possible positive impact of REACH on toxicological sciences. For instance, we have suggested that toxicological data generated under REACH may offer a unique chance to develop and validate techniques and methods that predict toxicity faster and more precisely than the conventional techniques and we have suggested strategies how an integrated scientific research programme could allow an enormous progress in toxicological sciences (Hengstler et al. 2006). Furthermore we initiated and contributed to a publication on the possible impact of (Q)SARs as alternative methods on improving risk assessment (Simon-Hettich et al. 2006).

We recognise that there has been considerable progress in the development of alternative methods for classification and labelling, e.g. for local toxicity and genotoxicity. However, modulation of a systemic toxic effect by the interplay of the complete organism cannot be exhaustively imitated *in vitro* or *in silico*, at least before the availability of powerful systems biology. Accordingly, it appears questionable whether acute toxicity testing can eventually do without *in vivo* experimentation. In other complex areas of toxicology such as repeated dose toxicity, reproductive toxicity and carcinogenicity, the situation is even less promising in our opinion. Certain features of long-term toxicity such as the slow replacement of parenchyma cells by the connecting tissue or the contribution of inflammation can at present best be determined by exposure of the integral animal. In addition, metabolites of the compound to be tested become increasingly important in long-term toxicity. As discussed above, the inclusion of metabolism in alternative methods is at present not yet satisfactory.

The time scale aspect in the context of REACH

The time scale aspect of the introduction of alternative methods is expected to be more critical the more complex the toxicological issue is and the more heterogeneous the tests contributing to an integrated test system are. It is assumed that using a set of different *in vitro* tests or other alternative methods with different reliability and relevance of the tests will result in a higher complexity of the validation procedure. Are the current validation procedures suited to cope with more complex demands? And if so, how much time will be required to perform the validation of a set of complementary *in vitro* tests, not to mention the validation of a complete integrated testing system, and to achieve its regulatory acceptance? Will the approach of Weight of Evidence (WoE) validation assessment be an appropriate and credible last resort? In the context of REACH and the seventh amendment to the Cosmetics Guideline, there is a fundamental drawback with respect to the abandonment of animal experimentation: Their time frame does not match the time frame for the expected supply of alternative methods for complex endpoints. Substances that are produced or imported in amounts of more than 100 tons per year require data on subacute/subchronic toxicity and on reproductive toxicity for their registration (see Table 2). *In vivo* tests consume a very high number of test animals. However, the time schedule of REACH assigns only 3–9 years for the registration of these substances and for proposals and decisions on additional testing *in vivo*. Therefore, it is foreseeable that the development, validation, scientific and regulatory acceptance of alternative methods will need too much time to contribute significantly to the reduction, refinement or replacement of *in vivo* tests. As has been

pointed out by Combes (2007), Annex XI, section 1.4 appears to cut the standards for acceptance of *in vitro* methods when making possible the use of results of tests that have not yet been scientifically validated but are identified as being “suitable”, i.e. meeting the ECVAM criteria for entry of the method into the pre-validation process. The phasing out date, March 2013, of the seventh amendment to the Cosmetics Guideline 76/768/EEC for repeated dose toxicity, reproductive toxicity and carcinogenicity appears too optimistic for the same reasons and in a similar manner as the REACH time frame. Hence, instead of trying to meet these time frames, the strategy of the development of alternative methods should be more directed in the current stage to the refinement and the reduction of *in vivo* tests instead of replacement as it has been pointed out by Prieto et al. (2006).

The complete replacement of animal testing for safety purposes including new and improved testing strategies that meet future challenges of toxicological risk assessment can be considered as a vision. For instance, the Committee on Toxicity and Assessment of Environmental Agents on behalf of the National Research Council has developed such a vision and a strategy for ‘Toxicity Testing in the Twenty-first Century’ (NAS 2007). They consider their vision and strategy as a long-term paradigm shift in toxicological testing and conclude that implementing the vision will require improvements and focused effort over a period of decades.

Conclusions

1. We are confident that during the REACH process large quantities of data, which can be utilized for the improvement of alternative methods, will be generated provided that suitable measures are established to guarantee standardised data quality. We call for the rapid implementation of high quality REACH databases enabling access by scientists involved in work on alternative methods.
2. Complex toxicological endpoints such as acute, subacute and chronic toxicity, sensitisation, carcinogenicity and reproductive toxicology can at present not be covered by *in vitro* methods and it is questionable whether this will be possible in the foreseeable future as such endpoints need the interplay of a complete organism. However, alternative testing and non-testing methods have a potential to reduce the number of animals used through their inclusion into integrated testing strategies.
3. Considering the time frame of REACH (2012 for production volumes of phase-in substances $\geq 1,000$ tons per year, 2016 for ≥ 100 tons per year, Table 1), the

development of alternative methods will not be sufficiently rapid to markedly lower the use of animal experiments for toxicological testing. We caution against negotiating this obstacle by using premature alternative testing tools (Annex XI, section 1.4 of REACH). The integration of alternative testing methods into the toxicological assessment process requires highly qualified expert judgement until fully developed and validated methods are available for routine use and beyond.

4. Registration under REACH as well as the improvement of alternative methods, the development of validation procedures for new and integrated testing strategies, the expert judgement on the inclusion of alternative methods in testing and the maintenance of REACH data-bases require the broad availability of toxicological expertise. We call for the preservation and expansion of educational capacities for toxicologists in academia, industry and authorities.

Acknowledgements We thank Dr. Rüdiger Bartsch, Technical University of Munich, Dr. Charlotte Esser, University of Düsseldorf, Dr. Klaus Becker, Dr Simone Müller, Bayer Schering Pharma AG, Berlin, Prof. Dr. Hans-Werner Vohr, Bayer AG, Wuppertal for valuable contributions and helpful discussions.

References

- Abbot A (2005) More than a cosmetic change. *Nature* 430:144–146
- Amberg A, Rosner E, Dekant W (2001) Toxicokinetics of methyl tert-butyl ether and its metabolites in humans after oral exposure. *Toxicol Sci* 61:62–67
- Anderson ME (2003) Toxicokinetic modelling and its application to in chemical risk assessment. *Toxicol Lett* 138:9–27
- Arbillaga L, Azqueta A, van Delft JH, López de Cerain A (2007) In vitro gene expression data supporting a DNA non-reactive genotoxic mechanism for ochratoxin A. *Toxicol Appl Pharmacol* 220:216–224
- Balls M (ed) (2002) Alternatives to animal experiments: Progress made and challenges ahead. In: The Proceedings of the ECVAM Status Seminar, held on 4–6 June 2002, in the Amphitheatre at the European Commission Joint Research Centre (JRC), Ispra, Italy. *Altern Lab Anim* 30(Suppl. 2):1–243
- Balls M, Blaauboer B, Brusick D, Frazier J, Lamb D, Pemberton M, Reinhardt C, Roberfroid M, Rosenkranz H, Schmid B, Spielmann H, Stamatii A-L, Walum E (1990) Report and recommendations of the CAAT/ERGATT workshop on the “Validation of Toxicity Test Procedures”. *Altern Lab Anim* 18:313–337
- Balls M, Blaauboer BJ, Fentem JH (1995) Practical aspects of the validation of toxicological test procedures. *Altern Lab Anim* 23:129–147
- Balls M, Amcoff P, Bremer S, Casati S, Coecke S, Clothier R, Combes R, Corvi R, Curren R, Eskes C, Fentem J, Gribaldo L, Halder M, Hartung T, Hoffmann S, Schechtman L, Scott L, Spielmann H, Stokes W, Tice R, Wagner D, Zuang V (2006). The principles of weight of evidence validation of test methods and testing strategies. *Altern Lab Anim* 34:603–620
- Bergström CAS (2005) In silico predictions of drug solubility and permeability: two rate-limiting barriers to oral drug absorptions. *Basic Clin Pharmacol Toxicol* 96:156–161
- Bernaer U., Birner G, Dekant W, Henschler D. (1996) Biotransformation of trichloroethene: dose-dependent excretion of 1,1,1-trichloro-metabolites and mercapturic acids in rats and humans after inhalation. *Arch Toxicol* 70:338–346
- Bernaer U, Oberemm A, Madle S, Gundert-Remy U (2005) The use of in vitro data in risk assessment. *Basic Clin Pharmacol Toxicol* 96:176–181
- Blaauboer BJ (2002) The necessity of biokinetic information in the interpretation of in vitro toxicity data. *Altern Lab Anim* 30(Suppl. 2):85–91
- Blaauboer BJ, Anderson ME (2007) The need for a new toxicity testing and risk analysis paradigm to implement REACH or any other large scale testing initiative. Letter to the editor *Arch Toxicol* 81:385–387
- Bolt HM, Foth H, Hengstler JG, Degen G (2004) Carcinogenicity categorization of chemicals—new aspects to be considered in a European perspective. *Toxicol Lett* 151:29–41
- Botham PA (2004) Acute systemic toxicity—prospects for tiered testing strategies. *Toxicol In Vitro* 18:227–230
- Bremer S, Hartung T (2004) The use of embryonic stem cells for regulatory developmental toxicology testing in vitro. *Curr Pharm Des* 10:2733–2747
- Bremer S, Balduzzi D, Cortvrindt R, Daston G, Eletti B, Galli A, Huh-taniemi I, Laws S, Lazzari G, Liminga U, Smits J, Spano M, Themmen A, Tilloy A, Waalkens-Behrends I (2005) The effects of chemicals on mammalian fertility. *Altern Lab Anim* 33:391–416
- Casati S, Aeby P, Basketter DA, Cavani A, Gennari A et al (2005) Dendritic cells as a tool for the predictive identification of skin sensitisation hazard. *Altern Lab Anim* 33:47–62
- Coecke S, Ahr H, Blaauboer BJ, Bremer S, Casati S et al (2006a). Metabolism: a bottleneck in in vitro toxicological test development. *Altern Lab Anim* 34:49–84
- Coecke S, Eskes C, Gartlon J et al (2006b) The value of alternative testing for neurotoxicity in the context of regulatory needs. *Environ Toxicol Pharmacol* 21:153–167
- Coecke S, Goldberg AM, Allen S, Buzanska L, Calamandrei G, Crofton K, Hareng L, Hartung T, Knaut H, Honegger P, Jacobs M, Lein P, Li A, Mundy W, Owen D, Schneider S, Silbergeld E, Reum T, Trnovec T, Monnet-Tschudi F, Bal-Price A (2007) Workgroup report: incorporating in vitro alternative methods for developmental neurotoxicity into international hazard and risk assessment strategies. *Environ Health Perspect* 115:924–931
- Combes R (2007) Why ‘suitable’ in vitro methods, as defined in the final EU REACH legislation, are an inappropriate basis for risk assessment. *Altern Lab Anim* 35:289–293
- Combes R, Balls M (2005) Intelligent testing strategies for chemicals testing—a case for more haste, less speed? *Altern Lab Anim* 33:289–297
- Combes R, Barratt M, Balls M (2003) An overall strategy for the testing of chemicals for human hazard and risk assessment under the EU REACH system. *Altern Lab Anim* 31:7–19
- Cooper RL, Lambs JC, Barlow SM et al (2006) A tiered approach to life stages testing for agricultural chemical safety assessment. *Crit Rev Toxicol* 36:69–98
- Corvi R, Ahr HJ, Albertini S, Blakey DH, Clerici L et al (2006) Validation of toxicogenomics-based test systems: ECVAM-ICC-VAM/NICEATM considerations for regulatory use. *Environ Health Perspect* 114:420–429
- CSTEE (2004) Opinion of the Scientific Committee on Toxicity, Ecotoxicity and the Environment on the BUAV-ECEAE report on “the way forward—Action to end animal toxicity testing”. Adopted by the CSTEE during the 41st plenary meeting of 8 January 2004. (C7/VR/csteop/anat/080104 D(04) (http://europa.eu.int/comm/health/ph_risk/committees/sct/documents/out217_en.pdf
- Dent MP (2007) Strengths and limitations of using repeat-dose toxicity studies to predict effects on fertility. *Regul Toxicol Pharmacol* 48:241–258

- ECB (2005) European Chemicals Bureau (ed). TAPIR Three point three—A Project for the Information Requirements of REACH. Final Report—2 August 2005. Scoping study on the development of a Technical Guidance Document on information requirements on intrinsic properties of substances (RIP 3.3–1). <http://ecb.jrc.it/reach/rip>
- ECB (2007) European Chemicals Bureau (ed). Technical Guidance Document to industry on the information requirements for REACH. Part 1–4. Final Draft for review. REACH Implementation Project (RIP) 3.3, Phase 2. <http://ecb.jrc.it/reach/rip>
- Ellinger-Ziegelbauer H, Stuart B, Wahle B, Bomann W, Ahr HJ (2005) Comparison of the expression profiles induced by genotoxic and non-genotoxic carcinogens in rat liver. *Mutat Res* 575:61–84
- Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A, Ahr HJ (2008) Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat Res* 637:23–39. Epub 5 July 2007
- Ernstgard L., Shibata E, Johanson G (2005) Uptake and disposition of inhaled methanol vapor in humans. *Toxicol Sci* 88:30–38
- Eskes C, Zuang V (eds) (2005) Alternative (non-animal) methods for cosmetics testing: current status and future prospects. *Altern Lab Anim* 33(Suppl 1):1–223
- EU (1986) Council Directive 86/609/EEC on the approximation of laws. Regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. *Off J Eur Comm L* 358, pp 1–29
- EU (2002) Corrigendum to the Commission communication pursuant to Article 13 of Council Directive 67/548/EEC of 27 June 1967 on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances, as amended by Directive 79/831/EEC—EINECS (European Inventory of Existing Commercial Chemical Substances) (Off J Eur Comm C 146 A of 15.6.1990). *Off J Eur Comm C* 54, 1 March 2002, pp 13–73
- EU (2003) Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products. *Off J Eur Comm L* 066, 11 March 2003, pp 26–35
- EU (2006) Regulation (EC) 1907/2006 of the European Parliament and the European Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Regulation 1999/45/EC and repealing Council Regulation (EEC) No 93/793 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/677/EEC, 93/105/EEC and 2000/21/EC. *Off J Eur Comm L* 369, 30 December 2006, pp 1–849
- European Commission (ed) (2006) Evaluation and predictions of dermal absorption of toxic chemicals (EDET0X). Contract No. QLK4-2000-00196, p 61. <http://www.ncl.ac.uk/edetox/current-position.html>
- Fennell TR, Sumner SC, Snyder RW, Burgess J, Spicer R, Bridson WE, Friedman MA (2005) Metabolism and hemoglobin adduct formation of acrylamide in humans. *Toxicol Sci* 85:447–459
- Filser JG, Denk B, Trnqvist M, Kessler W, Ehrenberg L (1992) Pharmacokinetics of ethylene in man; body burden with ethylene oxide and hydroxyethylation of hemoglobin due to endogenous and environmental ethylene. *Arch Toxicol* 66:157–163
- Forsby A, Blaauboer B (2007) Integration of in vitro neurotoxicity data with biokinetic modelling for the estimation of in vivo neurotoxicity. *Human Exp Toxicol* 26:333–338
- Garberg P, Ball M, Borg N, Cecchelli R, Fenart L, Hurst RD, Lindmark T, Mabondzo A, Nilsson JE, Raub TJ, Stanimirovic D, Terasaki T, Oberg JO, Osterberg T (2005) In vitro models for the blood-brain barrier. *Toxicol In Vitro* 19:299–334
- Gartlon J, Kinsner A, Bal-Price A, Coecke S, Clothier RH (2006) Evaluation of a proposed in vitro test strategy using neuronal and non-neuronal cell systems for detecting neurotoxicity. *Toxicol In Vitro* 20:1569–1581
- Gebhardt R, Hengstler JG, Müller D, Glöckner R, Buehning P, Laube B, Schmelzer E, Martina U, Utesch D, Hewitt N, Ringel M, Reder Hiltz B, Bader A, Langsch A, Koose T, Burger HJ, Maas J, Oesch F (2003) New hepatocyte in vitro systems for drug metabolism: Metabolic capacity and recommendations for application in basic research and drug development, standard operation procedures. *Drug Metab Rev* 35:145–213
- Gelbke HP, Hofmann A, Owens JW, Freyberger A (2007) The enhancement of the subacute repeat dose toxicity test OECD TG 407 for the detection of endocrine active chemicals: comparison with toxicity tests of longer duration. *Arch Toxicol* 81:227–250
- Gennari A, van den Berghe C, Casati S, Castell J, Clemenson C, Coecke S, Colombo A, Curren R, Dal Negro G, Goldberg A, Gosmore C, Hartung T, Langezaal I, Lessigiarska I, Maas W, Mangelsdorf I, Parchment R, Prieto P, Sintes JR, Ryan M, Schmuck G, Stützel K, Stokes W, Vericat JA, Gribaldo L (2004) Strategies to replace in vivo acute systemic toxicity testing. The report and recommendations of ECVAM Workshop 50. *Altern Lab Anim* 32:437–459
- Genschow E, Spielmann H, Scholz G, Pohl I, Seiler A, Clemann N, Bremer S, Becker K (2004) Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32:209–244
- Gerner I, Zinke S, Graetschel G, Schleder E (2000) Development of a decision support system for the introduction of alternative methods into irritancy/corrosivity testing strategies: creation of fundamental rules for a decision support system. *Altern Lab Anim* 28:665–698
- Gerner I, Schlegel K, Walker JD, Hulzebos E (2004) Use of physico-chemical property limits to develop rules for identifying chemical substances with no skin corrosion or irritation potential. *QSAR Comb Sci* 23:726–733
- Gerner I, Liebsch M, Spielmann H (2005) Assessment of the eye irritation properties of chemicals by applying alternatives to the Draize rabbit eye test: the use of QSARs and in vitro tests for the classification of eye irritation. *Altern Lab Anim* 33:215–237
- Greim H, Arand M, Autrup H, Bolt HM, Bridges J, Dybing E, Glomot R, Foa V, Schulte-Herrmann R (2006) Toxicological comments on the discussion about REACH. *Arch Toxicol* 80:121–124
- Gubbels-van Hal WLMG, Blaauboer BJ, Barentsen HM, Hoitink MA, Meerts EATM, van der Hoeven JCM. (2005) An alternative approach for the safety evaluation of new and existing chemicals, an exercise in integrated testing. *Regul Toxicol Pharmacol* 42:284–295
- Gunnare S., Ernstgard L., Sjogren B., Johanson G. (2006) Toxicokinetics of 1,1,1,2-tetrafluoroethane (HFC-134a) in male volunteers after experimental exposure. *Toxicol Lett* 167:54–65
- Halle W (2003) The Registry of Cytotoxicity: toxicity testing in cell cultures to predict acute toxicity (LD50) and to reduce animal testing. *Altern Lab Anim* 31:89–198
- Hareng L, Pelizzer S, Schwarz M, Hartung T (2005) The integrated project ReProTect: a novel approach to reproductive hazard assessment. *Reprod Toxicol* 20:441–452
- Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Roi AJ, Prieto P, Sabbioni E, Worth A, Zuang V (2003) ECVAM's response to the changing political environment for alternatives: consequences of the European Union chemicals and cosmetics policies. *Altern Lab Anim* 31:473–481
- Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, Prieto P, Sabbioni, Scott L, E, Worth A, Zuang V (2004) A modular approach to the ECVAM Principles on test validity. *Altern Lab Anim* 32:467–472
- Hengstler JG, Utesch D, Steinberg P, Ringel M, Swales N, Biefang K, Platt KL, Diener B, Böttger T, Fischer T, Oesch F (2000) Cryo-

- preserved primary hepatocytes as an in vitro model for the evaluation of drug metabolism and enzyme induction. *Drug Metab Rev* 32:81–118
- Hengstler JG, Bogdanffy MS, Bolt HM, Oesch F (2003) Challenging dogma: thresholds for genotoxic carcinogens? The case of vinyl acetate. *Annu Rev Pharmacol Toxicol* 43:485–520
- Hengstler JG, Foth H, Kramer PJ, Kahl R, Liliënblum W, Schulz T, Schweinfurth H (2006). The REACH concept and its impact on toxicological sciences. *Toxicology* 220:232–239
- Hewitt NJ, Lechon MJ, Houston JB, Hallifax D, Brown HS, Maurel P, Kenna JG, Gustavsson L, Lohmann C, Skonberg C, Guillouzo A, Tuschl G, Li AP, LeCluyse E, Groothuis GM, Hengstler JG (2007a) Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. *Drug Metab Rev* 39:159–234
- Hewitt M, Madden JC, Rowe PH, Cronin MTD (2007b) Structure-based modelling in reproductive toxicology: (Q)SARs for the placental barrier. *SAR QSAR Environ Res* 18:57–76
- Höfer T, Gerner I, Gundert-Remy U, Liebsch M, Schulte A, Spielmann H, Vogel R, Wettig K (2004) Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation. *Arch Toxicol* 78:549–564
- Huang Q, Jin X, Gaillard ET, Knight BL, Pack FD, Stoltz JH, Jayadev S, Blanchard KT (2004) Gene expression profiling reveals multiple toxicity endpoints induced by hepatotoxicants. *Mutat Res* 549:147–167
- Hulzebos ET, Walker JD, Gerner I, Schlegel K (2005) Use of structural alerts to develop rules for identifying chemical substances with skin irritation or skin corrosion potential. *QSAR Comb Sci* 24:232–242
- ICCVAM (1997) Validation and regulatory acceptance of toxicological test methods: a report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No: 97-3981, March 1997, National Institute of Environmental Health Sciences, (NIEHS), Research Triangle Park, NC, USA, p 105. <http://iccvam.niehs.nih.gov/>
- ICCVAM/NICEATM (2001a) US Interagency Coordinating Committee on the Validation of Alternative Methods and Interagency Center for the Evaluation of Alternative Toxicological Methods. Guidance document on using in vitro data to estimate in vivo starting doses for acute toxicity. NIH Publication No. 01-4500. <http://iccvam.niehs.nih.gov/>
- ICCVAM/NICEATM (2001b) Report of the international workshop on in vitro methods for assessing acute systemic toxicity. Results of an international workshop organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). August 2001. NIH Publication No. 01-4499. <http://iccvam.niehs.nih.gov/>
- ICCVAM/NICEATM (2006) ICCVAM Test method evaluation report of in vitro cytotoxicity test methods for estimating starting doses for acute oral systemic toxicity testing. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), November 2006, NIH Publication No. 07-4519. <http://iccvam.niehs.nih.gov/>
- IHCP (2004) Institute of Health and Consumer Protection, Joint Research Centre of the EU Commission, Ispra, Piedmont, Italy. Alternative methods to animal testing. Annual Review 2004, Report EUR 21665 EN, pp 19–26. <http://ihcp.jrc.cec.eu.int/press/activity.htm>
- IHCP (2005) Institute of Health and Consumer Protection, Joint Research Centre of the EU Commission, Ispra, Piedmont, Italy. REACH and the need for intelligent test strategies. <http://139.191.1.25/docs/20051107its.pdf>
- IPCS (2006) International Programme on Chemical Safety. Dermal absorption. Environmental Health Criteria No 235. World Health Organization, Geneva. <http://www.who.int/ipcs/en/>
- Janer G, Hakkert BC, Slob W, Vermeire T, Piersma AH (2007) A retrospective analysis of the two-generation study: what is the added value of the second generation? *Reprod Toxicol* 24:97–102
- Kashida Y, Takahashi A, Moto M, Okamura M, Muguruma M, Jin M, Arai K, Mitsumori K (2006) Gene expression analysis in mice liver on hepatocarcinogenesis by flumequine. *Arch Toxicol* 80:533–539
- Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F, Elhajouji A, Glatt H, Hastwell P, Hayashi M, Kasper P, Kirchner S, Lynch A, Marzin D, Maurici D, Meunier JR, Müller L, Nohynek G, Parry J, Parry E, Thybaud V, Tice R, van Benthem J, Vanparys P, White P (2007) How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop. *Mutat Res* 628:31–55
- Klimisch HJ, Andreae E, Tillmann U (1997) A systematic approach for evaluating the quality of experimental and ecotoxicological data. *Regul Toxicol Pharmacol* 25:1–5
- Kroes R, Renwick AG, Cheeseman M, Kleine J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothors F, Vos JG, Würtzen G (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem Toxicol* 42:65–83
- Lein P, Locke P, Goldberg A (2007) Meeting Report: alternatives for developmental neurotoxicity testing. *Environ Health Perspect* 115:764–768
- Le Ferrec E, Chesne C, Artusson P, Brayden D, Fabre G, Gires P, Guillou F, Rousset M, Rubas W, Scarino ML (2001) In vitro models of the intestinal barrier. *Altern Lab Anim* 29:649–668
- Liebsch M, Spielmann H (2002) Current available in vitro methods used in the regulatory toxicology. *Toxicol Lett* 127:127–134
- Mangelsdorf I, Buschmann J, Orthen B (2003) Some aspects relating to the evaluation of the effects of chemicals on male fertility. *Regul Toxicol Pharmacol* 37:356–369
- Nakayama K, Kawano Y, Kawakami Y, Moriwaki N, Sekijima M, Otsuka M, Yakabe Y, Miyaura H, Saito K, Sumida K, Shirai T (2006) Differences in gene expression profiles in the liver between carcinogenic and non-carcinogenic isomers of compounds given to rats in a 28-day repeat-dose toxicity study. *Toxicol Appl Pharmacol* 217:299–307
- NAS (2007) National Academy of Science (ed). Toxicity testing in the twenty-first century: a vision and a strategy (free executive summary). Committee on Toxicity and Assessment of Environmental Agents, National Research Council ISBN: 978-0-309-10988-8, p 146. <http://www.nap.edu/catalog/11970.html>
- OECD (1990) Organisation for Economic Co-Operation and Development. Scientific criteria for validation of in vitro toxicity tests. Environment Monographs No. 36, Paris
- OECD (1995) Guidance document for the development of OECD guidelines for testing of chemicals. Environment monograph No. 76. OCDE/GD(95)71, Paris
- OECD (1996) Final report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods (Solna Report). ENV/MC/CHEM(96)9, Paris, p 60
- OECD (2001) Detailed Review Paper on non-genotoxic carcinogens detection: the performance of in-vitro cell transformation assays. Environment Health and Safety Publications Series on Testing and Assessment No. 31, Paris
- OECD (2004) The report from the expert group on (quantitative) structure-activity relationships [(Q)SARs] on the principles of validation

- of (Q)SARs. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 49. ENV/JM/MONO(2004)24, Paris
- OECD (2005a) Guidance Document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 34. ENV/JM/MONO(2005)14, Paris
- OECD (2005b) Report of the OECD/IPCS Workshop on Toxicogenomics. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 50. ENV/JM/MONO(2005)10, Paris
- OECD (2006a) Report of the validation of the updated test guideline 407: repeat dose 28-day oral toxicity study in laboratory rats. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 59. ENV/JM/MONO(2006)26. 08-Sep-2006, Paris
- OECD (2006b) Annexes to the report of the validation of the updated test guideline 407: repeat dose 28-day oral toxicity study in laboratory rats. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 59. ENV/JM/MONO(2006)26/ANN, Paris, 26 September 2006
- OECD (2006c) Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 58. ENV/JM/MONO(2006)25, Paris, 15 February 2007
- OECD (2006d) Detailed review paper on cell transformation for detection of chemical carcinogens. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 31, Paris, 28 November 2006 (draft 4th version)
- OECD (2007a) Guidance Document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 69. ENV/JM/MONO(2007)2, Paris
- OECD (2007b) Guidance on grouping of chemicals. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 80. ENV/JM/MONO(2007)28, Paris
- OECD (2007c) Draft guidance document on mammalian reproductive toxicity testing and assessment. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 43, Paris, 18 December 2007
- Ohmori K, Sasaki K, Asada S, Tanaka N, Umeda M (2004) An assay method for the prediction of tumor promoting potential of chemicals by the use of Bhas 42 cells. *Mutat Res* 557:191–202
- Ohmori K, Umeda M, Tanaka N, Takagi H, Yoshimura I, Sasaki K, Asada S, Sakai A, Araki H, Asakura M, Baba H, Fushiwaki Y, Hamada S, Kitou N, Nakamura T, Nakamura Y, Oishi H, Sasaki S, Shimada S, Tsuchiya T, Uno Y, Washizuka M, Yajima S, Yamamoto Y, Yamamura E, Yatsushiro T (2005). An inter-laboratory collaborative study by the Non-Genotoxic Carcinogen Study Group in Japan on a cell transformation assay for tumour promoters using Bhas 42 cells. *Altern Lab Anim* 33:619–639
- Okamura M, Sumida K, Muto T, Kashida Y, Machida N, Watanabe T, Mitsumori K (2004) Analysis of gene expression profiles of forestomach tumors in rasH2 mice initiated with *N*-ethyl-*N*-nitrosourea. *Arch Toxicol* 78:688–696
- Pähler A., Parker J, Dekant W (1999) Dose-dependent protein adduct formation in kidney, liver and blood of rats and in human blood after perchloroethene inhalation. *Toxicol Sci* 48:4–13
- Patlewicz G, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MI, Aptula AO, Phillips RD, Niemela J, Madsen C, Wedebye EB, Roberts DW, Bailey PT, Mekenyayn OG (2007) TIMES-SS-A promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. *Regul Toxicol Pharmacol* 48:225–239
- Pedersen F, de Bruijn J, Munn S, van Leeuwen K (2003). Assessment of additional testing needs under REACH. Effects of QSARs, risk based testing and voluntary industry initiatives. JRC Report EUR 20863 EN, 33 pp. European Commission, Joint Research Centre, Ispra, Italy. <http://ecb.jrc.it/>
- Pellizzer C, Bremer S, Hartung T (2005) Developmental toxicity testing—from animals towards embryonic stem cells. *ALTEX* 22:47–57
- Pelkonen O, Turpeinen M, Uusitalo J, Rautio A, Raunio H (2005) Prediction of drug metabolism and interactions on the basis of in vitro investigations. *Basic Clin Pharmacol Toxicol* 96:167–175
- Piersma AH. (2006) Alternative methods for developmental toxicity testing. *Basic Clin Pharmacol Toxicol* 98:427–431
- Piersma AH, Genschow E, Verhoef A, Spanjersberg MQ, Brown NA, Brady M, Burns A, Clemann N, Seiler A, Spielmann H (2004) Validation of the postimplantation rat whole-embryo culture test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32:275–307
- Poiger H, Schlatte C (1986) Pharmacokinetics of 2,3,7,8-TCDD in man. *Chemosphere* 15:1489–1494
- Prieto P, Blaauboer B.J, de Boer A.G, Boveri M, Cecchelli R, Clemedson C, Coecke S, Forsby A, Galla H-J, Garberg P, Greenwood J, Price A, Tähti H (2004) Blood-brain barrier in vitro models and their application in toxicology. *Altern Lab Anim* 32:37–50
- Prieto P, Baird AW, Blaauboer BJ, Castell Ripoll JV, Corvi R, Dekant W, Dietl P, Gennari A, Gribaldo L, Griffin JL, Hartung T, Heindel JJ, Hoet P, Jennings P, Marocchio L, Noraberg P, Pazos P, Westmoreland C, Wolf A, Wright J, Pfaller W (2006) The assessment of repeated dose toxicity in vitro: a proposed approach. *Altern Lab Anim* 34:315–341
- Roberts DW, Aptula AO, Cronin MT, Hulzebos E, Patlewicz G (2007) Global (2007) QSARs for skin sensitisation: assessment against OECD principles. *SAR QSAR Environ Res* 18:343–365
- Russell WMS, Burch RL (1959) The principles of human experimental techniques. Methuen, London
- Sakai T, Takahashi M, Mitsumori K, Yasuhara K, Kawashima K, Mayahara H, Ohno Y (2000) Collaborative work to evaluate toxicity on male reproductive organs by repeated dose studies in rats—overview of the studies. *J Toxicol Sci* 25:1–21
- Sakai A, Iwase Y, Nakamura Y, Sasaki K, Tanaka N, Umeda M (2002) Use of a cell transformation assay with established cell lines, and a metabolic cooperation assay with V79 cells for the detection of tumour promoters: a review. *Altern Lab Anim* 30:33–59
- Sasaki K, Mizusawa H, Ishidate M (1988) Isolation and characterization of ras-transfected BALB/3T3 clone showing morphological transformation by 12-O-tetradecanoyl-phorbol-13-acetate. *Jpn J Cancer Res* 79:921–930
- Schauer UM, Volkel W, Heusener A, Colnot T, Broschard TH, von Landenberg F, Dekant W (2006) Kinetics of 3-(4-methylbenzylidene)camphor in rats and humans after dermal application. *Toxicol Appl Pharmacol* 216:339–346
- Schlede E, Aberer W, Fuchs T, Gerner I, Lessmann H, Maurer T, Rossbacher R, Stropp G, Wagner E, Kayser D (2003) Chemical substances and contact allergy—244 substances ranked according to allergenic potency. *Toxicol* 193:219–259
- Seiler A, Buesen R, Hayess K, Schlechter K, Visan A, Genschow E, Slawik B, Spielmann H. (2006) Current status of the embryonic stem cell test: the use of recent advances in the field of stem cell technology and gene expression analysis. *ALTEX* 23(Suppl.):393–399
- Simon-Hettich B, Rothfuss A, Steger-Hartmann T (2006) Use of computer-assisted prediction of toxic effects of chemical substances. *Toxicol* 224:156–162
- Spangenberg C, Lausch EU, Trost TM, Prawitt D, May A, Keppler R, Fees SA, Reutzel D, Bell C, Schmitt S, Schiffer IB, Weber A,

- Brenner W, Hermes M, Sahin U, Tureci O, Koelbl H, Hengstler JG, Zabel BU (2006) ERBB2-mediated transcriptional up-regulation of the alpha-5-beta-1 integrin fibronectin receptor promotes tumor cell survival under adverse conditions. *Cancer Res* 66:3715–3725
- Spielmann H, Genschow E, Brown NA, Piersma AH, Verhoef A, Spanjersberg MQ, Huuskonen H, Paillard F, Seiler A (2004) Validation of the rat limb bud micromass test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32:245–274
- Spielmann H, Seiler A, Bremer S, Hareng L, Hartung T, Ahr H, Faustman E, Haas U, Moffat GJ, Nau H, Vanparys P, Piersma A, Sintes JR, Stuart J (2006). The practical application of three validated in vitro embryotoxicity tests. *Altern Lab Anim* 34:527–538
- Stemmer K, Ellinger-Ziegelbauer H, Lotz K, Ahr HJ, Dietrich DR (2006) Establishment of a protocol for the gene expression analysis of laser microdissected rat kidney samples with affymetrix genechips. *Toxicol Appl Pharmacol* 217:134–142
- Stokes WS, Schechtman LM, Hill RN (2002) The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). In: Balls M (ed) A review of the ICCVAM test method evaluation process and current international collaborations with the European Centre for the Validation of Alternative Methods (ECVAM), pp 23–33
- Sul D, Kim H, Oh E, Phark S, Cho E, Choi S, Kang HS, Kim EM, Hwang KW, Jung WW (2007) Gene expression profiling in lung tissues from rats exposed to formaldehyde. *Arch Toxicol* 81:589–597
- Tamura K, Ono A, Miyagishima T, Nagao T, Urushidani T (2006) Profiling of gene expression in rat liver and rat primary cultured hepatocytes treated with peroxisome proliferators. *J Toxicol Sci* 31:471–490
- Trost TM, Lausch EU, Fees SA, Schmitt S, Enklaar T, Reutzel D, Brixel LR, Schmidtke P, Maringer M, Schiffer IB, Heimerdinger CK, Hengstler JG, Fritz G, Bockamp EO, Prawitt D, Zabel BU, Spangenberg C (2005) Premature senescence is a primary fail-safe mechanism of ERBB2-driven tumorigenesis in breast carcinoma cells. *Cancer Res* 65:840–849
- Umeda M (2006) Detection of non-genotoxic carcinogens using ras-transfected bhas 42 cells. *ALTEX* 23(Suppl.):450–455
- van der Jagt K, Munn S, Tørsløv J, de Bruijn J (2004) Alternative approaches can reduce the use of test animals under REACH. Addendum to the report “Assessment of additional testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives”. JRC Report EUR 21405 EN, p 25. European Commission, Joint Research Centre, Ispra, Italy. <http://ecb.jrc.it>
- van den Waterbeemd H (2005) Which in vitro screens guide the prediction of oral absorption and volume of distribution? *Basic Clin Pharmacol Toxicol* 96:162–166
- Vermeire TG, Aldenberg T, Dang Z, Janer G, de Knecht JA, van Loveren H, Peijnenburg WJGM, Piersma AH, Traas TP, Verschoor AJ, van Zijverden M, Hakkert B (2007) Selected integrated testing strategies (ITS) for the risk assessment of chemicals. RIVM-report 601050001/2007, Bilthoven, Netherlands
- Verwei M, van Burgsteden JA, Krul CA, van de Sandt JJ, Freidig AP (2006) Prediction of in vivo embryotoxic effect levels with a combination of in vitro studies and PBPK modelling. *Toxicol Lett* 165:79–87
- Völkel W, Colnot T, Csanady GA, Filser JG, Dekant W (2002) Metabolism and kinetics of Bisphenol A in humans at low doses following oral administration. *Chem Res Toxicol* 15:1281–1287
- Walker JD, Gerner I, Hulzebos TE, Schlegel K (2004) (Q)SARs for predicting skin irritation and corrosion: mechanisms, transparency and applicability of predictions. *QSAR Comb Sci* 23:721–725
- Walker JD, Gerner I, Hulzebos TE, Schlegel K (2005) The skin irritation corrosion rule estimation tool (SICRET). *QSAR Combin Sci* 24:378–384
- Waring JF, Jolly RA, Ciurlionis R, Lum PY, Praestgaard JT, Morfitt DC, Buratto B, Roberts C, Schadt E, Ulrich RG (2001) Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol Appl Pharmacol* 175:28–42
- Whitlow S, Bürgin H, Clemann N (2007) The embryonic stem cell test for the early selection of pharmaceutical compounds. *ALTEX* 24:3–8
- Williams FM (2004) EDETOX. Evaluations and predictions of dermal absorption of toxic chemicals. *Int Arch Occup Environ Health* 77:150–151
- Williams F (2006) In vitro studies—how good are they at replacing in vivo studies for measurement of skin absorption? *Environ Toxicol Pharmacol* 21:199–203
- Worth AP, Balls M (2001) The importance of the prediction model in the validation of alternative tests. *Altern Lab Anim* 23:135–143
- Worth AP, Balls M (eds.) (2002a) Alternative (non-animal) methods for chemicals testing: current status and future prospects. A report prepared by ECVAM and the ECVAM working group on chemicals. *Altern Lab Anim* 30(Suppl. 1):1–125
- Worth AP, Balls M (2002b) The principles of validation and the ECVAM validation process. In: Balls M (ed) pp 15–21
- Zuang V, Hartung T (2005) Making validated alternatives available—the strategies and work of the European Centre for the Validation of Alternative Methods (ECVAM). *Altern Anim Testing Exp* 11:15–26
- Zur Nieden NI, Kempka G, Ahr HJ (2004) Molecular multiple endpoint embryonic stem cell test—a possible approach to test for the teratogenic potential of compounds. *Toxicol Appl Pharmacol* 194:257–269