European Food Safety Authority

The EFSA Journal (2009) 1052, 1-77

Scientific Opinion of the Scientific Committee

"Existing approaches incorporating replacement, reduction and refinement of animal testing: applicability in food and feed risk assessment"[1]

(Question No EFSA-Q-2005-231)

Adopted on 8 April 2009

This opinion, published on 15 July, replaces the earlier version published on 6th June 2009

Scientific Committee Members

Sue Barlow, Andrew Chesson, John D. Collins, Albert Flynn, Anthony Hardy, Klaus-Dieter Jany, Ada Knaap, Harry Kuiper, John Christian Larsen, David Lovell[2], Pierre Le Neindre, Jan Schans, Josef Schlatter, Vittorio Silano, Staffan Skerfving and Philippe Vannier.

SUMMARY

The founding Regulation of EFSA[3] requires the Authority to contribute to a high level of protection of human life and health, and in this respect to take account of animal health and welfare. EFSA is committed to a proactive animal welfare approach, based on sound scientific principles and the need to ensure that adequate data are available for a reliable risk assessment. In this context, EFSA and its Scientific Committee recognise the importance to stimulate the use of food and feed assessment approaches that would not only minimise the number of experimental animals and any suffering, but also work towards their replacement.

The present Opinion gives an overview of the legislation and guiding principles on the use of animals for experimental purposes and explains the principles of scientific validation of toxicity tests for regulatory purposes. It summarises possibilities for replacement, reduction and refinement (Three Rs) of toxicological animal testing within the different areas of EFSA's activities. Finally, integrated testing and risk assessment strategies are considered and recommendations are given to implement better animal welfare within EFSA's activities.

Most of the assessments conducted by EFSA's Scientific Committee and Scientific Panels require experimental data. The complete replacement of animal experiments while maintaining the same level of food and feed safety is not possible at present. However in the meantime, considerable improvements can be made to promote better animal welfare. In this

[1]For citation purposes: Opinion of the Scientific Committee on a request from EFSA on existing approaches incorporating replacement, reduction and refinement of animal testing: applicability in food and feed risk assessment. *The EFSA Journal (2009)* 1052, 1-77.

[2] External expert of the Scientific Committee

[3] http://eur-lex.europa.eu/LexUriServ/site/en/oj/2002/l_031/l_03120020201en00010024.pdf

Opinion, the existing approaches for incorporating replacement, reduction and refinement methods for each toxicological endpoint are reviewed, and recent and future developments are discussed.

The status of the science can be summarised as follows:

Toxicokinetic studies: several *in vitro* methods have been developed, however they cannot at present replace *in vivo* methods.

Acute toxicity testing: the classical oral test has been replaced by *in vivo* alternatives that require fewer animals and reduce pain and suffering. *In vitro* methods are available for estimating starting doses.

Skin irritation and corrosion testing: for skin irritation, at present only an *in vivo* method is accepted by regulatory authorities[4]; however it is anticipated that an *in vitro* method will be accepted soon for regulatory purposes at OECD level. For corrosion, *in vitro* methods are available and must be used in Europe.

Skin sensitisation testing: only *in vivo* methods are accepted by regulatory authorities among which the Local Lymph Node Assay, LLNA, is the preferred refined method. An alternative *in vivo* method for the LLNA, and *in silico* and *in vitro* methods, are being developed.

Eye irritation testing: the *in vivo* Draize rabbit eye test is still the standard for eye irritation testing in Europe. However, four *in vitro* methods can be used to classify severe eye irritants without further testing on animals.

Testing for **acute systemic and local toxicity** is still required for plant protection products and animal feed additives to assess risk to workers. EFSA does not require such tests for risk assessment of food additives, food contact material or newly expressed proteins in GMO.

Genotoxicity testing: initial testing is often done using *in vitro* methods, but positive results may need to be confirmed by *in vivo* testing.

In other areas of toxicology, particularly those involving complex endpoints, such as those that are investigated in **repeated dose toxicity, reproduction and developmental toxicity studies,** the development of alternative methods is more difficult. The same applies for **ecotoxicity** endpoints such as acute and chronic toxicity in fish and birds and bioconcentration in fish. A number of methods, based on the Three Rs, are under development and some are undergoing validation. At present, they cannot yet provide the information that can be derived from currently used *in vivo* methods. In these areas, modifications of the individual tests and the implementation of integrated testing and risk assessment strategies can contribute to a reduction in the number of animal studies needed and, by the choice of test selected, may result in use of fewer animals. Various EFSA approaches implementing the Three Rs are described. They relate for example to tiered testing, which has been developed for food contact materials, and to application of the threshold of toxicological concern for flavouring substances and metabolites of plant protection products in groundwater. Another approach is the

[4] http://eur-lex.europa.eu/JOHtml.do?uri=OJ%3AL%3A2008%3A142%3ASOM%3AEN%3AHTML

implementation of the qualified presumption of safety which will lead to a significant reduction in the need for animal testing in the assessment of microorganisms.

The EFSA Scientific Committee notes that the legislation in force since 1986 requires that when validated, practical and accepted alternative methods, as described by the Three Rs are available they must be used. This should be communicated to the applicants that submit dossiers to EFSA as well as being fully reflected in any guidelines developed by EFSA. The Scientific Committee also recommends that in EFSA evaluations, all existing data should be reviewed before any additional *in vivo* studies are requested.

Recognising that the Commission and Member States have responsibility for agreement of new testing methods, it is important that the existing communication between EFSA and the corresponding Commission Services that lead in this area continue to be improved, so that EFSA is informed of latest developments on the validation and acceptance of new testing methods. Communication on implementing the Three Rs with other EU Agencies dealing with chemical risk assessment is also important.

The Scientific Committee recommends that EFSA follows up this Opinion with a review of progress in the field of alternatives in 3 years time.

Key words: **Food and feed safety, Three Rs, experimental animals, alternative methods, *in vivo* methods, animal welfare, alternative risk assessment testing strategies, replacement, reduction, refinement.**

TABLE OF CONTENTS

BACKGROUND

The concern regarding use of animals for experimental purposes in the European Community is laid down in the Council Directive 86/609/EEC of 24 November 1986 on the approximations of laws, regulations and administrative provisions of the Members States regarding the protection of animals used for experimental and other scientific purposes. This Directive is currently being revised. The principles regarding animal welfare found in this Directive are also found in the European Convention for the protection of vertebrate animals used for experimental and other scientific purposes (Strasbourg 1986).

In the founding Regulation of EFSA (Regulation EC No. 178/2002) it is stated that the mission of EFSA includes that "The Authority shall contribute to a high level of protection of human life and health, and in this respect take account of animal health and welfare…" (Article 22).

The EFSA Management Board on its meeting of 22 June 2004 supported EFSA's willingness to develop a proactive animal welfare approach provided that it was based on sound scientific principles. The Management Board also emphasised that the welfare of food producing animals is already adequately covered by the mission of the Scientific Panel on Animal Health and Animal Welfare (AHAW). The implementation of such a policy should stimulate the development of new food and feed assessment approaches that would not only minimise the numbers of experimental animals and their suffering, but also work towards their replacement through the use of alternative techniques (Replacement, Reduction and Refinement i.e. the Three Rs approach).

TERMS OF REFERENCE

It is recognised that the implementation of fundamental changes to improve the welfare of experimental animals in relation to EFSA's activities would take several years and can be achieved only step-by-step. Therefore a Working Group of the Scientific Committee on Animal Welfare should be established comprising representatives of the Scientific Committee, representatives of a number of Scientific Panels dealing with animal studies (e.g. the AFC Panel - now replaced by the ANS and CEF Panels -, the AHAW Panel, the CONTAM Panel, the GMO Panel, the FEEDAP Panel and the PPR Panel), EFSA staff with a background in hazard assessment based on animal studies, and external experts on an as-needed basis. The mission of the Scientific Committee and its new Working Group would be to develop a stepwise approach to define and incorporate elements in the process and procedures of risk assessment which are aimed at the replacement, reduction, and refinement of experimental animals without compromising the integrity of EFSA's risk assessment procedures and with the constant aim of achieving the highest scientific quality of the work. The activities of the Working Group would be limited to experimental vertebrates but the Working Group may alert, as appropriate, the AHAW Panel of issues that would require the Panel's attention.

In order to accomplish its mission, the Working Group will commence with a series of short-term projects, being the basis for one or more long-term projects that will be defined at a later stage. Therefore, the current mandate comprises only those tasks which can be completed within a reasonable time frame. These are:

1. To develop a comprehensive overview of all current EU legislative and guidance documents that address experimental animals and their welfare. It should also address methodologies officially accepted or in use in the EU or some countries although not formally validated;
2. To provide all Panels, Task Forces, Expert Groups and their Working Groups with the comprehensive overview once approved by the Scientific Committee as adequate and sufficiently covering the area, and ensure that all have access to the documents mentioned therein;
3. To determine which guidance documents and procedures are currently applied by the EFSA Panels that could have an impact on experimental animals and their welfare;
4. To propose a process to harmonise the application of guidance and legislative elements across EFSA Panels;
5. To make an inventory of all current activities of the Panels and Scientific Committee that relate to animal welfare, e.g. implementation of the Qualified Presumption of Safety approach, voluntary data sharing;
6. To draft a proposal for an action plan to improve sharing of information with organisations active in the area of animal welfare, e.g. bodies involved in the development of validation of methodologies for safety assessment, regulatory bodies requiring animal testing.

Longer-term projects could include proposals for the development and application of new approaches or concepts for chemical and microbiological risk assessment that would take better account of the Three Rs or, alternatively the evaluation of such new approaches developed by other expert groups.

As a continuous activity, it would be appreciated if the Scientific Committee, could:

1. Inform Panels of the latest scientific developments related to alternative methods to animal testing, and to internationally available, alternative approaches for hazard characterisation;
2. Contribute to the improvement of existing guidance documents and procedures where appropriate, in collaboration with the risk managers, in order to take account of developments in the use of alternative methods with regard to current requirements for testing and food/feed assessments;
3. Advise, as appropriate, on how to stimulate new research activities and new approaches in the field of risk assessment which would work towards the Three Rs policy.

In its work the Scientific Committee is requested to take into consideration:

- The numerous ongoing activities and approaches in this area at national and international levels (e.g. EMEA, ECVAM-IHCP-JRC, OECD, OIE, DG RTD, DG SANCO Scientific Committee on Consumer Products);
- The specific needs and the diversity of requirements of the EFSA Panels.

INTERPRETATION OF THE TERMS OF REFERENCE

The Scientific Committee has addressed the mandate as follows:

1- In the working document "Overview of the test requirements in the area of food and feed safety" (presented at the SC plenary in November 2007) the following Terms of Reference have been addressed:

- To develop a comprehensive overview of all current EU legislative and guidance documents that address experimental animals and their welfare. It should also address methodologies officially accepted or in use in the EU or some countries although not formally validated;
- To provide all Panels, Task Forces, expert groups and their Working Groups with the comprehensive overview once approved by the Scientific Committee as adequate and sufficiently covering the area, and ensure that all have access to the documents mentioned therein;
- To determine which guidance documents and procedures are currently applied by the EFSA Panels that could have an impact on experimental animals and their welfare;

The document was endorsed by the Scientific Committee as a **working document** on the basis of which the following draft Opinion has been built (see point 2).

2- The Opinion "Existing approaches incorporating replacement, reduction and refinement of animals testing: applicability in food and feed risk assessment" covers the following terms of reference:

- To make an inventory of all current activities of the Panels and Scientific Committee that relate to animal welfare, e.g. implementation of the Qualified Presumption of Safety approach, voluntary data sharing;
- To inform Panels of the latest scientific developments related to alternative methods to animal testing, and to internationally available, alternative approaches for hazard characterisation;
- To advise, as appropriate, on how to stimulate new research activities and new approaches in the field of risk assessment which would work towards the Three Rs policy.

3- The following Terms of Reference have not been pursued in depth and the reasons for this are explained below or in the general conclusions of the Opinion:

- To propose a process to harmonise the application of guidance and legislative elements across EFSA Panels.
- To draft a proposal for an action plan to improve sharing of information with organisations active in the area of animal welfare, e.g. bodies involved in the development of validation of methodologies for safety assessment, regulatory bodies requiring animal testing.
- To contribute to the improvement of existing guidance documents and procedures where appropriate, in collaboration with risk managers, in order to take account of

developments in the implementation of alternative methods with regard to current requirements for testing and food/feed assessments. ***This point of the Terms of Reference is done by the different EFSA Panels upon request from the Commission.***

ACKNOWLEDGEMENTS

The European Food Safety Authority wishes to thank the members of the Working Group for the preparation of this Opinion: Sue Barlow, Diane Benford, Erik Dybing (member of the working group until July 2008), Alberto Mantovani, Otto Meyer, David Morton, Iona Pratt, Horst Spielmann (Vice-Chair) and Pierre Le Neindre (Chair).
The contributions received by Marlies Halder (DG JRC), Marcelle Holloway (DG Environment) and Antonio Lacerda (DG Enterprise) were also highly appreciated.

1. Introduction

The mission of the European Food Safety Authority (EFSA) as laid down in its founding Regulation5 (Regulation No. 178/2002) includes the provision of "scientific and technical support for the Community's legislation and policies in all fields which have a direct or indirect impact on food and feed safety" (Article 22). This Article also states that "The Authority shall contribute to a high level of protection of human life and health, and in this respect take account of animal health and welfare…". Animal welfare considerations in the field of food and feed safety assessment notably include concerns about experimental animals.

EFSA receives documentation from animal studies as part of the dossiers provided by an applicant who applies for approval of the use of a substance. The content of dossiers usually relies on the legal basis under which the application was made, including EC Regulations and sometimes Guidelines. These Regulations or Guidelines generally require that studies on animals be performed.

In its 15th Management Board meeting6 (22 June 2004) EFSA decided on the adoption of a pro-active animal welfare approach summarised as follows: "While recognising that animal testing cannot be eliminated at present, EFSA could make every effort to stimulate, and participate in, the development of new food and feed assessment approaches that would minimise the use of experimental animals and would reduce to the extent possible the level of suffering of those animals that are still needed today".

The implementation of the above decision should stimulate the development of new food and feed assessment approaches that would not only minimise the numbers of experimental animals used and their suffering, but also work towards their replacement. At the same time, a high level of human health protection should be maintained.

2. Legislation and guiding principles on the use of animals for experimental purposes

The EU legally recognises the welfare requirements of animals. The protocol on protection and welfare of animals, annexed to the Treaty of Amsterdam, provides that "In formulating and implementing the Community's agriculture, transport, internal market and research policies, the Community and the Member States shall pay full regard to the welfare requirements of animals, while respecting the legislative or administrative provisions and customs of the Member States relating in particular to religious rites, cultural traditions and regional heritage".

In the area of animal experimentation the basis of European legislation on the welfare of animals used for scientific purposes is Council Directive 86/609/EEC on the protection of animals used in scientific experiments. Due to a variety of weaknesses in this Directive, some Member States have passed laws which have lead to a widening gap in standards in Europe. In order to rectify this imbalance and make further improvements to the welfare of animals used in scientific procedures, the Commission has published a proposal to revise Directive 86/609/EEC (COM/ 543/final/5 November 20087).

[5] http://eur-lex.europa.eu/LexUriServ/site/en/oj/2002/l_031/l_03120020201en00010024.pdf
[6]http://www.efsa.europa.eu/EFSA/AboutEfsa/WhoWeAre/ManagementBoard/efsa_locale-1178620753812_MeetingsMB.htm
[7] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52008PC0543:EN:NOT

Underpinning the Commission proposal are measures for compulsory ethical evaluation and authorisation of procedures and projects using animals, which will allow for maximum implementation of the Three Rs principle of replacement, reduction and refinement. The proposal explicitly mentions that the Three Rs must be taken into account in all aspects of animal use, including breeding, housing and care. It specifies that a validated alternative test method where recognised by Community legislation must be used (methods are listed in EC Regulation 440/2008/EC); Member States shall ensure that the number of animals used shall be reduced to the absolute minimum possible; and that there is a refinement of breeding, accommodation and care and of test methods. The compliance to these specific requirements is examined during the compulsory ethical evaluation of all projects using animals.
The Commission proposal also introduces a ban on the use of great apes in scientific procedures, however it also states that when survival of the species itself is at stake, or in the case of an unexpected outbreak of a life-threatening or debilitating disease, a Member State can exceptionally be granted permission for their use. The scope of the current Directive is widened to include specific invertebrate species and foetuses in their last trimester of development, as well as animals used for the purposes of basic research, education and training.
The Commission proposal was forwarded to the European Parliament and to the Council in November 2008 and will follow the 'co-decision' procedure where both Parliament and Council must agree on the final content of the text before it becomes legislation. The whole process may take more than 18 months. An additional 18 months transition period may be given to Member States to update their national legislation and comply with the new provisions. More information can be found at the website8.

2.1. The Three Rs: replacement, reduction and refinement of animal testing

All vertebrates are protected animals and should be treated the same as they are all assumed to be sentient and able to suffer and experience pain and distress. The Three Rs (Russell & Burch, 1959) make up an ethical framework by which the use of animals in research projects and for safety testing for regulatory purposes can be reviewed to help ensure humane experimentation. They pose a set of questions which have been subsequently refined by those with oversight of animal research, e.g. research directors, sponsors of research, licensing authorities, as well as those carrying out the research.

Assuming the work needs to be carried out, then the first ‘R’ question is whether the scientific objectives could be achieved without using animals at all, e.g. **Replacing** animals through the use of cell cultures *(in vitro)*, computer modelling *(in silico*) etc. If replacement alternatives are not available, and living sentient animals have to be used, then the next two Rs are invoked. It may be possible to maintain the scientific objective of the work by **Reducing** the number of animals to be used, or by **Refining** the work or both, so that less animal suffering is caused. The Three Rs are thus defined as Replacement, Reduction and Refinement. For all studies carried out on animals, the training and competence of scientific research and care staff is vitally important to be able to comply with the Three Rs, human and animal safety, and to carry out good science.

[8] http://ec.europa.eu/environment/chemicals/lab_animals/home_en.htm

2.1.1. Replacement

Replacement is mandated in Directive 86/609/EEC, Article 7.2 thus: "An experiment shall not be performed if another satisfactory method of obtaining the results sought, not entailing the use of an animal, is reasonably and practicably available". It may be further sub-divided into **absolute replacement** where animals are not required at all, e.g. using permanent cell lines, computer modelling; and **relative replacement** where animals are still required, e.g. using primary cells. Sometimes, both *in vitro* and *in vivo* experiments are required as different scientific questions are being addressed, and the results obtained *in vitro* may avoid the need for *in vivo* work.

Replacement methodologies include some of the following:

1. Improved storage, exchange and use of available information on animal experiments.
2. Mutual acceptance of data conducted according to recognised national and international guidelines.
3. Mutual recognition, whereby registration of a product by one regulatory authority is accepted by another regulatory authority, without the need for further data.
4. Predictions based on the physical and chemical properties of a substance.
5. Computer models, e.g. QSARs, molecular modelling of biochemical and other processes, the description of biologically active compounds.
6. Organisms of lower neurological sensitivity, e.g. some invertebrates that are not sentient and bacteria, plants.
7. *In vitro* methods, such as cell culture, tissue slices, isolated organs, whole embryos.

2.1.2. Reduction

Reduction is mandated in Directive 86/609/EEC, Article 7.3 as: "… In a choice between experiments, those which use the minimum number of animals,… and which are most likely to provide satisfactory results shall be selected".

Reduction can be described in terms of the number of animals used to obtain information of a given amount and precision. In practice, reduction has two aspects: first, a reduction of the number of animals used in the required tests; and second, a reduction of the number of tests required. Reduction is also applied through careful choice of species and strain of animal that is appropriate to the scientific objective, the experimental design and the use of more modern statistical methodology. Other ways to reduce the numbers of animals include the use of pilot studies, historical data, careful consideration of controls, staging experiments so that key experiments are carried out first possibly making further ones redundant, and critical interim review of an ongoing study may reveal that the scientific objective has been achieved at an earlier stage than predicted. Statistical approaches are being developed that could help minimise animal numbers with no loss in sensitivity (Festing et al., 2002; Stigler, 1982). It is also noteworthy that group sizes for fish and bird studies are often larger than for mammalian studies but no biological or scientific reason is given for this difference.

2.1.3. Refinement

Refinement is mandated in Directive 86/609/EEC, Article 7.3 as: "… In a choice between experiments, those which…, involve animals with the lowest degree of neurophysiological

sensitivity, cause the least pain, suffering, distress or lasting harm … and which are most likely to provide satisfactory results shall be selected".

Refinement can be described as those methods which aim to promote animal health and welfare, alleviate or minimise the potential pain, distress and other adverse effects, or that enhance their wellbeing during their routine husbandry and care.

Animals that are unhealthy through contagious disease and infections are likely to experience poor welfare and also make unreliable research subjects. However, over the past 20 years or so, the health status of laboratory animals has improved considerably (FELASA 1994, 2002). This improvement has also contributed to the scientific integrity of the work as disease in the past sometimes confounded experimental outcome measures. Furthermore, the accompanying reduction in outcome variance can contribute to a reduction in animal numbers.

Avoiding the unnecessary harm associated with performing an *in vivo* study not only reduces animal suffering but again reduces the variance due to unnecessary perturbation of animal homeostasis through an animal's response to pain, distress, malaise, suffering, etc. Early detection of unnecessary harm is crucial to avoid finding animals dead as well as providing important data relating to the test being undertaken. The implementation of humane endpoints in this regard is important (OECD Guidance Document 19, 2000). Finally, sharing data not only helps to reduce duplication of experiments that results in unnecessary use of animals, but can also contribute to improvement of animal welfare in other ways. For example, sharing information on failed experiments, on the recognition, alleviation and avoidance of unnecessary harm and on experimental models can contribute to the refinement of methods.

Refinement requires that any animal suffering is kept to the minimum to meet the scientific objective and this is both a legal and an ethical obligation. However, before that can be done the existence of any pain and distress has to be recognised and then assessed in some way. Some tests require the death of some animals, for example acute fish toxicity or potency testing of some biological preparations stipulated by pharmacopoeia monographs. Requirements for late endpoints that inflict significant suffering on animals have always to be rigorously scientifically justified and it has also to be shown that there are no alternative ways of achieving the same scientific objective not involving that degree of pain and distress. In addition, the anticipated benefits of the proposed test have to be realistically sought and proportionate to the degree of animal suffering that will be incurred.

In the end, the improvement in husbandry will enhance the wellbeing of the animals by meeting their needs in particular those relating to social interactions and space and resources to carry out natural behaviour. This factor is recognised through the revised guidance in Commission Recommendation 2007/526/EC for the accommodation and care of animals used for experimental and other scientific purposes and in Appendix A of the European Convention on the Protection of Vertebrate Animals (ETS 123). Many of these standards have been included in the proposal to revise Directive 86/609/EEC and will become binding once the proposal is adopted.

2.2. Principles of scientific validation of toxicity tests for regulatory purposes

Regulators will only accept alternatives to animal tests in toxicology if the new tests will allow them to assess hazards with at least the same level of reliability as the current animal tests. The Organization for Economic Cooperation and Development (OECD) plays a pivotal role in this

process by ensuring that toxicity tests can be adopted as OECD Test Guidelines (TG) only after successful validation. OECD TGs provide standardised protocols, which ensure that tests conducted in one country will be accepted for assessment in another country, under the OECD decision of 1981 on Mutual Acceptance of Data[9].

About two decades ago, various stakeholders (academia, regulators, industry and animal welfare organisations) began working together to define the validation process which would establish the reproducibility and relevance of a toxicity testing procedure (animal or non-animal based) for a particular purpose (Balls et al., 1990). The failure of some large-scale validation studies of tests for eye irritation at the beginning of the 1990s demonstrated that more detailed validation guidelines were needed. The at that time newly established European Centre for the Validation of Alternative Methods (ECVAM-IHCP-JRC) at the European Commission's Joint Research Centre with stakeholder involvement developed and published reports on validation principles (Balls et al., 1995; Curren et al., 1995; Worth A, 2002). They were endorsed in 1996 by US regulatory Agencies (NIEHS, 1997) and also by the OECD (1996).

The main points to be considered in the validation of new methods are:

a) The new method should be sufficiently developed, e.g. transferable to and reproducible in other laboratories.

b) For non-animal based methods, a biostatistically based prediction model should be defined which allows the extrapolation from the determined endpoints to *in vivo* animal or humane endpoints.

c) Experimental validation should be carried out under blind conditions in several laboratories.

d) The management of and the responsibilities in the validation process should be well-defined.

e) The biostatistical analysis, and the selection, coding and shipment of the test samples should be independent.

f) The results of the validation trial should be independently assessed.

Following the improved principles, several *in vitro* methods were successfully validated (Fentem et al., 1998; Liebsch et al., 2000; Spielmann et al., 1998a; Spielmann et al., 1998b) and accepted for regulatory purposes by the European Commission in 2000 (European Commission 2000a, 2000b) and by the OECD in 2004 (OECD TG 430, 431, 432 adopted in 2004).

To improve the acceptance of alternative methods at the international level, the OECD has recently published a guidance document on the validation and international acceptance of new or updated test methods for hazard assessment (OECD Guidance Document 34, 2005).

When conducting animal experiments for food and feed safety, there should be compliance with EU Dir 86/609/EEC with respect to requirements to replace, reduce or refine current testing in animals. In general, testing should be conducted according to established guidelines that are endorsed by services of the EU Commission, or EU Agencies (e.g. EFSA), or by other international Agencies [e.g. OECD or the International Organization for Standardization (ISO)]. Such documents help ensure that testing is conducted according to the highest standard

[9] http://webdomino1.oecd.org/horizontal/oecdacts.nsf/linkto/C(81)30

and that the results will be accepted at the international level and the tests need not be repeated. The concept of validation has been developed to ensure that new animal tests and non-animal tests can be accepted for regulatory purposes by regulatory Agencies around the world. However, Agencies in some countries have accepted validated *in vitro* toxicity tests only for positive classification and labelling. As a consequence, negative results *in vitro* must still be confirmed by additional testing in animals. For "stand alone *in vitro* tests" (full replacement) as for example the *in vitro* skin irritation test using human epidermis models (Spielmann et al., 2007) neither positive nor negative results need to be confirmed *in vivo* (see also chapter 4.1.2.1 for more details).

2.3. The European Partnership on Alternatives to Animal Testing (EPAA)

The European Partnership for Alternative Approaches to Animal Testing (EPAA[10]) is a joint initiative from the European Commission and a number of companies and trade federations active in various industrial sectors.

The Partnership was launched on 7 November 2005 by Commissioners and industry representatives from seven large industrial sectors. Its purpose is to promote the development of new Three Rs methods (replace, refine and reduce) as modern alternative approaches to regulatory testing. The Partnership's work focuses on mapping existing research, promoting the search for new alternative approaches and strategies, designing and implementing intelligent compliance testing strategies, and promoting communication, education, validation and acceptance of alternative approaches.

3. Possibilities for replacement, reduction and refinement (Three Rs) of animal testing within the different areas of EFSA's activities

As already stated in the Introduction, the implementation of a proactive animal welfare approach within the different areas of EFSA's activities is in part dependent on the development of new food and feed assessment approaches. These approaches should reduce the numbers and suffering of the experimental animals required, and should consider all possibilities for their replacement.

The requirements for animal tests as part of the data package or dossier needed by EFSA to carry out a risk assessment of a particular agent present in food or feed (e.g. zoonotic agents, food or feed additives, pesticides, contaminants) depend on the legal framework applicable to the agent, the potential exposure of consumers and the state of knowledge regarding the potential hazards. In turn, these factors have a significant impact on the possibilities for replacement, reduction and refinement of animal testing in the data required for risk assessment.

In many areas of EFSA's activities, for example the assessment of plant protection products, the legal framework triggers testing requirements for a wide range of toxicological and ecotoxicological endpoints. Possibilities for replacing, reducing or refining the use of animals used and the application of alternative methods thus vary from area to area in EFSA's risk assessment work. For example, in the assessment of food additives, there are currently relatively few possibilities for the application of alternative methods, since many of the

[10] http://ec.europa.eu/enterprise/epaa/brochure.htm

endpoints for which significant advances have been made, such as irritation/corrosivity and acute toxicity as outlined in Section 4, are not required endpoints for food additives. On the other hand, the approach used in the assessment of a new food additive is relatively flexible compared with that required, for example, for a plant protection product, and offers the possibility of waiving certain animal studies. The studies required depend on the chemical nature of the additive, its (proposed) uses and levels of use in food and consequently the likely exposure of consumers and whether it is a new additive or a re-examination of an existing additive, as outlined in the guidance on submissions for food additive evaluations of the Scientific Committee on Food (SCF, 2001). In contrast, the assessment of plant protection products may offer fewer opportunities to waive animal testing for certain endpoints such as reproductive toxicity, carcinogenicity and chronic toxicity. However, the Three Rs may be implemented for mandatory endpoints such as skin and eye irritation and corrosion.

Another aspect of EFSA work concerns animal testing related to vaccines and to diagnostic tests for animal diseases. In general, vaccines for animal diseases are tested both for safety and efficacy. EMEA and the national control authorities are responsible for the licensing and batch control of vaccines; i.e. each batch produced is tested for safety and efficacy (potency). In some circumstances, EFSA can be asked to give an opinion on the application of a vaccine as a way to control the outbreak of a disease. In this case, additional testing may be required to assess the efficacy of the vaccines. The testing of vaccines raises issues over humane endpoints, depending on the vaccine, as well as the use of surrogate species for target animals, and suitable controls for the assessment of test organism viability and dose. The use of animals to maintain invertebrate vectors is also an issue. Regarding diagnostic tests for animal diseases (e.g. Avian Influenza, Bluetongue, Newcastle Disease), the responsibility of their evaluation changes according to different diseases (Community Reference Laboratories, National Reference Laboratories or others). In special cases, EFSA can be asked to evaluate the relevance of the diagnostic tests. In this exercise, it may recommend research involving the use of animal testing for the evaluation of the validity of a diagnostic test.

In risk assessment it is important to consider any available human data derived from epidemiological studies, volunteer studies, or case reports, e.g. reports of poisoning from exposed individuals. Reliable human data are particularly useful because they do not carry the attendant uncertainties of cross-species extrapolation. Epidemiology studies are especially useful for identifying a relationship between human exposure and effects such as biological effect markers, early signs of chronic effects, disease occurrence, or mortality and may provide the best data for risk assessment. Several EFSA Scientific Panels make extensive use of human epidemiological data, case- reports and studies in volunteers, when available. For example, the Scientific Panel for Food Additives and Nutrient Sources (ANS), in its work on food supplements uses the results of human (clinical) studies on a range of nutrients already on the market and also available from the human diet to judge the safety and bioavailability of these sources. The Scientific Panel on Contaminants in the Food Chain (CONTAM) regularly uses human data, including from case reports of human poisoning (e.g. to set Acute Reference Doses for marine biotoxins), epidemiological studies of effects in exposed populations (e.g. aflatoxins) and biomonitoring (e.g. to consider margins of exposure between levels of a contaminant in blood of humans in the general population and of animals in pivotal toxicity studies).

Data sharing can also contribute to the implementation of the Three Rs. For example, EFSA works with other EU Agencies and Institutions active in closely related fields by exchanging information and cooperating on matters of mutual interest. To reinforce these relations EFSA seeks to sign Memoranda of Understanding with other EU Agencies on enhancing cooperation and information exchange. Moreover, at the EFSA Advisory Forum meeting in September 2006, representatives from European National food safety authorities signed a 'Declaration of Intent' aimed at strengthening scientific cooperation and information exchange on risk assessment and risk communications in the European Union.[11]

EFSA has signed a Memorandum of Understanding[12] with the European Centre for Disease Prevention and Control (ECDC) to increase cooperation and exchange scientific information on topics of mutual interest including food safety, control of communicable diseases, infectious diseases prevention and emergency response.

EFSA has also signed a collaboration agreement[13] with the European Commission's Joint Research Centre (DG JRC) to strengthen the cooperation in the field of food and feed safety, animal health and welfare, plant health and nutrition.

Conclusions

1. The application of the Three Rs to animal testing requirements in the field of food and feed risk assessment could result in a reduction of the numbers of animals and their suffering, and at times it may be possible to replace the use of animals altogether.
2. Sometimes, legal requirements inhibit the implementation of the Three Rs alternatives.
3. In some areas, validated alternative methods are already available and accepted internationally.

Recommendations

1. Possibilities for the application of the Three Rs in animal testing requirements should always be considered.
2. Dialogue between regulatory authorities and EFSA risk assessors should be encouraged.
3. Petitioners submitting dossiers for risk assessment should consider the use of alternative testing strategies that are already available. In a choice of *in vivo* tests, the one causing least suffering, followed by the one that uses fewest animals, should be chosen whenever this is scientifically appropriate.

4. Replacement, reduction and refinement of animal testing for regulatory purposes: current state of the science

The following chapters describe, for each endpoint, the current test methods in use followed by an overview of the recent and future developments in terms of the Three Rs.

[11] http://www.efsa.europa.eu/EFSA/PartnersNetworks/AdvisoryForum/efsa_locale-1178620753812_Declarationofintent.htm
[12] http://www.efsa.europa.eu/cs/BlobServer/General/efsa_ecdc_mou.pdf?ssbinary=true
[13] http://www.efsa.europa.eu/cs/BlobServer/General/corporate_collaboration_agreement_EFSA_JRC,0.pdf?ssbinary=true

4.1. Mammalian toxicokinetics, acute, local and systemic toxicity

4.1.1. Toxicokinetics

Toxicokinetic studies provide data on absorption, distribution, metabolism and elimination (ADME) of the test substance in the organism and aid in understanding its mode of action. The data can be used to assess the adequacy and relevance of the extrapolation of animal toxicity data to human hazard and risk assessment. They are useful for establishing a testing program i.e. the extent of toxicological testing (e.g. reduced testing may be possible for substances not absorbed or metabolised to normal dietary or body constituents), and the choice of an appropriate test species, including dose levels and duration of toxicity studies.
Toxicokinetics studies are mandatory for plant protection products and feed additives, but are also normally performed as core set of studies for evaluation of the safety of food additives.
Normally the rat is used, based upon the large database for this laboratory animal from toxicokinetic and toxicological studies. The design of metabolism and toxicokinetic studies must be flexibly adapted to the particular substance being tested. Not all aspects may need to be investigated in every case. In principle, whole animal studies using single and repeat dosing are needed. These will enable determination of gastrointestinal absorption and overall elimination rates, any changes in the kinetic behaviour of the substance with repeated administration, and, if comparative studies are available, evaluation of species differences (EU B. 36/OECD TG 417).

In vitro studies, employing enzymes, subcellular organelles, cell cultures and perfused organs, can also contribute useful information in the investigation of metabolic pathways, mechanisms of toxicity, effects on enzymes and other specific aspects. Toxicokinetic information in humans is also of considerable value for a number of reasons, including confirmation of the validity of the animal models used.

Animal welfare concerns: they include the numbers of animals used and variation according to species and refinement in the administration and sampling methods. Repeated administration can be stressful for animals. Some routes of blood sampling are less invasive and have fewer side effects than others, and the use of biomarkers, such as excreta, will have less impact on the animals. Whilst death may not be a required endpoint in many studies, some suffering in animals is anticipated in repeat dose studies with the toxic doses but 'severe' suffering should be avoided. It may be possible to reduce any suffering by emphasising that animals may give data when showing just mild adverse effects as opposed to producing 'severe' effects in the animals. More sensitive measurement techniques and the use of fewer animals can be considered at multiple time points, as well as the possibility with cannulation to use the same animal over multiple time points.

Recent and future developments:
In recent years, methods to integrate *in vitro* results into a prediction of ADME *in vivo* by the use of appropriate physiologically-based toxicokinetic (PBTK) models have been developed (Blaaboer et al., 2001, 2002). *In silico* models for toxicokinetics are mathematical models which can be used to understand how a chemical is handled by the body. Such models describe the body as a set of compartments through which chemicals travel or are transformed. These models have until now only been used to a limited extent as they are data- and resource-demanding. However, the potential of PBTK models to generate predictions from *in vitro* or *in vivo* information is one of their attractive features in the risk assessment of substances.

Several alternative *in vitro* methods have been developed and proposed to predict absorption/bioavailability (Pelkonen et al., 2001), distribution (Hinderling, 1997) and metabolism (Pelkonen et al., 2005), but these *in vitro* methods cannot at present be used as stand-alone methods to predict the toxicokinetic behavior of a substance.

4.1.2. Acute toxicity

Acute toxicity data are generated for hazard identification to be used in risk assessment for human health and, under some regulatory regimes, they are used to develop labelling requirements.

Acute toxicity data (oral, dermal and inhalation) are legally required for plant protection products and for feed additives because of worker exposure and sometimes in order to set an Acute Reference Dose (ARfD) for consumer exposure to pesticide residues in food. However, they are not required for food additives, food contact materials or newly expressed proteins in GMOs. For these types of substances they do not serve any regulatory purpose.

4.1.2.1. Acute oral toxicity

Acute oral toxicity tests on animals have in the past used mortality as the main observational endpoint, usually in order to determine LD_{50} or LC_{50} values. The conventional acute oral toxicity test (OECD TG 401) has been heavily criticised in terms of animal welfare and this concern was the driving force behind the development of three alternative tests for acute oral toxicity and for the deletion of the OECD TG 401.

Modifications to the classical LD_{50} test include the Fixed Dose Procedure (EU method B.1bis /OECD TG 420), the Acute Toxic Class method (EU method B. 1tris/OECD TG 423), and the Up-and-Down Procedure (OECD TG 425).

Guidance on the selection of the most appropriate test method for a given purpose can be found in the OECD Guidance Document 24 on Acute Oral Toxicity Testing (OECD, 2001). This guidance document also contains additional information on the conduct and interpretation of Testing Method B. 1*bis*/OECD TG 420.

All available information on the test substance should be considered prior to conducting the study. Such information will include the identity and chemical structure of the substance; its physico-chemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the substance; toxicological data on structurally related substances; and the anticipated use(s) of the substance. This information is necessary to satisfy all concerns that the test is relevant for the protection of human health, and will help in the selection of an appropriate starting dose.

It is a principle of the three current test methods that in the main study only moderately toxic doses are used, and that administration of doses that are expected to be lethal should be avoided. Also, doses that are known to cause marked pain and distress, due to corrosive or severely irritant actions, should not be administered. Moribund animals or animals obviously in pain or showing signs of severe and enduring distress shall be humanely killed (EFSA 2005), and are considered in the interpretation of the test results in the same way as animals that died on test. Criteria for making the decision to kill moribund or severely suffering animals, and guidance on the recognition of predictable or impending death, are the subject of a separate OECD Guidance Document (OECD GD 19, 2000).

Animal welfare concerns: It is noted that three alternative tests still accept clinical signs of pending death, predicted death and moribund conditions. However, these signs, given as guides for the termination of a study and to replace death, still indicate significant pain and distress for the test animals. Such suffering could be minimised through more refined protocols and earlier endpoints. The ability to recognise and implement pre-painful and early clinical signs of distress and, sometimes its alleviation or treatment, may enable a better discrimination of the actual toxicity of the test substance.

Recent and future developments for acute oral toxicity

There are currently no *in vitro* tests that have been officially adopted by EU or OECD for assessment of acute toxicity. However, taking into account the proposal to determine the starting dose for acute oral toxicity from cytotoxicity data (Halle et al., 1997), two *in vitro* tests for acute toxicity have been validated[14]: *in vitro* basal cytotoxicity assays, based on neutral red uptake (NRU) by cells (the BALB/c 3T3 mouse fibroblast NRU and normal human keratinocyte NRU assays). These methods could be used for predicting starting doses for *in vivo* oral toxicity tests and lethal concentrations in man and they have undergone peer review by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). ICCVAM recommended that the *in vitro* basal cytotoxicity test methods evaluated in the joint ICCVAM/ECVAM-IHCP-JRC validation study should be considered before using animal testing for acute oral toxicity, and that the methods should be used where determined appropriate. Data from the test methods should be used following a weight-of-evidence approach for determining starting doses for *in vivo* studies. Using these *in vitro* methods where appropriate is expected to reduce the number of animals required for each toxicity test. However, ICCVAM concluded that the *in vitro* test methods are not sufficiently accurate to replace animals for regulatory hazard classification purposes. In February 2008, ICCVAM forwarded recommendations on the use of *in vitro* test methods for estimating starting doses for acute oral systemic toxicity tests to federal USA Agencies, which responded positively to the proposal that aims to reduce the number of animals used in safety testing for regulatory purposes[15]. The *in vitro* basal cytotoxicity test methods to estimate starting dose for *in vivo* tests have also been included in the European Chemicals Agency "Guidance on information requirements and chemical safety assessment (chapter R.7a)" developed by ECHA (ECHA 2008b).

Acute toxicity studies in animals are considered necessary for pharmaceuticals intended for human use. The studies are carried out, usually in rodents, to support marketing of new drugs and to identify the minimum lethal dose. A European initiative including 18 companies has undertaken an evidence-based review of acute toxicity studies and assessed the value of the data generated. Preclinical and clinical information was shared on 74 compounds. The analysis indicated acute toxicity data was not used to (i) terminate drugs from development (ii) support dose selection for repeat dose studies in animals or (iii) to set doses in the first clinical trials in humans. The conclusion of the working group is that acute toxicity studies are not needed prior to first clinical trials in humans. Instead, information can be obtained from other studies, which are performed at more relevant doses for humans and are already an integral part of drug development. The conclusions have been discussed and agreed with representatives of regulatory bodies from the US, Japan and Europe (Robinson et al., 2008).

[14] http://iccvam.niehs.nih.gov/methods/invidocs/panelrptpaTpanelrpt.htm
[15] http://iccvam.niehs.nih.gov/methods/acutetox/inv_nru_recommend.htm

The ACuteTox project, which is funded within the 6th Framework Program (FP6) of the DG Research of the European Commission, has the ambitious overall objective of developing an *in vitro* test strategy sufficiently robust and powerful to completely replace *in vivo* testing of acute toxicity of chemicals (for more details see also Annex 1).

4.1.2.2. Acute dermal toxicity

For acute dermal toxicity testing, the only method accepted for regulatory purposes is the *in vivo* EU method B.3/OECD TG 402. A proposal to replace the current method for acute dermal toxicity testing by applying one or more of the sequential testing procedures that are used for acute oral toxicity testing has however been submitted to the OECD and is currently being evaluated. Moreover, the scientists who have successfully developed and validated the acute toxic class method (EU method B1.tris/OECD TG 423), have provided biostatistical evidence that the acute toxic class method has the potential to replace the current *in vivo* method for acute dermal toxicity testing (Holzhütter et al., 2003).

Animal welfare concerns: test substances may cause irritation of the skin where the nociceptors are present and cause pain and irritation that are unrelievable by the animal's normal responses e.g. licking. This will also lead to forms of mental distress.

4.1.2.3. Acute inhalation toxicity

For acute inhalation toxicity testing, the only method accepted for regulatory purposes is the *in vivo* EU method B.2/OECD TG 403. Taking into account the sequential testing procedures that allow a reduction in the number of animals in acute oral toxicity testing, the OECD has recently initiated a revision of TG 403 in order to reduce test animal numbers[16].

Similar to the situation described above for acute dermal toxicity, a draft OECD TG entitled "Test Guideline 436. Acute Inhalation Toxicity – Acute Toxic Class (ATC) Method" has been finalised by OECD experts and may be accepted in the near future both as a new OECD TG 436[17] and also as an EU method.

The general approach for acute inhalation toxicity testing has been reviewed by OECD experts and is described in a *draft* OECD "Guidance Document on Acute Inhalation Toxicity Testing" (OECD Guidance Document 39, OECD 2008). The suggestions indicated in this guidance document should ideally be followed in order to reduce the number of animals used for acute inhalation toxicity testing.

Animal welfare concerns: test substances may cause irritation of the upper respiratory tract where the nociceptors are present and cause pain and irritation that are unrelievable.

Conclusions on acute toxicity:

1- The classical oral LD_{50} method has been deleted from the list of OECD testing methods in 2001 and is not allowed to be used any more. Alternative methods for the estimation of acute oral toxicity are now available (EU method B.1 bis/OECD TG 420; EU method B.1 tris/OECD 423; OECD TG 425), which reduce the number of animal used and reduce pain and suffering.

[16] http://www.oecd.org/dataoecd/54/55/41761261.pdf
[17] http://www.oecd.org/dataoecd/54/11/41761436.pdf

2- To further reduce the use of test animals, the starting dose for acute oral toxicity testing may be determined by using *in vitro* cytotoxicity data as recommended in the "Guidance on information requirements and chemical safety assessment (chapter R.7a)" developed by ECHA.

3- A proposal to replace the current methods for acute dermal and inhalation toxicity testing by applying one or more of the sequential testing procedures in use for acute oral toxicity testing has been submitted to the OECD and is currently being evaluated.

4.1.3. Skin irritation/corrosion

Data on skin irritation and corrosion are legally required for plant protection products and for feed additives. In both fields, data on skin irritation are required to assess the safety for workers. The experimental data are used for hazard classification and labelling.

Animal welfare concerns: test substances may cause irritation of the skin where the nociceptors are present and cause pain and irritation that are unrelievable by the animal's normal responses e.g. licking. This will also to lead to frustration and other forms of mental distress.

4.1.3.1. Skin irritation

Skin irritation is defined as the production of "reversible damage of the skin following the application of a test substance for up to 4 hours". The only method accepted for regulatory purposes is EU method B.4/OECD TG 404 for "Acute dermal irritation/corrosion". This test assesses the potential of a substance to cause erythema and/or oedema after a single topical application on rabbit skin, based on the Draize score (Draize et al., 1944). This test may be soon replaced by the successfully validated *in vitro* method using human skin models. These models have been endorsed by the ECVAM Scientific Advisory Committee (ESAC) in 2007 and 2008 (for more information visit ECVAM-IHCP-JRC website[18]). The ESAC concluded that the performance of these assays met the criteria outlined to be considered to have sufficient accuracy and reliability for prediction of skin irritation according to the EU criteria for risk phrase R38 (classification and labelling for skin irritation).
At the OECD level, a *draft* TG entitled "*In vitro* skin irritation: human skin model test[19]" is currently being evaluated by national experts. It is anticipated that the *in vitro* test will be the only skin irritation test accepted for regulatory purposes. The European Commission intends to adopt this method (draft test method EU method B.46 "*In vitro* skin irritation: reconstructed epidermis model test") by incorporating it in the first adaptation to technical progress of the test methods Regulation[20] before summer 2009.

4.1.3.2. Skin corrosion

Skin corrosion tests assess the potential of a substance to cause irreversible damage to the skin, i.e. visible necrosis through the epidermis and into the dermis, following the application of a test substance. In the past, skin corrosion was assessed by using animal studies such as those according to EU method B.4/OECD TG 404 "Acute toxicity: dermal irritation/corrosion". However, recently two alternative methods have been accepted officially for regulatory purposes, EU method B.40/OECD TG 430, "*In vitro* skin corrosion: transcutaneous electrical resistance test (TER)" and EU method B.40 bis/OECD TG 431 "*In vitro* skin corrosion: human

[18] http://ecvam.jrc.ec.europa.eu/
[19] http://www.oecd.org/dataoecd/21/56/40793105.doc
[20] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:142:0001:0739:EN:PDF

skin model test". Another test for skin corrosion is the "*In Vitro* membrane barrier test method for skin corrosion" (OECD TG 435), based on the Corrositex® test method, which uses proprietary biomembrane to assess the dermal corrosive potential of chemicals and chemical mixtures.

In the interest of sound science and animal welfare, information on a substance relevant to its potential skin corrosivity/irritancy should be evaluated prior to considering *in vivo* testing. Sufficient evidence may already exist to classify a test substance as to its dermal corrosion or irritation potential without the need to conduct testing in laboratory animals. A weight-of-the-evidence analysis can be used to evaluate existing information regarding the likely skin irritation and corrosion of substances to determine whether additional studies, other than *in vivo* dermal studies, should be performed to help characterise such potential. Where further studies are needed, the sequential testing strategy could be utilised to develop the relevant experimental data. For substances which have no history of testing, the sequential testing strategy described below could also be utilised to develop the data set needed to evaluate its dermal corrosion/irritation potential (Supplement to the OECD TG 404). Although this sequential testing strategy is not an integral part of testing EU method B.4, it reflects the preferred approach for the determination of skin irritation/corrosion characteristics. This approach represents best practice for *in vivo* testing for skin irritation/corrosion.

4.1.3.3. Sequential testing strategies for skin irritation/corrosion

In 2002, a supplement to OECD TG 404 for skin irritation and corrosion was published, which recommends a sequential (stepwise) testing strategy for hazard identification. Some general refinement provisions have been introduced which allow for best scientific practices and cause less animal suffering. A validation study carried out by ECVAM-IHCP-JRC confirmed the usefulness of pH as a predictor of skin corrosion potential, and provided a new prediction model for identifying chemicals that are corrosive by a pH-dependent mechanism (Worth & Cronin, 2001). The evaluation of a two-step strategy, based on the sequential use of pH measurements and *in vitro* data, indicated that the use of pH data in addition to the transcutaneous electrical resistance (TER) test or the use of a reconstructed human epithelium *in vitro* data (EU method B.40/OECD TG 430 and EU method B.40bis/TG 431), improves the ability to predict corrosion potential (EU annex 2D to Commission Directive 2004/73/EC). An evaluation of a three-step strategy, based on the sequential use of structure-activity relationships (SARs), pH measurements and *in vitro* data, indicated that tiered approaches provide an effective means of classifying chemicals, while reducing and refining the use of animals. Various SARs for skin corrosion have been reported (Gerner et al., 2004). A decision support system developed by the German Federal Institute for Risk Assessment (BfR) uses physico-chemical exclusion rules to predict the absence of skin irritation/corrosion potential in combination with structural inclusion rules (SARs) to predict the presence of such potential (Gerner et al., 2000a and b). The exclusion rules are based on physico-chemical properties such as molecular weight, aqueous solubility, and log K_{ow}, whereas the inclusion rules are based on substructural molecular features. The physico-chemical rules implicitly take into account bioavailability (skin penetration) whereas the structural rules take reactivity into account. The physico-chemical and structural rulebases are designed to predict the EU risk phrases for skin irritation (R38) and skin corrosion (R34 and R35). Further details are given in QSAR Reporting Format for the BfR skin and eye irritation rulebases[21].

[21] http://www.ecb.jrc.it/QSAR

Conclusions for skin irritation/corrosion:

1 - For **skin irritation testing,** only *in vivo* testing according to EU method B.4/OECD TG 404 is accepted for regulatory purposes at the international level. However, the successfully validated *in vitro* method applying human skin models has been endorsed by the ECVAM-IHCP-JRC Scientific Advisory Committee in 2007 and 2008, and will be incorporated into the first adaptation to technical progress of the EU test methods Regulation in 2009. At the OECD level, this method, known as *draft* TG "*In vitro* skin irritation: human skin model test" (OECD, 2008) is currently being evaluated by national experts. It is anticipated that the *in vitro* test will be the only skin irritation test accepted for regulatory purposes.

2- For the area of **skin corrosion,** three *in vitro* alternatives test methods are available, which have undergone rigorous validation. Thus at present, only *in vitro* methods can be used in accordance with Directive 86/609 for skin corrosion (EU methods B.40/OECD TG 430 or B. 40bis OECD TG 431 or OECD TG 435). Among them, EU method B.40bis/OECD TG 431, which uses human skin models, is the most commonly used method.

4.1.4. Skin sensitisation

As for skin irritation and corrosion, data on skin sensitisation are legally required as core studies for the evaluation of the safety of the plant protection products (active substances) and for the safety assessment for workers in the field of feed additives.

At present, there are three *in vivo* methods available and accepted by regulatory authorities for identifying skin sensitising chemicals and for confirming that chemicals lack a significant potential to cause skin sensitisation: the mouse Local Lymph Node Assay (LLNA) (EU method B.42/OECD TG 429), the Magnusson Kligman Guinea-pig Maximisation Test and the Buehler Guinea-pig Test (EU method B.6/OECD TG 406). Since the LLNA uses fewer animals and causes less pain and distress it should be used as a preferred method as stated in the "Guidance on information requirements and chemical safety assessment (chapter R.7a)" developed by ECHA (ECHA, 2008b) where it is stated that "The LLNA is the first choice method for *in vivo* testing. Only in exceptional circumstances should another test be used. Justification for the use of another test shall be provided".
As mentioned in the EU B.42/OECD TG 429, the LLNA provides advantages with regard to both scientific progress and animal welfare compared to the other two methods. It studies the induction phase of skin sensitisation and provides quantitative data suitable for dose response assessment.

The LLNA is an *in vivo* method and, as a consequence, will not eliminate the use of animals in the assessment of contact sensitising activity. It has, however, the potential to reduce the number of animals required for this purpose. Moreover, the LLNA offers a substantial refinement of the way in which animals are used for contact sensitisation testing. The LLNA is based upon consideration of immunological events stimulated by chemicals during the induction phase of sensitisation (Kimber et al., 2002). Despite the advantages of the LLNA over traditional Guinea Pig maximisation tests, it should be recognised that there are certain limitations that may necessitate the use of traditional guinea pigs tests (e.g. false negative findings in the LLNA with certain metals, false positive findings with certain skin irritants).

Animal welfare concerns: test substances may cause sensitisation of the skin where the nociceptors are present and cause pain and irritation that are unrelievable by the animal's

normal responses e.g. licking. This will also to lead to frustration and other forms of mental distress.

Recent and future developments for skin sensitization

Refinements to the existing methods may also be possible. For example, it might be feasible to conduct a reduced version of the LLNA (rLLNA) in which assessments are made on the basis of results from a vehicle control and a single (highest) concentration of the test substance (Kimber et al., 2006). The ECVAM Scientific Advisory Committee (ESAC) established a peer review Panel to evaluate if there was the potential to minimise animal use by employing the rLLNA as a screening test as part of a tiered-testing strategy. In April 2007, ESAC concluded that the rLLNA can be used within tiered-testing strategies to reliably distinguish between substances that are skin sensitisers and non sensitisers. Subsequent to the ECVAM-IHCP-JRC peer review of the rLLNA, the US Interagency Coordinating Committee on the Validation of Alternative Methods – (ICCVAM) (Haneke et al., 2001) evaluated the performance the rLLNA with a larger set of data, coming to very similar conclusions[22]. The ICCVAM recommendations on the proposed application of the rLLNA are currently being finalised. Despite the positive outcome of two independent peer review processes, concerns were raised by the EU Member States Regulatory Authorities, about its slightly lower sensitivity with respect to the standard test and its inability to provide information on potency. Nevertheless this test is mentioned in the "Guidance on information requirements and chemical safety assessment (chapter R.7a)" developed by ECHA[23] .

Various *in silico* and *in vitro* models are being developed to assess the skin sensitising potential of chemicals and products. The two — not mutually exclusive — approaches for the substitution of *in vivo* sensitisation include computer-based expert or QSAR systems, and *in vitro* methods that are mostly based on chemically-induced responses of cell culture systems. Among these, (Q)SAR systems are the most developed and it is envisaged that further optimisation of such methods, achievable by extending their range of chemical knowledge, will lead to formal validation within the next few years. The development process of (Q)SARs is mainly carried out by consortia and companies that develop and distribute these computer programs. An approach to the evaluation of these systems has been defined at the OECD level (OECD 2004), but at present they are not yet accepted for regulatory purposes.

Although cell-based systems have, in some cases, been shown to be able to discriminate between sensitisers and non-sensitisers (and for this reason could be used for priority setting), they are still at the research/optimisation level. Advances in immunological research continue to increase the number of potential test parameters, such as the up-regulation or down-regulation of the expression of cell membrane proteins and cell–cell signalling molecules, such as interleukins, and changes in the antigen uptake process. Other promising screening approaches are those based on the measurement of protein/peptide reactivity. Despite these advances in the field, further research and refinement activities are envisaged before an alternative test battery for full replacement will be available.

The DG Research is funding this approach via the Integrated Project **Sens-it-iv** as described in the annex 1.

Conclusions for skin sensitisation:

[22] http://iccvam.niehs.nih.gov/methods/immunotox/LLNA-LD/IWGrecLLNA-LD07Jan08FD.pdf)
[23] http://guidance.echa.europa.eu/guidance_en.htm

1- The Local Lymph Node Assay (LLNA) has been recommended among others by the European Chemicals Agency as the method of choice for skin sensitisation. EFSA considers this method should be used whenever possible.

2- The reduced rLLNA has currently been rejected by EU Member States but is undergoing further evaluation at the OECD level.

4.1.5. Eye irritation/corrosion

Data on eye irritation and corrosion are legally required as core studies for the evaluation of the safety of plant protection products (active substances) and for the safety assessment for workers in the field of feed additives.

Animal welfare concerns: The cornea is particularly sensitive to substances that are irritant and corrosive. The animals respond severely through tear production and nictitating membrane activity and sometimes they vocalise with very irritant/corrosive substances. The animals may also show evidence of mental distress as they are unable to make a normal response to the irritation.

4.1.5.1. The Draize rabbit eye test

The conventional test for the eye irritant and corrosive potential of chemicals is the rabbit eye test, which was developed by Draize et al. (1944) and which has become the international standard assay for acute ocular toxicity (EU method B.5/OECD TG 405). A variety of different scoring systems for assessing the extent of injury to the corneal, the iridial and the conjunctival compartments are currently applied in different regulations, ranging from maximum single tissue scores to averaged weighted sum scores for all three tissues.

The test can be very painful to the rabbits (even if local anaesthetic can be used) and has two other major shortcomings: subjectivity in the allocation of the respective scores, i.e. low interlaboratory reproducibility (Weil & Scala, 1971), and differences in sensitivity to tested substances between rabbits and humans (Christian & Diener, 1996).

Reduction and refinement approaches are now included in a sequential testing strategy published as a Supplement to EU method B.5/OECD TG 405. Following that strategy, the *in vivo* rabbit eye test needs only to be performed as a last step, i.e. when the assessments in all the other tiers have produced negative results, to assess mildly to moderately irritant compounds. However, this strategy does not eliminate the need for an *in vivo* test, and there is potential for over-classification of the eye irritancy potentials of chemicals.

4.1.5.2. Recent and future developments for eye irritation

Success in developing and validating alternative tests to fully replace the Draize rabbit eye irritation test has remained elusive, despite major efforts by ECVAM-IHCP-JRC, industry trade associations, individual companies and academia. Six major evaluation studies were conducted between 1991 and 1997. The outcome of each of the validation studies was summarised by Balls et al. (1995 and 1999). No test was found to be capable of replacing the Draize rabbit eye test, but some of the assays showed considerable promise as screens for ocular irritancy. The main reason for this is the difficulty of comparing *in vitro* test results with historical animal data, since the subjective scoring of tissue lesions in the eye in the Draize test provides variable estimates of eye irritancy. Other possible contributing reasons for the outcomes of these evaluation studies are: a) the *in vitro* tests only partially modelled the complex *in vivo* eye irritation response; b) the protocols and prediction models might have been

insufficiently developed; and c) the choice of statistical approaches for analysing the data might not have been appropriate (Balls et al., 1999).

The alternative methods for eye irritation testing comprise organotypic models (Chamberlain et al., 1997), such as isolated eyes or eye components, tissue and cell culture systems, and physicochemical tests. The basis of the organotypic models is the *in vitro*/*ex vivo* exposure to chemical irritants of isolated eyes, corneas or lenses of bovine, porcine, chicken or rabbit origin. The most widely used organotypic models are the Bovine Corneal Opacity and Permeability (BCOP) assay, the Isolated Rabbit Eye (IRE) test, the Isolated Chicken Eye (ICE) and the Hen's Egg Test on the Chorio-Allantoic Membrane (HET-CAM). Although several of these *in vitro* tests have performed successfully in several of the international validation studies described above, none of them has to date officially been accepted for regulatory purposes at the international level.

Four cytotoxicity and cell/tissue function based assays such as the Red Blood Cell haemolysis test, the Neutral Red Uptake assay, the Fluorescein Leakage test and the Silicon/Cytosensor Microphysiometer assay are currently undergoing a retrospective evaluation by ECVAM-IHCP-JRC. More recently Human Cornea Epithelial (HCE) constructs have shown promising results and validation may be considered.

Most of the evaluation studies on alternatives for eye irritation have failed to identify a single replacement assay for the *in vivo* Draize rabbit eye test. It is generally recognised that the range of criteria for injury and inflammation covered by the Draize rabbit eye test is unlikely to be replaced by a single *in vitro* test. To overcome this, several national and international organisations have recommended the use of a tiered testing procedure for reducing the number of animals used. The proponents of tiered testing strategies include the OECD (2002), and the regulatory authorities in the USA (Silva et al., 1997) and in the EU Member States (Schlede & Gerner, 1995), as well as ECVAM-IHCP-JRC (Balls et al., 1999). Within the aim of total animal test replacement, these strategies could be designed to utilise the strengths of particular *in vitro* assay systems to address the required ranges of irritation potential and/or chemical classes.

Recently ICCVAM has published a report on "*In vitro* ocular toxicity testing methods for identifying severe irritants and corrosives". The report recommends the bovine cornea opacity and permeability (BCOP) and the isolated chicken eye (ICE) test methods, with specific limitations for certain chemical classes and/or physical properties, as methods to be used in a tiered strategy to determine ocular hazards. Substances that are positive in these tests can be classified as ocular corrosive or severe irritants without further testing in animals. This approach, if accepted by regulatory authorities, will result in a reduction of the number of animals used for the identification of permanent or temporary damage to the eye. Meanwhile the USA has submitted draft test guidelines for the BCOP and ICE to the OECD and they are currently being evaluated by national experts[24]. If the above-mentioned recommendations are followed and the tests prove to be valid, the most promising methods (BCOP, ICE, HET-CAM and HCE models) could also be accepted at more international level by the OECD and be endorsed for defined applications, within a few years. However, so far no *in vitro* test hold promise for assessing the most difficult *in vivo* endpoints, e.g. reversibility, persistence or mechanical irritation.

Various SAR(s) have been developed for eye irritation and corrosion. The decision support system developed by the German Federal Institute for Risk Assessment (BfR) uses physico-

[24] http://www.oecd.org/dataoecd/37/8/41896718.pdf.

chemical exclusion rules to predict the absence of eye irritation/corrosion potential in combination with structural inclusion rules (SARs) to predict the presence of such potential (Gerner et al., 2005). These rules are used analogously to those described in the skin irritation and corrosion section above. The physico-chemical and structural rulebases are designed to predict the EU risk phrases for eye irritation (R36) and severe eye irritation/corrosion (R41). Independent validation exercises by the former European Chemical Bureau-IHCP-JRC support the performance of the physico-chemical rulebase for predicting the absence of eye effects as well as the performance of the structural rulebase for predicting the occurrence of eye effects (Tsakovska et al., 2007).

Conclusions for eye irritation:

1- The Draize rabbit eye test (EU method B.5/OECD TG 405) is still the standard for eye irritation testing in Europe. However, positive results from four *in vitro* assays, the BCOP, ICE, IRE and HET-CAM are accepted by the EU to classify severe eye irritants. Substances that are positive in these tests can be classified as ocular corrosive or severe irritants without further testing in animals.

2- Four cytotoxicity and cell/tissue function based assays such as the Red Blood Cell haemolysis test, the Neutral Red Uptake assay, the Fluorescein Leakage test and the Silicon/Cytosensor Microphysiometer assay are currently undergoing a retrospective evaluation by ECVAM-IHCP-JRC.

3- The BfR decision support system can be used for predicting the absence of eye irritation/corrosion.

4.1.6. Genotoxicity

In the area of food and feed safety, genotoxicity testing is a key part of the core studies e.g. for the evaluation of food additives, smoke flavourings and food contact materials, plant protection products, feed additives and contaminants.

The aim of testing for genotoxicity is to assess the potential of substances to induce genotoxic effects which may lead to cancer or cause heritable damage in humans. The usual strategy used with genotoxicity testing is starting with *in vitro* testing, followed by *in vivo* testing if positive results are obtained *in vitro*.

For a comprehensive coverage of the potential genotoxicity of a substance, information on gene mutations (base substitutions and deletions/additions), structural chromosome aberrations (breaks and rearrangements) and numerical chromosome aberrations (loss or gain of chromosomes, defined as aneuploidy) is normally required by all EFSA Scientific Panels dealing with assessment of chemical substances for EU authorisation. Assays for DNA damage constitute a third group, including, among others Unscheduled DNA synthesis in mammalian cells *in vitro* (EU B.18/OECD TG 482). Normally three (or in some cases two) *in vitro* tests are required, namely a test for induction of gene mutation in bacteria, a test for induction of gene mutations in mammalian cells and a test for induction of chromosomal aberrations in mammalian cells (respectively EU B.13/14-OECD TG 471and B.15/OECD TG 480; B.17/OECD TG 476; B.10/OECD TG 473).

Another newer *in vitro* test for detecting potential genotoxicity of a substance is the *in vitro* micronucleus test (MNvit). The test has recently been validated by ECVAM-IHCP-JRC based on existing data and has been endorsed by the ECVAM-IHCP-JRC Scientific Advisory

Committee as a valid alternative to the *in vitro* chromosomal aberration test (CAT). The MNvit has been shown to be a robust test with greater accuracy and because thousands instead of hundreds of cells can be assessed, the statistical power is greater compared to the CAT (Corvi et al., 2008). In addition to clastogenic effects, it can also be used to identify aneugens. The CAT is less reliable for this purpose as it is difficult to distinguish between aneugenic cells and artefacts in which chromosomes might be displaced from cells. The OECD is in the process of drafting a guideline on the MNvit (OECD draft TG 487) that will most probably be finalised during 2009. The test has already been integrated in the guidance document of the REACH legislation (Guidance on information requirements and chemical safety assessment (chapter R.7a) developed by ECHA).

Differences exist in the testing requirements for genotoxicity among the different EFSA Scientific Panels. While two *in vitro* tests are considered to be enough to assess the genotoxicity of feed additives such as micro-organisms and enzymes, three *in vitro* tests are required for feed additives other than microorganisms and enzymes, for food additives, smoke flavourings, food contact materials and for plant protection products. Some of these differences are laid down in legal requirements, for example, those on feed additives.

Positive *in vitro* results (i.e. those showing a substance has genotoxic potential) normally require follow up with *in vivo* testing. The most commonly used *in vivo* genotoxicity assays, often use to confirm positive *in vitro* results are:

- Test for the detection of damage to chromosomes or the mitotic apparatus, namely the mammalian erythrocyte micronucleus test, EU B. 12/OECD TG 474, and the mammalian bone marrow chromosome aberration test, EU B. 11/OECD TG 475.
- Unscheduled DNA synthesis (UDS) test to investigate DNA damages in mammalian liver cells *in vivo* (EU method B. 39/OECD TG 486)
- *In vivo* COMET assay. There is no OECD test guideline for this test but a validation study is ongoing in Japan coordinated by the Japanese Centre for the Validation of Alternative Methods (JaCVAM)[25].

The EFSA PPR Panel has proposed that an *in vivo* study is not needed if the results of the *in vitro* studies are negative (EFSA 2007a).

Animal welfare concerns: would be those associated with *in vivo* tests in which the test substances may be given at toxic doses and will vary considerably. They include variation according to species and administration and sampling methods.

Recent and future developments:

There are several possibilities to be considered in the interest of ensuring that the number of animals used in genotoxicity testing is kept to a minimum (e.g. the harmonisation project "Mutagenicity testing for chemical risk assessment", WHO 2007).

Understanding the toxicokinetic properties of the test substance before undertaking animal tests will enable development of appropriate study designs for the *in vivo* tests. This applies especially with respect to tissue(s) to be investigated, the route of substance administration and the highest dose to be tested. If the systemic availability of a test substance is not understood at this stage, toxicokinetic investigations or modelling may be necessary.

[25] http://altweb.jhsph.edu/wc6/paper483.pdf

When performing animal tests, both males and females do not necessarily need to be used. In accordance with standard guidelines (e.g. EU B. 12/OECD TG 474; EU B. 11/OECD TG 475; EU B. 39/OECD TG 486) testing in one sex only is possible when the substance has been investigated for general toxicity and no sex-specific differences in toxicity have been observed. If the test is performed in a laboratory with substantial experience and historical data, it should be considered whether a concurrent positive control and a concurrent negative control for all time points (*e.g.* for both the 24h and 48 h time point in the micronucleus assay) will really be necessary (Krishna and Hayashi, 2000). The possibility to sample positive and negative controls at only single time points is already foreseen in the OECD test guidelines for the MNvit and CAT.

Another possibility for reduction of animal use would be to incorporate, if scientifically appropriate, the *in vivo* genotoxicity test into a short-term repeated-dose toxicity test (such as a 28-day oral toxicity test). All these recommendations have already been included in the "Guidance on information requirements and chemical safety assessment (chapter R.7a)" developed by ECHA.[26]

There are concerns that *in vitro* genotoxicity tests in mammalian cells produce a high number of irrelevant positive results (e.g. when compared with data on rodent carcinogenicity). As reported by Kirkland et al. (2005, 2006, 2007) the rate of positives in *in vitro* tests may trigger an increased number of *in vivo* genotoxicity testing. According to Kirkland et al. (2007) there would be a significant reduction in the number of animals used if *in vitro* tests were more predictive for *in vivo* genotoxicity and carcinogenicity (i.e. fewer false positives).

In April 2006, a workshop was held at ECVAM-IHCP-JRC with the aim to address the high rate of *in vitro* false positive results when performing genotoxicity testing (Kirkland et al., 2007). The workshop addressed the possibilities to improve the current battery of *in vitro* tests in order to have a lower false-positive rate (Kirkland et al., 2008). This is a crucial step in the development of a strategy to reduce unnecessary *in vivo* tests. The recommendations developed during the workshop are currently being followed up by different groups and organisations, including COLIPA (the European Cosmetics Association), ECVAM-IHCP-JRC and NC3Rs (the UK National Centre for Replacement, Reduction and Refinement of animals in research).

Conclusions for genotoxicity:

1- Several *in vitro* genotoxicity test methods are available and currently used as the first and sometimes the only tests necessary.

2- If there is evidence of genotoxicity *in vitro*, *in vivo* genotoxicity testing is usually required to demonstrate whether the genotoxic potential observed in *in vitro* tests is expressed *in vivo*. Ongoing research continues to refine the predictive ability of *in vitro* tests and to develop strategies for follow-up of substances showing genotoxic potential *in vitro*.

3- The possibility of reduction should be considered when performing an *in vivo* genotoxicity test. OECD test guidelines already allow the possibility to reduce the number of animals used in *in vivo* tests sampling positive and negative controls at a single timepoint and using only one sex instead of both sexes.

[26] http://reach.jrc.it/docs/guidance_document/information_requirements_en.htm#A

4.2. Mammalian systemic and repeated dose toxicity

The following sections discuss longer-duration mammalian toxicity tests which are designed to detect adverse effects in diverse biological systems. Exposure to chemicals can bring about many different types of disturbance to physiological, cellular and biochemical homeostasis, which may be of sufficient magnitude and duration to perturb the organism and result in pathological or other lasting adverse consequences. Because of the complexity of the underlying biology and the many possible mechanisms by which chemicals can cause damage, even within the same cell, tissue or organ, progress in developing alternative tests that may reduce, refine or replace existing *in vivo* methods has been slower than in the toxicity endpoints discussed earlier in sections 4.1.1 - 4.1.6. While a number of relevant *in vitro* tests have been developed, they are usually inadequate, by themselves, to fully cover the entire range of biological processes that need to be assessed.

Animal welfare concerns: would be those associated with giving substances at toxic doses and will vary considerably. They include variation according to species and administration and sampling methods. Repeated administration can be stressful for animals.

4.2.1. Repeated-Dose Toxicity

Data on repeated-dose toxicity are required for the safety assessment of substances in most of the areas covered by the EFSA Scientific Panels. They can be of various durations, most commonly 28 days and 90 days (subchronic test), 1 year or more (chronic tests). While sub-chronic studies provide information on the possible health hazards likely to arise from repeated exposure over a relative limited period of time, chronic studies give information on possible risks arising from prolonged, repeated exposure to the identified hazard, covering a major portion of the average lifespan of the experimental animals.

The administration can be oral, dermal, or via inhalation. The data from these studies provide information on the effects of cumulative exposure of target organs and on general health hazards likely to occur as a consequence of repeated exposure to a chemical. As these studies are aimed at detecting any systemic toxic effect, a wide variety of parameters are monitored, usually including clinical appearance, ophthalmological examination, body weight, food consumption, haematology, clinical biochemistry, and pathology including histopathology.

Repeated-dose studies should provide information, not only about adverse effects, but also about the amount of the substance that can be tolerated without adverse effect (e.g. identification of a no-observed-adverse-effect level - NOAEL) and, as such, are essential studies for risk assessment and for the setting of health-based guidance values both for occupational and consumer exposure. When no NOAEL can be identified, the benchmark dose (BMD) approach may be applied to use the study for deriving health-based guidance values rather than repeating the study (EFSA Opinion on BMD – under preparation). Subchronic studies can also provide valuable information about the design of subsequent chronic toxicity studies and carcinogenicity studies. Data from 28-day studies are not mandatory for any of the types of substance that EFSA Scientific Panels assess, but 90-day studies are normally required in most areas (e.g. food additives, smoke flavourings, plant protection products and some feed additives).

In cases where testing of the safety and nutritional value of whole (GM) foods and feed is required, a subchronic, 90-day rodent feeding study may be conducted (EFSA, 2008).

All laboratory animals can be used in repeated-dose toxicity studies, including rodents (most often rats and mice) and non-rodents, of which the mini-pig and the dog are the preferred species (however, for non-rodents, specific test guidelines within the EU and OECD programmes only exist for a 90-day oral study).

The available test guidelines for repeated-dose toxicity tests are the following:
Repeated Dose 28-day Oral Toxicity Study in Rodents - EU B.7/OECD TG 407;
Repeated Dose Dermal Toxicity: 21-28 day - EU B.9/OECD TG 410;
Repeated Dose Inhalation Toxicity: 28 day or 14 day study - EU B.8/OECD TG 412;
Repeated Dose 90-day Oral Toxicity Study in Rodents - EU B.26/OECD TG 408;
Repeated Dose 90-day Oral Toxicity Study in Non-Rodents - EU B.27/OECD TG 409;
Subchronic Dermal Toxicity: 90-day Study - EU B.28/OECD TG 411;
Subchronic Inhalation Toxicity: 90-day Study - EU B.29/OECD TG 413;
Chronic Toxicity Studies - EU B.30/OECD TG 452.

Animal welfare concerns: would be those associated with giving substances at toxic doses and will vary considerably. They include variation according to species and administration and sampling methods. Repeated administration can be stressful for animals.

Recent and future developments:
The *in vivo* testing requirements for plant protection products were laid down some years ago in Council Directive 91/414/ECC, which is currently being revised, and are more extensive than for most of the other types of substance considered by EFSA Scientific Panels. The EFSA Panel on Plant Protection Products (PPR) has issued three Opinions (EFSA, 2006, 2007a, 2007b) commenting on the draft data requirements proposed for the revision of the Directive. The approach described below might also be considered and discussed by other EFSA Scientific Panels.

For plant protection products, data from a 90-day oral toxicity study on the active substance performed in rodents, usually the rat, is required, according to Directive 91/414/EEC. The use of a different rodent species must be justified. Likewise, data from subchronic or chronic oral toxicity studies on non-rodents must always be reported (EFSA, 2007a).

The PPR Panel has proposed that for plant protection products for which both subchronic and chronic studies are always available, a 90-day study in rodents may not be needed if a full 28-day study (EU B. 7/OECD TG 407) in rodents is available. The short-term toxicity of the active substance to non-rodents (usually dogs) should always be reported from a 90-day study. However, the Panel has proposed that there is no need to require a 12-month dog study in addition to a 90-day study in dogs. Any use of non-rodents other than dogs should be justified. If not all studies are used, there may be advantages in increasing the power of those studies that are conducted by using larger numbers of animals per study (while overall numbers are still reduced).

The rationale behind the PPR Panel proposal is that the 90-day study in rodents rarely has an impact on health-based guidance values when an adequate 28-day study is available. Based on extensive evidence (Doe et al., 2006), the PPR Panel has concluded that extension of a dog toxicity study beyond a 13-week duration provides little additional information, and that a 12-

month dog study therefore constitutes an unnecessary use of animals (EFSA, 2007a). In the draft data requirements of the revision of the Council Directive 91/414/EEC for plant protection products, it is proposed that a functional test (e.g. T-cell dependent antibody assay) to assess immune response after an adequate period of exposure should be included in the study plan for the 90-day rodent study (EFSA, 2007a). The PPR Panel supported the proposal only where there is a prior cause for concern about immunotoxicity – e.g. because of observed pathology of immune tissues, because the plant protection product belongs to a group of chemicals with known immunotoxicity, or based on the mechanism underlying other observed toxic effects.

A new "Joint Research Initiative (JRI) between COLIPA (the European Cosmetics Association) and the European Commission in the field of replacement of animal tests in repeated dose systemic toxicity" has been recently announced. For the first time the DG (Directorate General) Research of the EU Commission and a major industry organisation, COLIPA, will in a joint venture sponsor a major research program, which is aimed at solving a central problem in *in vitro* toxicology "repeated dose toxicity". The call for submission of project proposals will be published by DG Research in July 2009 and the work on the projects will probably start in July 2010.

Conclusions for repeated dose toxicity:

1- Repeated-dose toxicity studies are often pivotal for the risk assessment of substances and cover many endpoints, some of which could indicate the critical toxicological effect to be looked for at a later testing stage, such as effects on specific organs, indication of alterations in the immune system, potential for endocrine disruption, cell proliferation, pre-neoplastic foci, etc. A single *in vitro* test cannot reflect all these endpoints, and even a battery of *in vitro* tests is unlikely to have the potential to mimic the complexity of the physiological and toxicological responses of the whole organisms. Thus, no *in vitro* alternatives that would replace *in vivo* testing for repeated-dose toxicity are available at present.

2- The EFSA Scientific Panel on plant protection products has proposed some amendments to the legal requirements for plant protection products testing. If those suggestions were to be taken up during the revision of Council Directive 91/414/EEC, a different testing strategy would decrease the number of animals used.

4.2.2. Reproductive toxicity

Reproductive toxicity testing aims at identifying and quantifying hazards to any part of the whole reproductive cycle, from production of gametes through to fertilization, pregnancy, prenatal and early post-natal development of the progeny.

An important issue is the assessment of long-term developmental effects, i.e., those effects that are not readily apparent in the progeny but can lead to impaired functioning of critical systems later in life, e.g., the nervous, immune and reproductive systems. Indeed, the assessment of late emerging developmental effects is a major aspect of the assessment for chemicals that might impact on developmental programming, e.g., through endocrine-mediated mechanisms (see e.g., IPCS, 2002; Mantovani et al., 2006).

Four main studies exist to assess reproductive and developmental toxicity:

a. *Prenatal developmental toxicity study* (EU B. 14/OECD TG 414): this method provides a detailed assessment at the end of pregnancy of effects caused by a prenatal exposure: the specificity of effects on the offspring must be assessed in comparison with maternal toxicity. The lack of postnatal follow-up reduces the sensitivity of the protocol to some groups of chemicals, e.g., endocrine disrupters. In general, the prenatal developmental toxicity study is considered relevant especially to short term, peak exposures, such as those occurring in occupational assessment of chemicals. However, the results of such studies are also relevant for food safety, e.g., to establish an Acceptable Daily Intake for a substance, or an Acute Reference Dose for plant protection products (EFSA, 2007). Developmental toxicity is one of the few instances where a rodent and a non-rodent species (rabbit) are required. A problem in using the rabbit, however, is the lack of background information on kinetics and systemic toxicity as well as the high vulnerability to maternal stress. Whether two species are always required and whether the rabbit is the most adequate second species is still an unresolved issue, requiring further analysis of available data. Proposals for a tiered testing strategy for prenatal developmental toxicity only using one animal species, when also accompanied by a reproduction study, have been made (Cooper et al., 2006) and in the case of veterinary products, already adopted (VICH, 2002; Hurtt et al., 2003).

b. *One- and two-generation study* (EU B.34/OECD TG 415 and EU B.35/OECD TG 416, respectively): these tests measure the full impact of a chemical on reproductive function, throughout the production of one or two successive generations. They are usually performed by dietary exposure and are especially relevant to low-level, continuous exposures such as those from chemicals present in foods. The one-generation test uses fewer animals and is less time consuming compared to the two-generation reproduction toxicity study, but could also be less sensitive; however, this test has not been updated for quite a long period and the database of recent studies is rather limited. The two-generation method (for which OECD TG 416 was updated in 2001) is currently regarded as a major piece of information for toxicological risk assessment of chemicals of importance in relation to dietary or environmental exposure. Most importantly, it is the only standard protocol where laboratory animals of the F1 generation are exposed through their full lifecycle, i.e., from zygote stage through to sexual maturity. Thus, EU B.35/OECD TG 416 is the only current protocol providing information on juveniles (post-weaning to adolescence) which otherwise would be missing from current testing strategies (Seely 2008). Due to the complexity of endpoints and targets, its critical role in the overall testing strategy as well as the intensive use of animals (i.e. dams plus their litters) reproductive toxicity is a focus for Three Rs implementation.

c. *Developmental Neurotoxicity Study* (OECD TG 426): the test guideline for this study was adopted in 2007. The developmental neurotoxicity study provides information on the effects of repeated exposure to a substance during *in utero* and early postnatal development. This TG is important for testing effects on the nervous system as a critical target in the developing organism, due to its prolonged maturation and unique vulnerability to subtle perturbations during the dosing period. The preferred test species is the rat and the dosing period spans

from implantation through to end of lactation; the observation period lasts up to young adulthood. A battery of observations is performed to detect neurological and behavioural effects as well as neuropathology; although the battery is fairly comprehensive, the assessment of some relevant endpoints are not included in the guideline at present, e.g., social interaction, mating behaviour, and neuroendocrine effects. The OECD TG 426 is not a standard requirement, but is performed whenever a potential for neurotoxicity is envisaged, e.g., from systemic toxicity studies or from chemical structure similarity to known neurotoxicants, such as certain groups of agrochemicals. The OECD TG 426 can be conducted as a separate study or incorporated as a satellite study into EU B.34/OECD TG 415 and EU B.35/OECD TG 416.

d. The *in vivo* uterotrophic (OECD TG 440) and Hershberger (draft OECD TG) assays are screens for, respectively, estrogenic and (anti)androgenic activities (Gelbke et al., 2004). The uterotrophic bioassay has been recently adopted as OECD TG 440, and the adoption of a test guideline for the Hershberger bioassay may be foreseen in the near future. These tests, by measuring straightforward endpoints, can give information on potential endocrine disruption and be used in clarifying the mode of action of effects found *in vivo.* The role of both tests including their potential impact on the Three Rs in a tiered testing strategy has to be established.

Animal welfare concerns: would be those associated with giving test substances at toxic doses and will vary considerably. They include variation according to species and administration and sampling methods. Repeated administration can be stressful for animals. Welfare of the animals might concern not only the treated animals but also the offspring.

Recent and future developments:
The approaches mentioned below are relevant to possible reduction and refinement of the use of animals for reproductive toxicity testing.
The application of *in vivo* screening tests for reproductive/developmental toxicity alone or combined with systemic toxicity (OECD TG 421 and 422; Reuter et al., 2003) have been recommended for use for industrial chemicals for which no data exist.

The updated 28-day repeated-dose toxicity assay (EU B.7/OECD TG 407) will be able to identify relevant hazards in young adult organisms, such as target toxicity in reproductive tissues, sperm parameters, ovulatory cycles and endocrine disruption (Hartmann et al., 2008). This information is valuable for both prioritization for further testing and hazard identification/characterization. A significant shortcoming of the updated EU B.7/OECD TG 407 is that it is performed on young rodents that have just reached full sexual maturity, which may not be the most vulnerable phase of the lifecycle.

The most interesting development is the implementation of the *extended one-generation test* (Cooper et al., 2006). An OECD project on the validation of the extended one-generation test is ongoing and the OECD draft protocol is not yet ready for regulatory acceptance (OECD, 2008b[27])

[27] http://www.oecd.org/dataoecd/22/20/40899803.pdf

The proposal stems from meta-analysis of two-generation test data. According to the scientists proposing the extended one-generation study, the production of the second generation (F2) could be triggered instead of being mandatory. Instead, significant additional information in terms of either type of effects or dose-response, could be gained by extending the observation of the F1 generation, allowing a detailed, targeted observation of those endpoints that can be considered more relevant (e.g., neurobehavioral, immune, endocrine). Such a protocol would maintain, and even improve, the potential to identify long-term impacts on development as well as effects on juveniles. It would also provide a substantial reduction in numbers of animals used: if (in the current two-generation protocol) 20 pairs of F1 breeders x 4 dose levels (3 treatment levels plus control) produce 12 F2 pups each, avoiding the production of F2 would lead to an approximate reduction in the range of 800-1000 rats for each assessment.

Most important, reduction/refinement might be achieved successfully from more efficient *strategies*, as well as through new tests. Cooper et al. (2006) recommend a tiered approach for agricultural chemicals that makes full use of metabolism and systemic toxicity information. The proposal is based on the extended one-generation test in the rat plus a stand-alone developmental toxicity study in the rabbit. The approach aims at integrating in an efficient way several endpoints that cover the whole reproductive cycle (from gametogenesis through to maturation of the following generation), namely:

i) the use of an extended one-generation test in the rat to assess fertility and reproductive function and short- to long-term developmental effects from exposure during pregnancy, lactation and pre-pubertal phases. In this study, effects induced prenatally will be identified by such parameters as neonatal weight, number of pups and their viability, which can also identify potential developmental toxicants and support the derivation of a NOAEL; however, they do not provide a full characterisation of developmental toxicity;

ii) the use of a prenatal developmental toxicity study (with specific parameters on malformations, minor anomalies, embryo-foetal death, intrauterine growth retardation) in the rabbit; this study will detect also any developmental toxicant to which rat is not or poorly susceptible.

A previous meta-analysis by Hurtt et al. (2003), using data on active substances used in veterinary products, showed that the rat and rabbit had similar sensitivities with regard to the detection of teratogens but different specificity and that a tiered approach to prenatal developmental toxicity testing could be followed, in which, for some circumstances, testing a second species would not be necessary. Although this and the Cooper et al. (2006) papers focus on specific types of substance in food (i.e. residues of active substances from agricultural and veterinary products), the principles they explore could be applicable to other substances present in food.

As for *Replacement*, it is evident that a single, or a couple, of *in vitro* assays, however refined, could never properly identify an appropriate range of reproductive or developmental hazards.

Many *in vitro* assays targeting a number of different relevant endpoints do exist (see e.g., Bremer et al., 2005). Assays that are validated or are in the process of validation address:

a) developmental toxicity, e.g. 1) the embryonic stem cell assay, identifying effects on cell differentiation that can be relevant to early embryogenesis but also to late programming and 2) the whole-embryo rodent culture assay, identifying effects on late

embryogenesis, including embryonic adnexa and 3) the limb bud micromass test (Piersma 2004; Spielmann et al., 2004).

b) binding and transactivation of steroid nuclear receptors (ERalpha and AR) in intact cells (Sonneveld et al., 2006). The OECD Working Group of National Coordinators (WNT) has recently approved a new test guideline (TG 455) "Stably transfected human estrogen receptor-α transcriptional activation assay" for the detection of estrogenic agonist-activity of chemicals. The new TG will be soon published on the OECD website and therefore made available for future use.

Notwithstanding the many assays available, several areas are not completely covered, for instance, mechanisms of endocrine disruption other than those mediated by ERalpha and AR and effects on late foetal development. Whereas such gaps deserve due consideration, it is essential at this stage to make the best possible use of the several tests that are already available. Thus, the two main issues are:

i) to assess the potential of available tests for prevalidation or validation;
ii) to develop a *testing strategy* by *integrating* different tests that cover a range of priority endpoints.

The 6th Framework Programme Integrated Project ReProTect (Hareng et al., 2005; www.reprotect.eu) is aiming at the optimisation of these tests in order to prepare them for formal validation studies (see also Annex 1).

Conclusions on reproductive and developmental toxicity:

1- *In vivo* testing methods are still required to assess the reproductive and developmental toxicity of compounds.

2- Proposals have been made for a tiered approach to prenatal developmental toxicity testing, which under some circumstances may avoid the need for testing in a second species.

3- The two-generation and one-generation studies (EU B. 34/OECD TG 415 and EU B. 35/TG 416) are the methods of choice for risk assessment of prolonged, low-level exposures that are most relevant in the field of food safety. In particular, the two-generation study is currently the most widely used test for reproductive toxicity. Nevertheless, for scientific and animal welfare reasons, the OECD is working on a draft guideline for an *extended one-generation study protocol*. This new test will have at least the same potential as the two-generation study for identification/characterisation of long-term effects on reproduction and development, including endocrine disruption; at the same time, it would lead to a considerable reduction in the number of animals used for reproductive toxicity testing.

4- The *in vivo* uterotrophic and Hershberger assays represent potentially effective screens for, respectively, estrogenic and (anti)androgenic activities. The uterotrophic bioassay has been recently adopted as OECD TG 440, and the adoption of a test guideline for the Hershberger bioassay is expected soon.

5- A number of *in vitro* testing methods for studying specific steps in the reproduction process (e.g. cell differentiation, receptor transactivation) have already been developed. Their potential

relevance for risk assessment relies on the capacity to integrate them in a testing strategy covering a number of critical targets as well as exploiting the new molecular biology tools.

4.2.3. Carcinogenicity

Among the types of substances assessed by EFSA Scientific Panels, carcinogenicity studies are required only under Council Directive 91/414/EEC on plant protection products. For other types of substances, the cancer bioassay is not mandatory and, depending on the circumstances, it may or may not be performed.

The principal objective of carcinogenicity testing is to identify substances that may cause an increase in human cancer at any site by any mechanism. The recommended test protocol for carcinogenicity is the EU B.32/OECD TG 451. Both rats and mice are commonly used in risk assessment for humans. All routes of exposure can be applied. The chemicals are normally administered for a period of not less than 24 months for rats, or 18 months for mice and hamsters. Each dose groups employed in the test include a minimum of 50 animals of each sex, with one control group and three test groups of rodents that are no more than 6 weeks of age at the start of the dosing period.

Animal welfare concerns: would be those associated with giving test substances at toxic doses and will vary considerably. Additional concerns include the high number of animals used and the duration of the experiments.

Recent and future developments

Where possible, the animal models used for carcinogenicity testing should be biologically appropriate for the assessment of possible human risk. However, the selection of test species is usually limited, by practical considerations, to laboratory rats and mice. There is an ongoing discussion within the scientific community about whether the use of both rodent species is necessary. It has been stated, based upon a survey of databases covering studies with pharmaceuticals, that in general the rat is more sensitive than the mouse, i.e. a higher proportion of pharmaceuticals tested caused tumours in the rat than in mouse (van Oosterhout et al., 1997). Several reviews of the contribution of long-term studies in the mouse to the establishment of health-based guidance values have indicated that the use of this species does not provide any additional contribution to risk assessment (Doe et al., 2006). Hence, there is a strong case for deleting this requirement, and focusing more on understanding modes of action when assessing risks of carcinogenicity (e.g. Boobis et al., 2006). Carcinogenic risk assessment on the basis of a life span study in a single rodent species in combination with short-term genotoxicity tests and mechanistic information has also been suggested (van Oosterhout et al., 1997). Accordingly, a long-term study in rat supplemented by a short- or medium-term *in vivo* rodent test, such as a model of initiation-promotion, or a carcinogenesis model using neonatal (Flammang et al., 1997; McClain et al., 2001) or transgenic mice, have been proposed as an alternative to life-span bioassays in a two rodent species for human pharmaceuticals (ICH S1B, 1997; see below). Use only the male rat and the female mouse has also been suggested as an alternative to a standard two species, two sexes bioassay (Ashby 1996). Reduced bioassay testing may well be sufficient for the detection of genotoxic carcinogens as these are more often found to give rise to tumours in more than one species, in more than one sex and more than one site, compared to non-genotoxic carcinogens (Ashby and Tennant, 1991; Ashby and Paton, 1993).

Generally, the transgenic mouse models have been reported to reliably predict the carcinogenic potential of compounds and importantly, the number of false positives has been reduced

significantly (RIVM report, 2004). In this case, 3 to 4-fold fewer animals are used in about a three-fold shorter time period. However, it has been shown that the transgenic models were not able of identifying all known human carcinogens when applied as single assays. The most promising models that have been considered are: $p53^{+/-}$, $Xpa^{-/-}$, Tg.AC and rasH2 (RIVM report, 2004). Using a short-term transgenic mouse assay in combination with a rat lifetime bioassay, however, completely eliminated the occurrence of false negatives results; moreover, the overall accuracy of detecting carcinogens and non-carcinogens (85%) increased compared to using only lifetime bioassay (69%). These marked advantages of transgenic mouse models have resulted in the Food and Drug Administration in US and the European Committee for Proprietary Medicinal Products (CPMP) now allowing the use of transgenic mice in the regulatory testing of pharmaceutical compounds as an alternative to a second lifetime bioassay, when a two-year bioassay with rats was carried out. Nonetheless, a more extensive evaluation of transgenic mice using newly developed models and more compounds with different modes of action needs to be carried out in the future (RIVM report, 2004).

Limited data are available to evaluate the results of transgenic rodent assays in known target tissues for carcinogenicity. A case-by-case analysis of instances in which discrepancies are apparent suggests that in the majority of cases, factors such as non-genotoxic mechanism of action, inappropriate mode of administration or inadequate study design may account for the observed negative results in the tissue of interest (Draft Detailed Review Paper on Transgenic Rodent Mutation Assays, OECD, May 2008d).

Many compounds shown to be mutagenic in whole animals have not been tested for carcinogenicity but, of those that have, the vast majority are carcinogenic in laboratory rodents (COC 1991). Results from genotoxicity tests are included in the assessment of putative carcinogenicity and a positive *in vivo* genotoxicity test may be sufficient to identify a chemical as a potential human carcinogen [Kasper et al., 2007; "Guidance on information requirements and chemical safety assessment (chapter R.7a)" developed by ECHA (ECHA 2008b)].

Concerning non-genotoxic carcinogens and based on current understanding, some tests could serve as indicators of some events in the process of carcinogenesis caused by non-mutagenic effects. The *in vitro* cell transformation assays like the "Syrian Hamster Embryo cells (SHE) assay", the "BALB/c 3T3 cell transformation assay" and the test for "Gap Junctional Intercellular Communication (GJIC)" (Maurici et al., 2005), have, among others, been proposed as tests for the risk assessment of chemicals. One proposed application of these tests could be screening for non-mutagenic, tumorigenic effects, which would be carried out simultaneously with mutagenicity tests. Thus, *in vitro* evidence on both mutagenicity and "promotor"-type effects would be available when the need for further testing for carcinogenicity is considered. The SHE assay is the best-developed test of the three mentioned tests so far. At present, a validation of the SHE and BALBc/3T3 assays is ongoing at international level involving Japan, USA and European Commission. The OECD is already drafting a test guideline for the cell transformation assays that will be finalized once the validation studies have been satisfactorily completed. At the EU level, a test guideline for the cell transformation tests already exists (EU B.21).

Conclusions on carcinogenicity:

1- The process of carcinogenicity is recognised as resulting from a sequence of stages and complex biological interactions. The modelling of such complex adverse effects cannot be accomplished at present by the use of non-animal testing. As a general limitation, cancer is a long-term process that is difficult to mimic with relatively short term *in vitro* tests.

2- The use of transgenic animals may offer opportunities for the reduction and refinement of animal tests. However, at present these assays cannot be considered on their own but may be used as an alternative for a lifetime bioassay in a second species, when a two-year bioassay with rats is carried out as has been accepted for the testing of pharmaceutical compounds by some regulatory authorities. The use of transgenic animals should be integrated with available data and considered as part of a weight-of-evidence approach for risk assessment purposes. Still, a more extensive evaluation of transgenic mice models newly developed, including compounds with different modes of action, need to be carried out in the future.

3- Many genotoxic compounds have been shown to be carcinogenic in laboratory rodents. Results from genotoxicity tests are included in the assessment of putative carcinogenicity and a positive *in vivo* genotoxicity test may be sufficient to identify a chemical as a potential human carcinogen.

4- For non-genotoxic carcinogens, *in vitro* cell transformation assays (Syrian Hamster Embryo cells assay and BALB/c 3T3 assay) are currently undergoing validation. The role of these tests in risk assessment has yet to be clarified.

4.3. Ecotoxicity

The safety of substances for the environment, i.e. their impact on the aquatic or terrestrial environment, is important for two EFSA Scientific Panels, FEEDAP and PPR.
Test species are algae, higher plants, invertebrates and vertebrates representing the aquatic and terrestrial environment. This chapter is dealing only with testing on vertebrate animals, fish and birds, and the respective methods/strategies to reduce, refine or replace these tests.
In general, the testing schemes are exposure driven, i.e. if a substance does not enter the environment, testing can be waived.
The testing of *feed additives* for the environment follows such a tiered strategy: in Phase I substances with a significant exposure are identified and only those enter Phase IIA, where acute effects on fish (EU C.1/OECD TG 203) are tested. If considered necessary, Phase IIB testing is conducted which might require testing of long-term effects on fish (e.g. OECD TG 210, EU C. 15/OECD TG 212 or EU C.14/OECD TG 215) and birds (OECD TG 206), and fish bioconcentration (EU C. 13/OECD TG 305).
The testing of *plant protection products* (revision of Council Directive 91/414/EEC is ongoing and will result in a new Regulation) requires tests for the active substances (Annex II) and formulations (Annex III). According to Annex II, two acute fish tests (EU C.1/OECD TG 203) in rainbow trout and a warm water fish species are required. A short-term or chronic fish test must be carried out unless it can be justified that continued or repeated exposure is unlikely to occur. Which of the three possible tests, growth of juvenile fish (EU C.14/OECD TG 215), fish early life stage test (OECD TG 210) or fish full life cycle test (EU C. 15/OECD TG 212) should be carried out depends on expert judgment and should be discussed with the competent authorities. A fish bioconcentration (EU C. 13/OECD TG 305) test might be required for substances with a log coefficient n-octanol/water (log Kow) >3. For formulations, additional testing might be required according to Annex III. If for example conclusions cannot be extrapolated from the active substance to the formulation, an acute test with the formulation is required using the aquatic species (fish, algae or daphnia), which was the most sensitive for the active substance.

For active substances, data on acute oral avian toxicity [as outlined in SETAC/OECD workshop on avian toxicity (OECD 1996) or in the draft OECD TG 223], and short-term dietary toxicity according to OECD TG 205 are required. If birds are exposed to a substance during the breeding period, possible effects on reproduction are tested according to OECD TG 206. Additional testing might be required for particular formulations, e.g. acute oral toxicity of the formulation, food avoidance testing, cage or field trials, secondary poisoning, habitation and feeding behaviour in treated areas.
In the light of the revision of Directive 91/414/EEC, the EFSA PPR Panel issued an Opinion (EFSA, 2007b) on ecotoxicity assessment. Some of the recommendations deal with animal welfare aspects and are included in the following. Further details of these proposals related to birds and mammals are described in an Opinion regarding risk assessment for birds and mammals (EFSA, 2008c).

4.3.1. Fish toxicity

The most promising alternative methods are dealing with the endpoints acute fish toxicity and fish bioconcentration. In the following, only methods which have been accepted by regulators or have entered the validation process are described in detail.

4.3.1.1. Short-term fish toxicity

With the currently used *in vivo* test according to EU C.1/OECD TG 203, the concentration which kills 50% of fish (LC50) exposed to a substance for 96 hours is determined. For this purpose, 7-10 fish/concentration are exposed to at least five concentrations of the test substance. The following fish species can be used: zebrafish (*Brachydanio rerio*), fathead minnow, (*Pimephales promelas*), common carp (*Cyprinus carpio*), Japanese medaka (*Oryzias latipes*), guppy (*Poecilia reticulata*), bluegill (*Lepomis macrochirus*), and rainbow trout (*Oncorhynchus mykiss*).
The OECD TG 203 already gives the possibility to significantly reduce the number of fish for acute toxicity testing by allowing testing with the "limit test" at the given concentration of 100 mg/l (active ingredient). Seven to ten fish are exposed to 100mg/l of the test substance and if no mortality occurs the LC50 is reported as LC50 > 100mg/l, which indicates that there is no acute fish toxicity.

Animal welfare concerns: in contrast to mammalian acute toxicity testing, short-term (acute) fish toxicity testing is still based on measuring the lethal concentration (LC50) to 50% of the fish. Moreover, the group sizes of test and control fish are rather large with 7-10 animals. Fish have a nociceptive system and have homologous and analogous structures to mammals, however their ability to suffer any pain and distress due to testing protocols is not well documented. Moreover, the developmental stage at which fish may be able to experience such adverse effects has hardly been studied at all. Nevertheless, they are equally protected under the law, and deserve equal consideration until it is proven they are not able to suffer in the same way as mammals.

Recent and future developments:
The approach promoted in the guidance documents developed by ECHA to fulfil the requirements of the REACH legislation (Guidance on information requirements and chemical safety assessment – Chapter R 7.b and R 7c) and described in the following might be also

applicable for feed additives and plant protection products. In the light of reducing *in vivo* testing, all available information should be evaluated to determine the need for acute fish toxicity testing. Significant information might be gained from experimental data from standard, non-standard studies or from non-standard species, data generated with QSARs, read-across, etc. For example, since for the risk assessment of chemicals only the data from the most sensitive species of the three trophic levels in the aquatic environment, i.e. algae for plants, daphnia for invertebrates and fish for vertebrates are used, it might be possible to waive the fish test if there is compelling evidence to suggest that the fish is likely to be at least a factor of about 10 less sensitive than invertebrates or algae.
If acute fish data are needed, the following methods already used for safety testing of chemicals could be considered, where methods "a" and "b" (see below) are already included in tiered testing strategies and whereas method "c" is in the process of being validated at OECD level. Promising research work combing fish cell culture methods and fish embryo tests is ongoing in the CellSense[28] project funded by CEFIC LRI (European Chemicals Industries Council Long-range Research Initiative) but not further described here.

a) In silico methods

For a long time *in silico* methods or computer-based models such as QSARs have been used for evaluating the toxic potential of substances to fish. The available tools are summarised in the "Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals" developed by ECHA (ECHA, 2008a). Both, the OECD and the European Commission Joint Research Centre give access to QSAR tools on their websites[29].
A specific QSAR tool predicting effects of plant protection products on rainbow trout (96h short-term fish toxicity) has been developed within the framework of the FP6 funded project "Development of environmental modules for evaluation of toxicity of pesticide residues in agriculture (DEMETRA)"[30].

b) Threshold approach

The threshold approach is a tiered strategy which can significantly reduce the number of fish to be used in acute fish toxicity testing. It takes into consideration that often only the lowest value of the acute toxicity in species of three trophic levels (daphnia, algae, fish), i.e. the most sensitive test species, is considered for regulatory purposes and for about 80% of the substances, fish is not the most sensitive test species (Weyers et al., 2000, Hutchinson et al., 2003, Jeram et al., 2005).
The threshold approach was originally described as threshold/step-down approach by Hutchinson et al. (2003) for pharmaceuticals, then applied to chemicals and other substances (Jeram et al., 2005; Hoekzema et al., 2006). It was further developed at the European Commission's Joint Research Centre taking into account existing guidelines and reflecting the requirements for the limit test of OECD TG 203. The ECVAM Scientific Advisory Committee has endorsed in 2006 the scientific validity of the threshold approach for chemicals following the advice of its peer review Panel.
As Hoeger et al. (2006) showed, the threshold approach could also be useful for plant protection products (data provided by the European Centre for Ecotoxicology and Toxicology of Chemicals - ECETOC or retrieved from DG SANCO databases on new and existing plant protection products).

[28] http://www.cefic-lri.org/; CEFIC Long-range Research Initiative - ECO8: Development of a strategy to predict acute fish lethality using fish cell lines and fish embryos
[29] www.oecd.org/env/existingchemicals/qsar and http://ecb.jrc.it/qsar/.
[30] The software and detailed description is available on http://www.demetra-tox.net/.

In the threshold approach, the lowest EC50 obtained from algae (e.g. EU C.3/OECD TG 201) and Daphnia (e.g. EU C.2/OECD TG 202) tests is used as Threshold Concentration, at which a fish limit test according to EU C.1/OECD TG 203 is carried out. In the case that no mortality is observed, no further testing is needed and the acute fish toxicity result is reported as LC50 > Threshold Concentration. In the case that mortality is observed, a full LC50 test according to OECD TG 203 should be performed.
The threshold approach is part of the testing strategy proposed in the guidance document for REACH implementation (ECHA, 2008a) and currently being discussed at OECD level[31].

c) Fish embryo tests
The use of zebrafish (*Danio rerio*) embryos for testing of chemical substances and waste water has been described by Schulte and Nagel (1994), Nagel (2002), Braunbeck et al. (2005) and their suitability to replace acute fish toxicity tests in the regulatory context is under discussion. Germany has replaced the fish test for routine whole effluent testing with a zebrafish embryo test, for which DIN and ISO guidelines are available (DIN, 2001; ISO, 2007).
In the fish embryo test, newly fertilised zebrafish eggs are exposed to chemicals for up to 48 hours or if considered necessary also post hatch. Four endpoints are recorded as indicators of acute lethal toxicity: coagulation of fertilised eggs, lack of somite formation, lack of detachment of the tail bud from the yolk and lack of heart beat. Embryos are considered dead, if one of these endpoints is recorded as positive.
Under the OECD umbrella, a collaborative study started in January 2009 to evaluate the within- and between-laboratory reproducibility of the zebrafish embryo test taking into account various exposure scenarios.

Conclusions on short-term fish toxicity testing:
1- The EU C.1/OECD TG 203 already gives the possibility to significantly reduce the number of fish for acute toxicity testing by allowing testing with the "limit test" at the given concentration of 100 mg/l (active ingredient). Whenever possible, e.g. for substances unlikely to be toxic to fish, the limit test should be applied since only 14-20 fish are used compared to 42-60 in the full LC50 test.
2- The various alternative methods (e.g. QSARs, the threshold approach, the fish embryo test) hold the potential of reducing the number of fish used for acute aquatic toxicity testing. It should be evaluated whether they could be used for the safety assessment of feed additives and plant protection products.

4.3.1.2. Long-term and chronic fish toxicity

At present, it depends on expert judgment which of the three possible tests should be carried out: growth of juvenile fish [EU C.14/OECD TG 215; recommended species: rainbow trout (*Oncorhynchus mykiss*), zebrafish (*Brachydanio rerio),* Japanese medaka (*Oryzias latipes*)], fish early life stage test [OECD TG 210; recommended species: see above and fathead minnow (*Pimephales promelas*)] or fish full life cycle test [EU C.15/OECD TG 212; recommended species: see above and common carp (*Cyprinus carpio*)]. However, if effects are evident in the fish early life stage test, further testing will be triggered.
The EFSA PPR Panel recommends (EFSA 2007b) performing a fish full life cycle test rather than separate tests on different life stages since exposure in early life stages may trigger effects

[31] Rolling work plan for the OECD test guidelines program 2006/2008 - Project 2.23: New Guidance Document on Application of the Threshold Approach as a Limit Test for Acute Fish Toxicity Testing; Lead European Commission.

only in later phases which would not be detected in an early life stage test (Nagel and Isberner, 1998).

Recent and future developments:

Despite the fact that long-term and chronic fish toxicity testing uses most of the fish in environmental hazard assessment, alternative methods are not very advanced yet, e.g. the use of fish embryos or "omics" based approaches. This might be due to the complexity of the endpoint.

Conclusions on long-term and chronic fish toxicity testing:
1- To date, alternative methods for long-term and chronic fish toxicity testing are at an early stage of development and not ready for regulatory use.
2- Due to the high numbers of animals used, projects on alternative methods/strategies for long-term and chronic fish toxicity testing should be initiated/supported.

4.3.1.3. Bioconcentration and bioaccumulation in fish

Information on the bioaccumulative potential of a substance in aquatic organisms is vital for understanding its environmental behaviour. Bioconcentration is defined as the net result of the uptake, distribution and elimination of a substance in an organism due to waterborne exposure, whereas bioaccumulation includes all routes, i.e. air, water, soil and food. The bioconcentration factor (BCF) for fish can be determined by using computional tools or by testing in fish. At least 108 fish are used for the fish bioconcentration test according to EU C.13/OECD TG 305 (recommended species: (*Brachydanio rerio*), fathead minnow, (*Pimephales promelas*), common carp (*Cyprinus carpio*), Japanese medaka (*Oryzias latipes*), guppy (*Poecilia reticulata*), bluegill (*Lepomis macrochirus*), rainbow trout (*Oncorhynchus mykiss*) and Three-spinned stickleback (*Gasterosteus aculeatus*).
Generally, it is assumed that hydrophilic substances with a partition coefficient n-octanol/water (log Kow) < 3 and/or low potential to cross biological membranes do not bioaccumulate and tests in fish are not required. In this context, the EFSA PPR Panel (EFSA 2007) proposes waiving of the fish bioconcentration test for substances which rapidly degrade in the environment.

Recent and future developments:
Various tiered testing strategies have been developed to determine the bioaccumulative potential of a substance (ECETOC, 2005; de Wolf et al., 2007). These combine read-across, chemical grouping, *in silico* models (Tier 1), *in vitro* methods to determine adsorption, distribution, metabolism and excretion parameters (Tier 2), reduced fish test (Tier 3) and full OECD fish bioconcentration test (Tier 4).
They are reflected in the "Guidance on information requirements and chemical safety assessment" developed by ECHA (ECHA 2008d) and might also be useful for other substances/products.

a) Read-across, chemical grouping, expert systems (Tier 1)
Existing information might be retrieved from databases listed by Weisbrod et al. (2006) and included in the "Guidance on information requirements and chemical safety assessment. Chapter R.7c" developed by ECHA (ECHA, 2008d).

b) In silico models (Tier 2)

Existing models have recently been reviewed by Pavan et al. (2006) and in the "Guidance on information requirements and chemical safety assessment. Chapter R.7c" developed by ECHA (ECHA, 2008d). The most common and simplest QSAR models are based on correlations between BCF and the chemical hydrophobicity (log Kow) of substances; others use solubility as parameter or molecular descriptors. More complex *in silico* models include correction factors to account for biotransformation (e.g. BCFWIN Software developed by Meylan et al., 1999). The recently developed software (BCFBAF) for predicting BCF and bioaccumulation factors (BAF) is now available on the EPI-suite website[32] and combines various QSAR models. Before using *in silico* models, it should be carefully evaluated whether the substance is covered by the applicability domain for which the model has been developed.

c) In vitro methods (Tier 3)

In order to improve the existing BCF models, it is important to consider the uptake, distribution, metabolism and excretion of a substance, e.g. substances which are metabolised most likely do not accumulate in the organism. For example, two projects funded by ECVAM-IHCP-JRC and CEFIC are evaluating the potential of *in vitro* methods (*in vitro* trout S9 assay) to predict the *in vivo* metabolism. By improving the modelling scenarios, less *in vivo* bioconcentration tests would need to be performed.

A list of promising methods is given in ECHA (2008d); however, none of them is yet validated, nevertheless they can provide important information on the metabolic stability of compounds and might be valuable in testing strategies.

d) Reduction/refinement of the fish concentration test (Tier 4)

Other *in vivo* approaches, which reduce the number of fish needed and test duration, are under discussion/evaluation. The most promising are the fish dietary bioaccumulation test and modifications of the EU C.13/OECD TG 305 regarding the exposure scenario, number of samples collected, test duration and number of exposure (ECHA, 2008c).

Conclusions on fish bioconcentration and bioaccumulation:

1- The current requirements for safety assessment of feed additives and plant protection products allows waiving of the fish bioconcentration test for substances with a coefficient n-octanol/water (log Kow) < 3 and/or a low potential to cross biological membranes or for substances which rapidly degrade in the environment.

2- Tiered strategies for fish bioconcentration testing are available for industrial chemicals and might be applicable for other substances as well, e.g. before *in vivo* tests are performed existing information (chemical properties, read-across, bioconcentration databases, metabolic stability etc), valid *in silico* models and *in vitro* data might be used and might allow conclusions on the bioaccumulative potential of a substance.

4.3.2. Avian toxicity

Avian toxicity testing is mainly required for plant protection products. In contrast to fish or mammalian toxicity testing, *in vitro* methods for avian toxicity testing are not available so far.

[32] http://www.epa.gov/oppt/exposure/pubs/episuite.htm (United States Environmental Protection Agency)

However, in the Opinion of the EFSA PPR Panel (EFSA, 2007; 2008), there are several options to reduce the number of birds tested or avoid unnecessary testing as described below.

4.3.2.1. Acute avian oral toxicity

Acute avian oral toxicity studies (OECD 1996) determine the LD50 dose and use at least 50 birds for each testing [recommended species are a quail species (Japanese quail, *Coturnix coturnix japonica* or bobwhite quail, *Colinus virginianus*) or mallard duck (*Anas platyrhynchos*)].

Animal welfare concerns: since lethality is required, the degree of adverse effects is likely to increase. There is considerable scientific literature on the ability of birds to suffer pain and distress and so they must be given equal consideration under the law until it is proven they are not able to suffer in the same way as mammals.

Recent and future developments:
A new guideline for acute avian toxicity is under development at OECD (draft OECD TG 223), which would require 12-24 animals. Further, the EFSA PPR Panel (EFSA 2007b) states that additional useful information should be gathered during the oral toxicity studies, e.g. measurement of food consumption on the day of dosing, and the approximate time of onset and disappearance of overt clinical symptoms. As reported by that Panel, such information could be used for a refined risk assessment when food avoidance responses and metabolism of the pesticide occur.
The DEMETRA tool[33] (see 4.3.1.1) includes models for predicting the acute oral toxicity and dietary toxicity to bobwhite quail with the help of QSARs.

Conclusions on acute avian oral toxicity:
1- At present there are no alternative methods available. However, a new guideline (draft OECD TG 223) is under development at OECD which significantly reduces the number of animals used. It should be incorporated into guidance documents as soon as finalised.

4.3.2.2. Short-term avian dietary toxicity test

The short-term dietary toxicity test (OECD TG 205) should provide information on short-term dietary toxicity (LC50 values, lowest lethal concentration), where possible no observed effect concentrations (NOEC), time courses of response and recovery and include relevant gross pathological findings. LC50 and NOEC values are converted to daily dietary dose (LDD50) and no observed effect daily dose (NOEDD). Recommended species are the mallard duck (*Anas platyrhynchos*), bobwhite quail (*Colinus virginianus*), pigeon (*Columba livia*), Japanese quail (*Coturnix coturnix japonica*), ring-necked pheasant (*Phasianus colchicus*) and redlegged partridge (*Alectoris rufa*). The test is carried out on a second species, when the oral toxicity or short-term dietary toxicity is below 500 mg/bodyweight or mg/kg food respectively.

Recent and future developments:
The EFSA PPR Panel describes in its Opinions (EFSA, 2007b; EFSA, 2008c), the scientific limitations (test design and purpose, results difficult to interpret) and associated animal welfare concerns (lethal test on 50 birds) of the short-term dietary test and recommends restricting the use of this test as far as possible. It should only be performed when scientifically justifiable,

[33]http://www.demetra-tox.net/

e.g. “only for those pesticides where the mode of action and/or results from mammalian studies indicate a potential for the LDD50 measured by the short term study to be lower than the LD50 based on an acute oral study. This would apply, for instance, to many of the organochlorine compounds and anticoagulants”. In addition, it should be performed only in one species.

Conclusions on the short-term avian dietary toxicity test:
1- At present, there are no alternatives available.
2- The advice of the EFSA PPR Panel should be followed and the short-term dietary toxicity test (OECD TG 205) should not be performed on a routine basis but only when scientifically justifiable and only in one species.

5. Integrated testing and risk assessment strategies

Chapter 4 described the current situation with respect to replacement, reduction and refinement of testing methods for specific areas of mammalian toxicology and ecotoxicology. This chapter addresses the wider issue of strategies that can contribute to achieving better testing and risk assessment approaches that incorporate the Three Rs. The purpose of integrated testing and risk assessment strategies is to ensure that the most appropriate information is generated to adequately assess any risks and to limit animal use to the minimum necessary to make that assessment. Elements that can be used in such strategies are tiered approaches to testing, computer models for exploring (quantitative) structure-activity relationships, use of physicochemical data to predict properties such as bioaccumulation, grouping or read-across from toxicity data on structurally related substances, qualified presumption of safety for certain microorganisms, and the use of human data.

5.1. Tiered testing approaches

Tiered approaches to toxicity testing are based either on the extent of exposure to a substance and/or on the nature of the chemical structure of the substance. In exposure-driven approaches, the amount of toxicity data needed is related to the extent of known or estimated human exposure to the substance, more data being required as exposure increases. When exposure is very low, it may even allow an assessment of the safety-in-use of a substance to be completed with little or no toxicity data, as described below. In chemical structure-driven approaches, knowledge of the general toxicity of the class of substance being considered influences the amount of toxicity data required. Examples of such approaches used by EFSA Scientific Panels include the reduced requirements for assessment of enzymes in animal feed because their protein structure indicates the likelihood of low toxicity, nutritional substances used as additives in food or feed, substances such as certain additives that are known to be normal metabolites in the body or which break down in the body to known endogenous substances, and substances such as polymeric additives in food contact materials that are very unlikely to be absorbed because of their high molecular weight. Tiered approaches for assessing the bioaccumulative potential of substances in aquatic organisms (e.g. bioconcentration factor) have already been described in 4.3.1.3.

5.1.1. Tiered testing for food contact materials

An example of the use of a tiered approach that is based on exposure considerations is the assessment by EFSA of substances used in food contact materials. In this scheme (SCF, 2001), toxicity data requirements are grouped into three tiers based on an assessment of migration of the test substance out of the contact material and into the food (and hence potential exposure). The approach starts with a requirement for only *in vitro* genotoxicity tests at the lowest exposures, scaling up at intermediate exposures by addition of a requirement for a 90-day oral toxicity test and data to demonstrate the absence of potential for accumulation in man (for which an octanol-water partition coefficient of <3 may suffice), and only for substances with the highest exposures is a full data set required (90-day oral study in a second species, studies on absorption, distribution, metabolism and excretion, reproduction and developmental toxicity studies and long-term toxicity/carcinogenicity studies). In practice, very few substances used in food contact materials migrate into food in quantities that require a full data set; most require only *in vitro* genotoxicity tests.

5.1.2. The Threshold of Toxicological concern (TTC) approach

An example of a tiered testing approach based on both exposure and chemical structure considerations is the Threshold of Toxicological Concern (TTC) approach (Kroes et al., 2000, 2004). The TTC approach is applicable for substances where intakes are anticipated to be low. This approach is used by EFSA for the assessment of flavouring substances and for plant protection product metabolites in ground water. It allows a toxicological evaluation, even in the absence of experimental data and thus reduces the requirement for experimental animal studies. The EFSA Scientific Committee has recently adopted a mandate on "Exploring options for providing preliminary advice about possible human health risks based on the concepts of the Threshold of Toxicological Concern". A working group has been established that will address the mandate and the Committee anticipates publishing an Opinion during 2010. The PPR Panel has also launched an art. 36[34] call on "Applicability of threshold of toxicological concern in the dietary risk assessment of metabolites, degradation and reaction products of active substances of plant protection products". The contractor will complete the work by the end of 2009.

The TTC concept is based on the principle that generic human exposure threshold values for (groups of) chemicals can be established, below which there would be no appreciable risk to human health. The concept proposes that such a value can be identified for many chemicals, including those of unknown toxicity when considering their chemical structures. Under the TTC approach, chemicals are assigned (on the basis of accumulated knowledge about chemical structure and the likelihood of toxicity), to one of 3 structural classes – Class 1, presumed low toxicity, Class 2, structures less innocuous than Class I but not containing features suggestive of toxicity, or Class 3**,** no strong initial presumption of safety, potential for significant toxicity or the presence of reactive groups. The European Commission Joint Research Centre has published a software application (Toxtree[35]) that includes a facility for allocating chemicals to the three structural classes (Cramer classes). Human exposure threshold values have been established for each of the 3 structural classes by considering extensive databases of toxicity data generated on a wide range of chemicals in the past (Munro et al., 1996, 1999; Cheeseman et al., 1999). Estimated intakes of individual flavouring substances are compared with the relevant human exposure threshold value to indicate whether further data, including toxicological data may be required.

In its safety assessment of food flavouring substances, EFSA's former Scientific Panel for Food Additives, Flavourings and Food Contact Material (AFC) in work now continued by the Panel for Food Contact Materials, Enzymes and Flavourings (CEF) has placed emphasis on the predicted metabolism, based on available toxicokinetic studies and theoretical considerations, and where it is predicted that the substance would be metabolised to innocuous substances, it is possible to reach a decision that a flavouring substance or group of flavourings are safe for human consumption (see for example, Annex 1 in EFSA, 2004). Emphasis is put on the evaluation of any available genotoxicity data including structural considerations, since genotoxic substances may pose a risk to health, even at low intakes, and substances have been assessed by EFSA as being of concern, following this step in the process. In consideration of both the metabolism and genotoxicity data, individual flavouring substances are assessed by reference to closely structurally-related substances where data may be available, even if there are no data on the particular substance under consideration.

[34] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:031:0001:0024:EN:PDF
[35] http://ecb.jrc.it/documents/QSAR/QSAR_TOOLS/Toxtree_user_manual.pdf

In cases where the intake of a flavouring substance is estimated to be below the relevant human exposure threshold value and consideration of the metabolic and genotoxicity profiles do not give rise to concern, further data are not required. For flavouring substances where the intake is above the relevant threshold value, use is made of the No Observed Adverse Effect Level (NOAEL) established in an available repeat dose toxicity study, carried out either on the flavouring substance under evaluation or on a closely related substance, to establish a margin of safety. If the margin of safety is considered sufficient it can be concluded that the substance is not of safety concern. Only when the toxicological data are judged to be inadequate is the substance (or group of substances) judged to be a candidate for further work, including possible animal testing. This approach has minimised the requirements for animal testing on these flavouring substances. However, when undertaking these evaluations, EFSA has in several cases identified data gaps that require additional information on toxicity and testing *in vitro* and *in vivo,* in order to confirm that the substance or substances are safe when used as flavourings.

5.1.3. Structure-activity relationships (SARs) and Quantitative structure-activity relationships (QSARs), grouping and read- across

SAR and QSAR approaches are methods for estimating the properties of a chemical from its molecular structure. The properties described may be physicochemical attributes, environmental fate or a specific effect on human health or an environmental species. They are potentially useful at the outset of designing studies to investigate the safety of a substance and in weight-of-evidence approaches to risk assessment. SARs and QSARs have been developed for the prediction of mammalian and ecotoxicity endpoints (see chapter 4).

OECD has developed guidance material and a 'Toolbox' for application of QSARs. The QSAR Application Toolbox was released in early 2008 and recently updated (OECD, 2008c). It comprises a software system for systematically grouping chemicals according to the presence or potency of a particular effect, allowing read-across or estimation of missing values for members of the group that have not been tested. The US Environmental Protection Agency (EPA) has also developed SAR approaches for the grouping of industrial chemicals in the High Volume Production programme (EPA, 2008). The European Chemicals Agency has also published extensive guidance on the use of QSARs for evaluation of industrial chemicals under the REACH legislation (ECHA, 2008a).

In mammalian toxicity, the use of SARS and QSARs has so far largely been restricted to prediction of pathways of metabolism, or of toxicity endpoints where there is a single mode of action, such as genotoxicity caused by direct interaction with DNA or chromosomes (Benigni et al., 2007). SARs for prediction of skin irritation and skin corrosion have also been developed (Gallegos Saliner et al., 2006; Rorije et al., 2005). However, development of SARs for other more complex endpoints, such as those revealed in repeat-dose toxicity, reproductive or developmental toxicity tests has been slower and those that exist have low predictivity, or are only predictive for a particular type of effect within the large number of endpoints or modes of action that can be investigated in *in vivo* studies.

Grouping of substances that are closely structurally related (e.g. an homologous series of alcohols or fatty acids) and read-across from a substance or substances within the group that have an extensive toxicity profile to those lacking data can be used to avoid the need to test all members in the group.

Read-across is used by EFSA Scientific Panels to evaluate substances that are structurally-related or have the same mode of action, such as certain food or feed additives and certain substances used in food contact materials.

5.1.4. Qualified Presumption of Safety (QPS)

The concept of the Qualified Presumption of Safety has been developed by EFSA to provide a generic assessment system for use within EFSA that could be applied to all requests received for the safety assessments of microorganisms deliberately introduced into the food chain.
A wide variety of microbial species are used in food and feed production. Some have a long history of apparent safe use, while others are less well understood and their use may represent a risk for consumers. Experience has shown that there was a need for a tool for setting priorities within the risk assessment of those microorganisms used in food/feed production referred to EFSA and consequently the subject of a formal assessment of safety. To meet this need a system was proposed for a pre-market safety assessment of selected groups of microorganisms leading to a "Qualified Presumption of Safety (QPS)". In essence, this proposed that a safety assessment of a defined taxonomic group (e.g. genus or group of related species) could be made based on four pillars (establishing identity, body of knowledge, possible pathogenicity and end use). If the taxonomic group did not raise safety concerns or, if safety concerns existed, but could be defined and excluded (the qualification) the grouping could be granted QPS status. Thereafter, any strain of microorganism the identity of which could be unambiguously established and assigned to a QPS group would be freed from the need for further testing and safety assessment other than satisfying any qualifications specified. Microorganisms not considered suitable for QPS would remain subject to a full safety assessment.
EFSA adopted an Opinion on the introduction of the Qualified Presumption of Safety approach for assessment of selected microorganisms in 2007 (EFSA 2007b).

The EFSA Biological Hazards Panel has recently reviewed the list of QPS microorganisms and updated the anti-microbial resistance criteria used to judge the safety of food/feed use microorganisms (EFSA, 2008b).

The implementation of the QPS approach will lead to a reduction in the need for animal testing for the safety assessment of microorganisms.

Conclusions on integrated testing strategies:

1- Adoption of integrated testing and risk assessment strategies can reduce the need for animal testing and refine *in vivo* testing when it is required
2- The use of strategies such as tiered testing approaches for low-exposure chemicals, the use of SARs and QSARs and QPS can eliminate or reduce animal testing.

6. General conclusions and recommendations

EFSA and its Scientific Committee are committed to a proactive animal welfare approach based on sound scientific principles and the need to ensure that adequate data are available for reliable risk assessment. EFSA's particular interest focuses on following closely the developments in replacement, reduction and refinement of animal testing methods used for the risk assessment of chemical substances and microbiological agents that are used in food and feed. This includes not only the developments in replacement of animal testing with *in vitro*, *in*

silico or other alternative methods, but also, where animals need to be used, refinement of *in vivo* methods to avoid or minimise any pain and distress that may be caused. Reduction in the number of animals used is also important.

To that end, the present Opinion has reviewed the current possibilities for replacement, reduction and refinement of animal testing covering the various toxicological and ecotoxicological endpoints on which EFSA Scientific Panels receive data for the purpose of hazard and risk assessment. Detailed conclusions and recommendations concerning the individual areas of testing can be found at the end of each section in the preceding text. In some areas, substantial advances already have been made and validated alternative methods are available. Improved methods, based on the Three Rs, with internationally validated and accepted protocols, are now available in the following areas of **mammalian toxicology**:

- acute toxicity testing,
- skin corrosion and skin irritation testing, skin sensitisation testing, and
- tests for severe eye irritants.

Improved methods are now available in the following areas of **ecotoxicology**:

- short-term fish toxicity, and
- bioconcentration and bioaccumulation in fish.

In addition, the EFSA PPR Panel has made recommendations for refinements and reduction in the approaches for avian toxicity tests.

The Committee is of the opinion that the available, validated methods should be used wherever possible and use of outdated methods that do not conform to best practice with respect to the Three Rs should be strongly discouraged. It is **recommended** that EFSA's Scientific Panels and the Scientific Committee include the use of the above-mentioned validated methods, in the elaboration of any future guidelines for testing and submission of dossiers on substances to be evaluated by EFSA. However, it should be noted that EFSA may not have much opportunity, other than through guidelines, to influence decision-making on the detailed content of safety testing programmes that are conducted by applicants prior to submitting a dossier to EFSA. It should also be noted that EFSA will continue to consider the results of animal tests conducted to earlier protocols, both in the interests of minimising any further animal testing and recognising that some tests will have been conducted earlier, perhaps several years before submission of a dossier.

In other areas of toxicology, particularly those involving complex endpoints, such as those that are investigated in repeated dose toxicity, reproduction and developmental toxicity studies, the development of alternative methods is more difficult. The same applies for ecotoxicology endpoints such as acute and chronic toxicity in fish and birds and bioconcentration. A number of methods, based on the Three Rs, have been proposed and further developed but, at present they cannot yet provide the wealth of information that can be derived from currently used *in vivo* methods. In these areas, the implementation of integrated testing and risk assessment strategies can contribute to reductions in the number of animals studies needed and, by the choice of test selected, may result in use of fewer animals. It is **recommended** that applicants

that submit dossiers to EFSA and EFSA Scientific Panels utilise such strategies wherever possible. When human data are available, as is the case for some environmental contaminants in food, such as dioxins or heavy metals, they may be used for risk assessment, possibly avoiding the need for further animal testing.

Recommendations for additional testing in laboratory animals to fill gaps in the animal toxicological databases should not be made unless there is clear potential for improving protection of human health, and until other approaches to obtaining relevant information have been fully explored. These would include assessment of knowledge on physicochemical properties of the substance being evaluated and its metabolism, use of structure-activity relationships, read-across from data on similar chemicals, results from *in vitro*, *in silico* and other *in vivo* studies, including observations or studies on humans, where available, and consideration of the extent of exposure.

Genotoxicity is one area of toxicology testing where *in vitro* approaches often obviate the need for any *in vivo* testing. Where the necessity for *in vivo* testing has to be considered, assessment strategies can assist in reducing that need to the minimum. Since assessment of the potential for genotoxicity is common to many of the chemical substances that EFSA evaluates, this would be a fruitful topic for EFSA to follow in the future, as ongoing research continues to refine the predictive ability of *in vitro* tests and develop strategies for follow-up of substances showing genotoxic potential *in vitro*.

It is desirable that the various areas of EFSA's activities adopt a common approach to replacement, reduction and refinement of animal testing, where such testing is needed and is common across a number of areas. The EFSA Scientific Committee notes that legislation in force since 1986 requires that when validated and accepted alternative methods are available they should be used in line with the principle of the Three Rs. This should be communicated to applicants submitting dossiers to EFSA as well as reflected in any guidelines developed by EFSA. However, it is recognised that the various guidelines already established covering the different areas of EFSA's remit have evolved over time and contain some justifiable differences in approach and data requirements. Thus complete harmonisation of all guidelines and their data requirements across EFSA's remit is not foreseen. It is also noted that some guidelines are now incorporated into EU legislation, while others remain as EFSA guidance. It is clear that revision of guidelines can be a time-consuming process. For this reason, it is **recommended** that a dialogue should be initiated between EFSA and the European Commission on best ways to address inclusion into guidelines of new, validated testing methods based on the Three Rs principle that would avoid undue delay.

Recognising that the Commission and Member States have responsibility for agreement of new testing methods, there needs to be good communication between EFSA and the corresponding Commission Services that lead on this area, so that EFSA can be informed of latest developments on the validation and acceptance of new testing methods. Communication on implementing the Three Rs with other EU Agencies dealing with chemical risk assessment is equally important.

EFSA, in its regular dialogue with DG Research should also ensure that where alternative methods need to be developed, such research is flagged up for inclusion in future framework programmes.

The Scientific Committee recommends that EFSA follows up this Opinion with a review of progress in the field of alternatives in 3 years time.

REFERENCES

Ashby J. and Tennant R.W. 1991. Definitive realtionshiips among chemical structure, carcionogencity and mutagenicity for 301 chemicals tested by the US NTP. Mutat Res. 257: 229-306. Erratum in: Mutat Res 1994 317:175.

Ashby J. and Paton D. 1993. The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. Mutat Res. 286:3-74.

Ashby J. 1996. Alternatives to the 2-species bioassay for the identification of potential human carcinogens. Human and Experimental Toxicology, 196, 15: 183-202.

Balls M., Blaauboer B.J., Brusick D., Frazier J., Lamb D., Pemberton M., Reinhardt C., Roberfroid M., Rosenkranz H., Schmid B., Spielmann H., Stammati A.L. and Walum E. 1990. Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity testing procedures. ATLA 18, 313-336.

Balls M., Blaauboer B.J., Fentem J., Bruner L., Combes R.D., Ekwall B., Fielder R.J., Guillouzo A., Lewis R.W., Lovell D.P., Reinhardt C.A. , Repetto G., Sladowski D., Spielmann H. and Zucco F. 1995. Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM workshop 5. ATLA 23, 129-147.

Balls M., Botham P.A., Bruner L.H. & Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. Toxicology *in Vitro* 9, 871–929.

Balls M., Berg N., Bruner L.H., Curren R., de Silva O., Earl L.K., Esdaile D.J., Fentem J.H., Liebsch M., Ohno Y., Prinsen M.K., Spielmann H. & Worth, A.P. 1999. Eye irritation testing: the way forward. The report and recommendations of ECVAM workshop 34. ATLA 27, 53–77.

Benigni R., Netzeva T.I., Benfenati E.., Bossa C., Franke R., Helma C., Hulzebos E., Marchant C., Richard A., Woo Y.T., Yang C. 2007. The expanding role of predictive toxicology: an update on the (Q)Sar models for mutagens and carcinogens. J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev., 25, 53-97.

Blaaubour B.J. 2001. Toxicodynamic modelling and the interpretation of *in vitro* toxicity data. Toxicol letter 31, 120:111-123.

Blaaubour B.J. 2002. The necessity of biokinetic information in the interpretation of *in vitro* toxicity data. ATLA 30 suppl 2: 85-91.

Braunbeck T., Boettcher M., Hollert H., Kosmehl T., Lammer E., Leist E., Rudolf M., Seitz N. 2005. Towards an alternative for the acute fish LC(50) test in chemical assessment: the fish embryo toxicity test goes multi-species -an update. ALTEX, 22:87-102.

Bremer S., Cortvrindt R., Daston G., Eletti B., Mantovani A., Maranghi F., Pelkonen O., Ruhdel I., Spielmann H. 2005. Reproductive and Developmental Toxicity. ATLA, 33 (Suppl. 1):183–209

Burg B. 2006. Comparison of *in vitro* and *in vivo* screening models for androgenic and estrogenic activities. Toxicol Sci. Jan; 89(1):173-87.

Boobis A.R., Cohen S.M., Dellarco V., McGregor D., Meek M.E., Vickers C., Willcocks D., Farland W. 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit Rev Toxicol 36(10): 781-792.

Chamberlain M., Gad S.C., Gautheron P., Prinsen M.K. 1997. IRAG (Interagency Regulatory Alternatives Group) working group 1. Organotypic models for the assessment/prediction of ocular irritation. Food and Chemical Toxicology 35, 23–37.

Cheeseman M.A., Machuga E.J. and Bailey, A.B. 1999. A tiered approach to threshold of regulation. Food and Chemical Toxicology, 37, 387-412.

Christian M.S. & Diener R.M. 1996. Soaps and detergents — alternatives to animal eye irritation Chapter 3.3: eye irritation 73 tests. Journal of the American College of Toxicology 15, 1–44.

COC - Committee on Carcinogenicity of Chemicals in Food, Consumer Products and Environment. 1991. Report on Health and Social Subjects 42, Guidelines for the Evaluation of Chemicals for Carcinogenicity. London, HMSO.

Cooper R.L., Lamb J.C., Barlow S.M., Bentley K., Brady A.M., Doerrer N.G., Eisenbrandt D.L., Fenner-Crisp P.A., Hines R.N., Irvine L.F.H., Kimmel C.A., Koeter H., Li A.A., Makris S.L., Sheets L.P., Speijers G.J.A. and Whitby K.E. 2006. A tiered approach to life stages testing for agricultural chemical safety assessment. Crit Rev Toxicol 36, 69-98.

Corvi R., Albertini S., Hartung T., Hoffmann S., Maurici D., Pfuhler S., van Benthem J., Vanparys P. 2008. "ECVAM retrospective validation of *in vitro* micronucleus test (MNT)". Mutagenesis 23(4): 271-283.

Council Directive 86/609/EEC of 24 November 1986 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. Official Journal L 358, published on 18.12.1986, 1-28.

Council Directive 91/414/EEC of 15 July 1991 concerning the placing of plant protection products on the market. Official Journal L230, published 19.08.1991.

Council Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Official Journal of the European Union L142, published on 31.5.2008.

Curren R.D., Southee J.A., Spielmann H., Liesch M., Fentem J.H. and Balls M. 1995. The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217.

de Wolf W., Comber M., Douben P., Gimeno S., Holt M., Léonard M., Lillicrap A., Sijm D., van Egmond R., Weisbrod A., Whale G. 2007. Animal use replacement, reduction, and refinement: development of an integrated testing strategy for bioconcentration of chemicals in fish. Integr Environ Assess Manag. 3, 3-17.

DIN 2001. German standard methods for the examination of water, waste water and sludge – Subanimal testing (group T) – Part 6: Toxicity to fish. Determination of the nonacute-poisonous effect of waste water to fish eggs by dilution limits (T 6). DIN 38415-6; German Standardization Organization.

Draize J.H., Woodard G. & Calvery H.O. 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. Journal of Pharmacology and Experimental Therapeutics 82, 377–390.

Doe J.E., Boobis A.R., Blacker A., Dellarco V., Doerrer N.G., Franklin C., Goodman J.I., Kronenberg J.M., Lewis R., Mcconnell E.E., Mercier T., Moretto A., Nolan C., Padilla S., Phang W., Solecki R., Tilbury L., van Ravenzwaay B., Wolf D.C. 2006 "A Tiered Approach to Systemic Toxicity Testing for Agricultural Chemical Safety Assessment". Critical Reviews in Toxicology, 36: 37-68.

ECETOC 2005. Alternative testing approaches in environmental safety assessment. Technical Report No 97. ISSN-0773-8072-97.

ECHA 2008a. Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?vers=20_08_08

ECHA 2008b. Guidance on information requirements and chemical safety assessment. Chapter R.7a: Endpoint specific guidance. http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7a_en.pdf?vers=20_08_08

ECHA 2008c. Guidance on information requirements and chemical safety assessment. Chapter R.7b: Endpoint specific guidance. http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7b_en.pdf?vers=20_08_08

ECHA 2008d. Guidance on information requirements and chemical safety assessment. Chapter R.7c: Endpoint specific guidance. http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7c_en.pdf?vers=20_08_08

EFSA 2004. Opinion of the Scientific Panel on Food Additives, Flavourings, Processing Aids and Materials in contact with Food (AFC) on a request from the Commission related to Flavouring Group Evaluation 11 (FGE.11): Aliphatic dialcohols, diketones, and hydroxyketones from chemical group 10 (Commission Regulation (EC) No 1565/2000 of 18 July 2000) (Adopted on 9 December 2004). http://www.efsa.europa.eu/EFSA/efsa_locale-1178620753812_1178620762424.htm

EFSA 2005. Scientific report. Aspects of the biology and the welfare of animals used for experimental and other scientific purposes. Annex to the EFSA Journal 292, 1-136.

EFSA 2006. "Opinion of the Scientific Panel on Plant Protection Products and their residues on a request from the Commission related to the revision of Annexes II and III to Council Directive 91/414/EEC concerning the placing of plant protection products on the market – residues". The EFSA Journal, 364, 1-27.

EFSA 2007a. Opinion of the Scientific Panel on Plant Protection Products and their residues on a request from the Commission related to the revision of Annexes II and III to Council Directive 91/414/EEC concerning the placing of plant protection products on the market – toxicological and metabolism studies. The EFSA Journal, 449, 1-60.

EFSA 2007b. "Opinion of the Scientific Panel on Plant Protection Products and their residues on a request from the Commission related to the revision of Annexes II and III to Council Directive 91/414/EEC concerning the placing of plant protection products on the market – ecotoxicological studies". The EFSA Journal 461, 1- 44.

EFSA 2007c. Opinion of the Scientific Panel on Plant protection products and their residues (PPR) on the Acute Reference Dose (ARfD) for imazalil, The EFSA Journal 460, 1-15.

EFSA 2007d. Opinion of the Scientific Committee on the Introduction of a Qualified Presumption of Safety (QPS) approach for assessment of selected microorganisms referred to EFSA - The EFSA Journal 587, 1-16.

EFSA 2008a. GMO Panel Report 'Safety and Nutritional Assessment of Genetically Modified Plants and Derived Food and Feed: The Role of Animal Feeding Trials, Food & Chemical Toxicology 46, supplement 1.

EFSA 2008b. The maintenance of the list of QPS microorganisms intentionally added to food or feed. Scientific Opinion of the Biological Hazards Panel. The EFSA Journal 923, 1-48.

EFSA 2008c. Scientific Opinion of the Panel on Plant protection products and their Residues (PPR) on the Science behind the Guidance Document on risk assessment for birds and mammals. The EFSA Journal (2008) 734, 1-181.

EPA 2005. The juvenile/peripubertal assays, also known as the "Female and Male Pubertal Assays". Available at: http://www.epa.gov/endo/pubs/edmvac/pubertal_story_111605.pdf

EPA 2008. The use of structure-activity relationships (SAR) in the high production Volume Challenge Program. Available at: http://www.epa.gov/oppt/chemrtk/pubs/general/sarfinl1.htm.

European Commission 2000a. Commission Directive 2000/33/EC of 25 April 2000 adapting to technical progress for the 27th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Annex I – B.40 Skin corrosion. Official Journal of the European Communities L 136; 91-97.

European Commission 2000b. Commission Directive 2000/33/EC of 25 April 2000 adapting to technical progress for the 27th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Annex II – B.41 Phototoxicity. Official Journal of the European Communities L 136; 98-107.

European Commission 2004. Annex 2D to Commission Directive 2004/73/EC adapting to technical progress for the 29th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances.

European Commission 2007. Commission recommendation 2007/526/EC on guidelines for the accommodation and care of animals used for experimental and other scientific purposes. Official Journal of the European Union L197/1, published on 30.7.2007.

European Convention on protection of vertebrate used for experimental and other scientific purposes – Strasburg 18.03.1986, amended 2 .12 2005 (ETS 123).
http://conventions.coe.int/Treaty/en/Treaties/Html/123.htm

FELASA 1994. Recommendations for the health monitoring of mouse, rat, hamster, guinea-pig and rabbit breeding colonies. Report of the Federation of European Laboratory Animal Science Associations Laboratory Animals 28, 1-12.

FELASA 2002. Recommendations for the health monitoring of rodent and rabbit colonies in breeding and experimental units. Laboratory Animals 36, 20–42.

Fentem J.H., Archer G.E.B., Balls M., Botham P.A., Curren R.D., Earl L.K., Esdaile D.J., Holzhütter H.G., Liebsch M. 1998. The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the management team. Toxicology in Vitro 12, 483-524.

Festing M.F.W., Overend P., Gaines Das R., Cortina Borja M., Berdoy M. 2002. The Design of Animal Experiments: Reducing the Number of Animals in Research through Better Experimental Design. Laboratory Animal Handbooks Series, 14. London: Royal Society of Medicine Press.

Flammang T.J., von Tungeln L.S., Kadlubar F.F., and Fu P.P.1997. Neonatal mouse assay for tumorigenicity: Alternative to the chronic rodent bioassay. Regul. Toxicol. Pharmacol. 26, 230-240.

Gallegos Saliner A., Patlewicz G., Worth A.P. 2006. Review of literature-based models for skin and eye irritation and corrosion. JRC report EUR 22320 EN. European Chemicals Bureau, Ispra, Italy. Accesible from: http://ecb.jrc.it.

Gelbke HP, Kayser M, Poole A. 2004. OECD test strategies and methods for endocrine disruptors. Toxicology. Dec 1; 205(1-2):17-25.

Gerner I., Graetschel G., Kahl J. & Schlede E. 2000a. Development of a decision support system for the introduction of alternative methods into local irritancy/corrositivity testing strategies. Development of a relational database. ATLA 28, 11–28.

Gerner I., Zinke S., Graetschel G. & Schlede E. 2000b. Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Creation of fundamental rules for a decision support system. ATLA 28, 665–698.

Gerner I., Spielmann H., Hoefer T., Liebsch M., Herzler M. 2004. Regulatory use of (Q)SARs in toxicological hazard assessment strategies. SAR QSAR Environ Res.;15(5-6):359-6.

Gerner I., Liebsch M., Spielmann H. 2005. Assessment of the eye irritating 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. ATLA 33: 351-67.

Halle W., Liebsch M., Traue D., Spielmann H. 1997. Reduction of the numbers of animals used for the classification of the acute oral toxicity of chemicals by taking into account cytotoxicity data from the Registry of Cytotoxicity. ALTEX 14 (1), 8-15.

Haneke K.E., Tice R.R., Carson B.L., Margolis B.H. & Stokes W.S. 2001. ICCVAM evaluation of the murine local lymph node assay. III. Data analysis completed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. Regulatory Toxicology and Pharmacology 34, 274–286.
Hinderling P.H. 1997. Red blood cells: a neglected compartment in pharmacokinetics and pharmacodynamics. Pharmacol Rev.; 49(3):279-95.

Hareng L., Pellizzer C., Bremer S., Schwarz M., Hartung T. 2005. The integrated project ReProTect: a novel approach in reproductive toxicity hazard assessment. Reprod Toxicol.; 20(3):441-52.

Hartmann E, Strauss V, Eiben R, Freyberger A, Kaufmann W, Loof I, Reissmueller. 2008. "ESTP comments on the draft updated OECD test guideline 407". Expo Toxicol Pathol, 59 (5): 297-300.

Hinderling PH. 1997. Commentary 'Prediction of absolute bioavailability for drugs using oral and renal clearance following a single oral dose: a critical review' (I. Mahmood). Biopharmaceutic and Drug Disposition 18 (9), 821-4.

Hoeger B., Jeram S., Holt M., Douben P., Halder M. 2006. Reduction of animal use in acute aquatic toxicity testing: Further development of the threshold approach and its application to existing chemicals and plant production products. Poster presentation at 16th Annual SETAC Europe meeting The Hague, NL.

Hoekzema C., Murk J.A., van de Waart B.J., van der Hoeven J.C.M., de Roode D.F. 2006. Alternative approaches can greatly reduce the number of fish used for acute toxicity testing. Environ. Toxicol. Chem. 22, 1322-1325.

Holzhütter H.G., Genschow E., Diener W., Schlede E. 2003. Dermal and inhalation acute toxic class methods: test procedures and biometric evaluations for the Globally Harmonized Classification System. Arch Toxicol 77, 243–254.

Hurtt M.E., Cappon G.D., Browning A. 2003. Proposal for a tiered approach to developmental toxicity testing for veterinary pharmaceutical products for food-producing animals. Food Chem Tox 41, 611-619.

Hutchinson T.H., Barrett S., Buzby M., Constable D., Hartmann A., Hayes E., Huggett D., Länge R., Lillicrap A.D., Straub J.O., Thompson R.S. (2003). A strategy to reduce the numbers of fish used in acute ecotoxicity testing of pharmaceuticals. Environ. Toxicol. Chem. 22, 3031-3036.

ICH 1997. S1B. Testing for Carcinogenicity of Pharmaceuticals. Available at: http://www.ich.org/cache/compo/276-254-1.html.

ISO 2007. Water quality -- Determination of the acute toxicity of waste water to zebrafish eggs (Danio rerio). ISO 15088:2007, International Organization for Standardization.

IPCS. Global Assessment of the State-of-the-Science of Endocrine Disruptors. International Programme on Chemical Safety. 2002. Editors: T Damstra, S Barlow, A Bergman, R Kavlock, and G van der Kraak. World Health Organization, Geneva. WHO/IPCS/EDC/02.2, 180 p. Also at http://ehp.niehs.nih.gov/who

Jeram S., Riego Sintes J.M., Halder M., Baraibar Fentanes J., Sokull-Klüttgen B., Hutchinson T.H. 2005. A strategy to reduce the use of fish in acute ecotoxicity testing of new chemical substances notified in the European Union. Regulatory Toxicology and Pharmacology 42, 218-224.

Kasper P, Uno Y, Mauthe R, Asano N, Douglas G, Matthews E, Moore M, Mueller L, Nakajima M, Singer T, Speit G; 2007. IWGT Workgroup. “Follow-up testing of rodent carcinogens not positive in the standard genotoxicity testing battery: IWGT workgroup report”. Mutat Res, 627 (1): 106-16.

Kimber I., Dearman R.J., Basketter D.A., Ryan C.A. & Gerberick, G.F. 2002. The local lymph node assay: past, present and future. Contact Dermatitis 47, 315–328.

Kimber I., Dearman R.J., Betts C.J., Gerberick G.F., Ryan C.A., Kern P.S., Patlewicz G.Y., Basketter D.A. 2006. The local lymph node assay and skin sensitization: a cut-down screen to reduce animal requirements? Contact Dermatitis; 54(4):181-5.

Kirkland D., Aardema M., Henderson L., Müller L. 2005. ”Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity” Mutat Res, 584 (1-2): 1-256.

Kirkland D., Aardema M., Müller L., Makoto H. 2006. Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens II. Further analysis of mammalian cell results, relative predictivity and tumour profiles. Mutat Res, 608 (1): 29-42.

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 J.R., 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(1): 31-55.

Kirkland D., Kasper P., Müller L., Corvi R., Speit G. 2008. Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: a follow-up to an ECVAM workshop. Mutat Res. 31;653(1-2):99-108. Epub 2008 Apr 11.

Krishna G. and Hayashi M. 2000. "*In vivo* rodent micronucleus assay: protocol, conduct and data interpretation". Mutat Res, 455 (1-2): 155-66.

Kroes R., Galli C., Munro I., Schilter B., Tran L.A., Walker R., Wurtzen G. 2000. Threshold of toxicological Concern for chemical substances present in the diet: a practical tool for assessing the need for toxicity testing. Food and Chemical Toxicology, 38, 255-312.

Kroes R., Renwick A.G., Cheeseman M., Kleiner J., Mangelsdorf I., Piersma, A., Schilter B., Schlatter J., van Schothorst F., Vos J.G. and Wurtzen G. 2004. Structure-based Thresholds of Toxicological Concern (TTC): Guidance for application to substances present at low levels in the diet. Food and Chemical Toxicology, 42, 65-83.

Liebsch M., Traue D, Spielmann H., Uphill P., Wilkins S., Wiemann C., Kaufmann T., Remmele M. & Holzhuetter H-G. 2000. The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. ATLA 28, 371-401.

Mantovani A. 2006 Risk assessment of endocrine disrupters. The role of toxicological studies. Annals of the New York Academy of Sciences 1076: 239–252.

Maurici D., Aardema M., Corvi R., Kleber M., Krul C., Laurent C., Loprieno N., Pasanen M., Pfuhler S., Phillips B., Prentice D., Sabbioni E., Sanner T., Vanparys P. 2005. Carcinogenicity. ATLA 33, suppl 1, 177-182.

McClain R.M., Keller D, Casciano D., Fu P., MacDonald J., Popp J., and Sagartz J. 2001. Neonatal Mouse Model: Review of Methods and results. Toxicologic.Pathology, vol 29 (Suppl) 128-137.

Meylan W.M., Howard P.H., Boethling R.S., Aronson D., Printup H., Gouchie S.1999. Improved method for estimating bioconcetration/bioaccumulation factor from octanol/water partition coefficient. Environ Toxicol Chem.;18:664–672.

Munro I.C., Ford R.A., Kennepohl E. and Sprenger J.G. 1996. Correlation of a structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. Food and Chemical Toxicology, 34, 829-867.

Munro I.C., Kennepohl E., Kroes R. 1999. A procedure for the safety evaluation of flavouring substances. Food and Chemical Toxicology, 37, 207-232.

Nagel R. 2002. DarT: The embryo test with the zebrafish *Danio rerio* – a general model in ecotoxicology and toxicology. Altern Tierexp 19:38-48.

Nagel R. and Isberner K. 1998. Testing of chemicals with fish – a critical evaluation of tests with special regard to zebrafish. In: Braunbeck T, Hinton DE and Streit B: Fish Ecotoxicology. Basel, pp 337-352.

National Institute of Environmental Health Sciences (NIEHS - 1997). Validation and Regulatory Acceptance of Toxicological Test Methods: A report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). NIH Publication No. 97-3981, NIEHS, Research Triangle Park, NC, USA.
http://iccvam.niehs.nih.gov/docs/about_docs/validate.pdf.

NIH 1999. Corrositex®: an *in vitro* test method for assessing dermal corrosivity potential of chemicals. NIH Publication No. 99-4495, 236pp. Research Triangle Park, NC, USA: NIEHS.

OECD 1996. Report of the SETAC/OECD Workshop on Avian Toxicity Testing. OECD/GD (96) 166.

OECD 2000. Environmental Health and Safety Monograph Series on Testing and Assessment No. 19: Guidance Document on the Recognition Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation.

OECD 2001. Environmental Health and Safety Monograph Series on Testing and Assessment No. 24: Guidance Document on Acute Oral Toxicity Testing, 24pp. Paris, France: Organisation for Economic Cooperation and Development.

OECD 2004. OECD principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationships Models.

OECD 2005. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Series on Testing and Assessment No. 34, ENV/JM/MONO(2005) 14, 96 pp. Paris, France: OECD.
http://appli1.oecd.org/olis/2005doc.nsf/linkto/env-jm-mono(2005)14
http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.html#OECD_Principles_for_the_Validation

OECD 2008a. Environmental Health and Safety Monograph Series on Testing and Assessment No. 39: Draft guidance document on acute inhalation toxicity testing. Paris, France: Organisation for Economic Cooperation and Development.

OECD 2008b. Draft Extended One-Generation Reproductive Toxicity Test. Test Guideline. Merged draft version 4, 2008-06-18.
Available at: http://www.oecd.org/dataoecd/22/20/40899803.pdf

OECD 2008c. OECD Quantitative Structure-Activity Relationships [(Q)SARs] Project. Available at: http://www.oecd.org/document/23/0,3343,en_2649_34379_33957015_1_1_1_1,00.html

OECD 2008d. Draft Detailed Review Paper on Transgenic Rodent Mutation Assays. Prepared by Lambert IB et al., Environmental Directorate, OECD, 2nd draft May 2008. Available at: http://www.oecd.org/dataoecd/43/38/40830957.pdf.

Pavan M., Netzeva T.I. and Worth AP 2006. EUR Technical Report 22327 EN. Review of QSAR Models for Bioconcentration. Ispra, Italy.

Pelkonen O., Boobis AR, Gundert-Remy U. 2001. *In vitro* prediction of gastrointestinal absorption and bioavailability: an experts' meeting report. Eur J Clin Pharmacol; 57(9):621-9.

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(3):167-75.

Piersma AH. 2004.Validation of alternative methods for developmental toxicity testing. Toxicol Lett. 149(1-3):147-53.

RIVM 2004. RIVM report 340700001/2004. Transgenic mice as alternatives in carcinogenicity testing: current status. Available at: http://rivm.openrepository.com/rivm/bitstream/10029/8911/1/340700001.pdf

Reuter U, Heinrich-Hirsch B, Hellwig J, Holzum B, Welsch F. 2003. Evaluation of OECD screening tests 421 (reproduction/developmental toxicity screening test) and 422 (combined repeated dose toxicity study with the reproduction/developmental toxicity screening test). Regul Toxicol Pharmacol.;38(1):17-26.

Robinson S., Chapman K., Delongeas JL., Donald E., David Dreher D., Festag M., Kervyn S., Lampo A., Nahas K., Nogues V., Ockert D. Quinn K., Old S., Pickersgill N., Somers K., Stark C., Stei P., Waterson L. 2008. A European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development, Regul. Toxicol. Pharmacol. 50, 345-52

Rorije E. & Hulzebos E. 2005. Evaluation of (Q)SARs for the prediction of Skin Irritation/Corrosion Potential. Physico-chemical exclusion rules. Final report for ECB contract IHCP.B430206. European Chemicals Bureau, Joint Research Centre, Ispra, Italy. Accessible from: http://ecb.jrc.it.

Russel W.M.S. & Burch R.L.(1959). The principles of humane experimental techniques, 238pp. London UK: Methuen

SCF 2001. Guidelines of the Scientific Committee on Food for the Presentation of an Application for Safety Assessment of a Substance to be Used in Food Contact Materials Prior to its Authorisation. Adopted 18 May 1990. Reports of the Scientific Committee for Food, Twenty-Sixth Series. Commission of the European Communities, Luxembourg. Available at: http://ec.europa.eu/food/fs/sc/scf/out82_en.pdf.

Schlede E. & Gerner I. (1995). The Draize eye test and progress in development and acceptance of alternatives to this test in Europe. In *The World Congress on Alternatives and Animals Use in the Life Sciences: Education, Research, Testing* (ed. A.M. Goldberg & L.F.M. van Zutphen), pp. 333–336. New York, USA: Mary Ann Liebert, Inc.

Schulte C., Nagel R. 1994. Testing acute toxicity in the embryo of zebrafish (*Brachydanio rerio*) as an alternative to the acute fish test: preliminary results. ATLA 22:12-19.

Seely JC. 2008. Toxicology and pathology considerations for the design of juvenile animal studies. Lab Anim; 37(5):206-9.

Silva O. de, Cottin M., Dami N., Roguet R., Catroux P., Toufic A., Sicard C., Dossou K.G., Gerner I., Schlede E., Spielmann H., Gupta K.C. & Hill R.N. 1997. Evaluation of eye irritation potential: statistical analysis and tier testing strategies. Food and Chemical Toxicology 35, 159–164.

Spielmann H., Balls M., Dupuis J., Pape W.J., Pechovitch G., de Silva O., Holzhütter H.G., Clothier R., Desolle P., Gerberick F., Liebsch M., Lovell W.W., Maurer T., Pfannenbecker U., Potthast J.M., Csato M., Sladowski D., Steiling W., Brantom P. 1998a. The International EU/COLIPA *In Vitro* Phototoxicity Validation Study: Results of Phase II (Blind Trial). Part 1: The 3T3 NRU Phototoxicity Test. Toxicology *In Vitro* 12, 305-327.

Spielmann H., Balls M., Dupuis J., Pape WJW., de Silva O., Holzhütter H.G., Gerberick F., Liebsch M., Lovell WW, Pfannenbecker U. 1998b. A Study on UV Filter Chemicals from Annex VII of European Union Directive 76/768/EEC, in the *in vitro* 3T3 NRU phototoxicity test. ATLA 26, 679-708.

Spielmann H., Genschow E., Liebsch M. and Halle W. 1999. Determination of the starting dose for acute oral toxicity (LD50) testing in the up and down procedure (UDP) from cytotoxicity data. ATLA 27, 957-966.

Spielmann H., Hoffmann S., Liebsch M., Botham P., Fentem J., Eskes C., Roguet R., Cotović J., Cole T., Worth A., Heylings J., Jones P., Robles C., Kandárová H., Gamer A., Remmele M., Curren R., Raabe H., Cockshott A., Gerner I. and Zuang, V. 2007. The ECVAM International Validation Study on *In Vitro* Tests for Acute Skin Irritation: Report on the Validity of the EPISKIN and EpiDerm Assays and on the Skin Integrity Function Test. ATLA 35, 559-601.

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: 245-74.

Sonneveld E., Riteco JA., Jansen HJ., Pieterse B., Brouwer A., Schoonen WG., van der Burg, B. 2006. Comparison of *in vitro* and *in vivo* models for androgenic and estrogenic activities. Toxicological Sciences 89, 173-187.

Stigler SM 1982, Thomas Bayes's Bayesian inference, J. Roy. Statist. Soc. Ser. A 145 (2), 250-258.

Tsakovska I., Saliner AG., Netzeva T., Pavan M., Worth AP. 2007. Evaluation of SARs for the prediction of eye irritation/corrosion potential: structural inclusion rules in the BfR decision support system. SAR QSAR Environ Res. 18(3-4):221-35.

van Oosterhout JP, Van der Laan JW, De Waal EJ, Olejniczak K, Hilgenfeld M, Schmidt V, Bass R. 1997. The utility of two rodent species in carcinogenic risk assessment of pharmaceuticals in Europe. Regul Toxicol Pharmacol 25: 6-17.

VICH 2002. Studies to evaluate the safety of residues of veterinary drugs in human food: Developmental toxicity testing. VICH GL 32. October 2002. Available at: http://www.vichsec.org/pdf/11_2002/Gl32_st7.pdf

Weil CS & Scala RA 1971. Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. Toxicology and Applied Pharmacology 19, 276–360

Weisbrod A.V., Burkhard L.P., Arnot J., Mekenyan O., Howard P.H., Russom C., Boethling R., Sakuratani Y., Traas T., Bridges T., Lutz C., Bonnell M., Woodburn K., Parkerton T. 2007. Workgroup report: review of fish bioaccumulation databases used to identify persistent, bioaccumulative, toxic substances. Environ Health Perspect: 115, 255-61.

Weyers A., Sokull-Klüttgen B., Baraibar-Fentanes J., Vollmer G., 2000. Acute toxicity data: a comprehensive comparison of results of fish, Daphnia and algae tests with new substances notified in the EU. Environ. Toxicol. Chem. 19, 1931-1933.

Worth A. and Balls M. 2002. The principles of validation and the ECVAM validation process. ATLA Suppl 2, 5-21.

Worth A.P. & Cronin M.T.D. 2001. The use of pH measurements to predict the potential of chemicals to cause acute dermal and ocular toxicity. Toxicology 169, 119–131.

ABBREVIATIONS:

AHAW:	Scientific Panel on Animal health and Animal Welfare
ANS:	Scientific Panel on Food Additives and Nutrient Sources added to food
ARfD:	Acute Reference Dose
ADI:	Acceptable Daily Intake
ADME:	Absorption Distribution Metabolisms and Excretion
BAF:	Bioaccumulation Factors
BfR:	German Federal Institute for Risk Assessment
BCF:	Bioconcentration Factor
BCOP:	Bovine Cornea Opacity and permeability
BIOHAZ:	Scientific Panel on Biological Hazards
BMD:	Benchmark Dose
CEF:	Scientific Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids
CEFIC:	European Chemicals Industry Council
COLIPA:	European Cosmetics Association
CONTAM**:**	Scientific Panel on Contaminants in the food chain
ECB:	European Chemical Bureau
ECDC:	European Centre for Disease Prevention and Control
ECHA:	European CHemicals Agency
ECVAM:	European Centre for the Validation of Alternative Methods
EMEA:	European Medicines Agency
EFSA:	European Food Safety Authority
EPAA:	European Partnership on Alternatives to Animal testing
ESAC:	ECVAM Scientific Advisory Committee
F1:	First generation
F2:	Second generation
FEEDAP:	Scientific Panel on Additives and Products or Substances used in Animal Feed
GLP:	Good Laboratory Practice
GMO:	Genetically Modified Organisms
HET-CAM:	Hen's Egg Test on Chorio-Allantoic Membrane
HCE:	Human Corneal Epithelia
IC50:	Inhibitory concentration 50% (concentration that causes death of 50% of the cells)
ICCVAM:	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE:	Isolated Chicken Eye
ICH:	International Conference on Harmonisation
ILSI:	International Life Science Institute
IPCS:	International Programme on Chemical Safety
JECFA:	Joint FAO/WHO Committee on Food Additives
JRC:	Joint Research Centre
LC50:	Lethal Concentration 50% (concentration that causes death of 50% of the test animals)
LD50:	Lethal Dose 50% (dose that causes or predicts the death of 50% of the test animals)
LDD50:	Lethal Daily Dietary Dose
MRL:	Maximum Residue Level
NDA:	Scientific Panel on Dietetic Products, Nutrition and Allergies
NIEHS:	National Institute of Environmental Health Sciences
NOEL:	No Observed Effect Level
NOAEL:	No Observed Adverse Effect Level
NOEC:	No Observed Effect Concentrations

OIE:	World Organisation for Animal Health
OECD:	Organization for Economic Cooperation and Development
PBPK:	Physiologically-Based PharmacoKinetics models
PBTK:	Physiologically-Based ToxicoKinetics models
PLH:	Scientific Panel on Plant Health
PPR:	Scientific Panel on Plant Protection Products and their Residues
QSAR:	Quantitative Structure-Activity Relationship
REACH:	Registration Evaluation Authorisation of Chemicals
SCF:	Scientific Committee on Food
TTC:	Threshold of Toxicological Concern

GLOSSARY

Term	Definition
Alternative method	Refers to all of the Three Rs. An alternative method can Replace the use of living sentient animals (Replacement alternative); Reduce the number of animals required for the same purpose (Reduction Alternative); Refine the animal test by reducing any pain, suffering, distress and lasting harm or enhancing animal well-being (Refinement Alternative).
Controls	The use of animals as negative controls e.g. vehicle alone, and positive control to establish the viability of test organisms and provide a comparator with treated groups.
Endpoint	The biological or chemical effect, response or process assessed by a test.
Laboratory/experimental animal	An animal used in a test or in a research experiment.
Humane endpoint	The earliest biological response in an animal that can be used to achieve the scientific objective in order to avoid or reduce animal pain, distress, suffering, or lasting harm.
In vitro	Literally meaning in glass, an experimental technique that may involve organs, tissues and cells (lines) derived from humans or animals.
In vivo	Scientific procedures involving a living animal with its whole body systems intact in order to study what happens in the body itself.
In silico	*In silico* is an expression used to mean using computer databases to predict scientific outcomes in safety testing (i.e. performed on computer or via computer simulation).
Regulatory testing	Testing required by national, European or international legislation.
Relevance	Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method.
Reliability	Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability.

Risk Assessment	A process intended to calculate or estimate the risk to a given target organism, system or (sub) population, including the identification of attendant uncertainties, following exposure to a particular agent, taking into account the inherent characteristics of the agent of concern as well as the characteristics of the specific target system. The risk assessment process includes four steps: hazard identification, hazard characterization (related term: dose response assessment), exposure assessment and risk characterization.
Scientific Procedure	A combination of one or more technical acts carried out on an animal for an experimental or other scientific purpose, with known or unknown outcome, and which may cause that animal pain, suffering, distress or lasting harm.
Surrogate species	A species other than the target animal in which the vaccine is to be used.
Target species	The species for which information on the potential toxicity or effects of a substance is sought.
Test	An experimental system used to obtain information on the adverse effects of a substance. It is used interchangeably with assay.
Three Rs	Is an ethical framework for humane experimentation that comprises three methodological principles: Replacement of living sentient animals in scientific procedures; Reduction in the number of animals used; and Refinement to cause less pain, distress, suffering or lasting harm, or enhance animal well-being.
Tiered approach	An approach which uses test information in a sequential testing framework such that the test methods selected in each succeeding level are determined by the results obtained in the previous level of testing.
Genetically modified (Transgenic) animal	A genetically modified organisms is an organism, with the exception of human being, in which the genetic material has been altered in a way that does not occur naturally by mating and/or recombination[36]. A transgenic animal can be designed to express foreign genes or to overexpress or silence endogenous genes.
Toxicological endpoint	see endpoint
Validation	The process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose.

[36] Definition taken from Directive 2001/18 art 2 (2)

ANNEX 1 – EU INTEGRATED PROJECTS:

A-Cute-Tox

The ACuteTox Integrated Project (http://www.acutetox.org), which is funded within the 6th Framework Program (6FP) of the DG RTD of the EU Commission, is designed to replace the animal testing for acute systemic toxicity, used today for regulatory purposes, by *in vitro* and *in silico* alternatives. In spite of the fact that earlier studies on acute systemic toxicity demonstrated a good correlation between *in vitro* basal cytotoxicity data (the 50% inhibitory concentration, IC50) in human cell lines and rodent LD50 values, and an even better correlation between IC50 values and human lethal blood concentrations, only very few tests have been accepted for general use. Therefore the A-Cute-Tox project is aiming to adapt new testing strategies, e.g. implementation of new endpoints and new cell systems for toxicity screening, organ-specific models, metabolism dependent toxicity, tissue absorption, distribution and excretion, computer-based prediction models. The challenge is to find the most appropriate and simple testing strategy for the prediction of human acute systemic toxicity, and to select a robust *in vitro* test battery for cytotoxicity testing of chemicals. Anticipated consequences results of this overall objective are decreased testing costs and improved scientific validity of the results. In order to realise such an *in vitro* test strategy, a number of building blocks are required.

Sens-it-iv

The overall goal of *Sens-it-iv* (http://www.sens-it-iv.eu) is to develop strategies to replace animal experimentation by *in vitro* assays for identifying skin and respiratory sensitisers to ensure with the use of safe ingredients by the chemical, cosmetic and pharmaceutical industry.

Allergies to sensitising agents are steadily increasing. Risk assessment for potential skin- or lung-sensitisers, mainly depends on animal testing. The overall objective of ***Sens-it-iv*** is to produce *in vitro* alternatives for these assays, and develop them up to the level of pre-validation. Besides reducing animal experimentation, an increase the accuracy of predicting sensitising potencies is expected.

In vitro mechanisms, relevant for *in vivo* sensitisation, will be identified at the level of human lung or skin epithelial cells (EC), dendritic cells (DC) and T-cells. These efforts imply specific scientific *(S)* and technologic *(T)* objectives:

- Existing data on sensitising, irritating and toxic compounds are collected *(S)*.
- *In vivo* changes induced by selected compounds in the specified cell types are described using functional genomics *(S)*.
- Similarly, the impact of compounds on individual cells, and the interaction between these cells is assessed *in vitro (S)*.
- The physico-chemical properties of chemicals responsible for metabolic activation and hapten-formation are determined (*S*).
- The data are collected in an *Inductive Database* allowing queries for data patterns and predictive models *(T)*.
- Mechanisms specifically involved in skin and respiratory sensitisation are identified using bio-informatics *(S)*.

- The information is used to adapt/improve existing techniques, and to develop organotypic models derived from human cells assays *(T)*.
- A proof of principle is established on a set of selected skin and respiratory sensitisers, irritants and toxic compounds *(T)*.

Sens-it-iv is innovative primarily by the coordinated and extensive characterisation of the impact of compounds on cell-cell interactions for identification of the key mechanisms of sensitisation.

Consistency with existing and ongoing projects and optimal exploitation of these achievements is ensured by the involvement of ECVAM-IHCP-JRC, COLIPA, ECOPA, IVTIP and OECD in various levels of the project.

ReProTect

In certain areas of reproductive toxicity testing, a number of useful and promising *in vitro* models are already available but they need to be converted into tests with a predictive power for toxicological safety testing. The 6th Framework Programme project **ReProTect** (Hareng et al., 2005; www.reprotect.eu) is aiming to optimize these tests in order to prepare them for formal validation studies. Targeting a limited set of effects/mechanisms is a potential shortcoming of individual *in vitro/in silico* assays for complex events such as reproductive and developmental toxicity. Thus, ReProTect aims at optimizing an integrated set of tests as a basis for a reproductive/developmental battery. *In vitro* test batteries will provide more detailed understanding of the main chemical target tissues or targeted biological mechanism in the reproductive cycle such as gametogenesis, steroidogenesis, embryogenesis etc. and will so support regulatory decisions.

The ReProtect project is divided into 4 workpackages and it exploits a network of 36 partners (including industry) throughout EU, whose final outcomes are foreseen in 2009. Most experimental units use either established cell lines or materials from human specimens (placenta, sperm) or veterinary *in vitro* fertilization (sperm, oocytes).

Three workpackage areas (*Fertilization* –dealing with male and female fertility-; *implantation* –dealing with placental function; *prenatal development*) develop a range *of cell- and tissue-specific tests* assaying targeted functional endpoints; thus the assays do target leyding/Sertoli cells, sperm, granulose cells, oocytes, trophoblasts, embryonic stem cells, etc.

The last workpackage (*cross-cutting technology*) aims at exploiting recent progresses of molecular biology to develop *mechanism-based* assays that may be potentially high-throughput: receptor binding/transactivation, transcriptomics, metabolic activation as well as *in silico* screen for structure-activity.

Thus the final objective, besides prevalidation of individual tests, is the integration of a range of assays targeting major reproductive endpoints with cross-cutting technologies identifying mechanisms.

Carcinogenomics

The major aim of the EU FP6 project **CARCINOGENOMICS** (http://www.carcinogenomics.eu/) is to develop *in vitro* methods for assessing the carcinogenic potential of compounds, as an alternative to current rodent bioassays for genotoxicity and

carcinogenicity. The major goal is to develop a battery of mechanism-based *in vitro* tests accounting for various modes of carcinogenic action. These tests will be designed to cover major target organs for carcinogenic action e.g. the liver, the lung, and the kidney. The novel assays will be based on the application of ”omics” technologies (i.e. genome-wide transcriptomics as well as metabonomics) to robust *in vitro* systems (rat/human), thereby also exploring stem cell technology, to generate “omic” responses from a well-defined set of model compounds causing genotoxicity and carcinogenicity. Phenotypic markers for genotoxic and carcinogenic events will be assessed for the purpose of anchoring gene expression modulations, metabolic profiles and mechanism pathways. Through extensive biostatistics, literature mining, and analysis of molecular-expression datasets, differential genetic pathways will be identified capable of predicting mechanisms of chemical carcinogenesis *in vivo*. Furthermore, generated transcriptomic and metabonomic data will be integrated into a holistic understanding of systems biology, and applied to build an iterative *in silico* model of chemical carcinogenesis. Subsequently, predictive gene expression profiles, typically consisting of some 150-250 genes, will be loaded onto high throughput dedicated DNA-chips, thus accelerating the analysis of transcriptomic responses by a factor of 100. It is expected that the outcome of this project will generate a platform enabling the investigation of large numbers of compounds for their genotoxic and carcinogenic potential, as for example envisaged for the new chemical policy REACH . This will contribute to speeding the identification of potential harmful substances to man, while lowering costs and reducing animal tests.

ANNEX 2:

INDICATIVE NUMBER OF ANIMALS REQUIRED ACCORDING TO OECD TEST GUIDELINES (TG)[37]

Please note that the table refers only to the TGs cited in the present Opinion and is not a comprehensive list of all the OECD TGs.

OECD/ EU test guideline	Endpoint	Number of animals Min - max	Comments
OECD 203/ EU C.1	Fish, Acute Toxicity Test	60-70	Limit test for fish toxicity >100mg/l: 14-21 fish; LC50 42-60 fish.
OECD 204	Fish, Prolonged Toxicity Test, 14 day study	20 - 50	Number of concentrations not given, 10 animals/dose, control group.
OECD 205	Avian Dietary Toxicity Test	150	A minimum of five test diets required. Two control groups and one treatment group for each of the five dietary levels. Limit test at 5000 ppm (parts per million) in the diet (30 animals with limit test)
OECD 206	Avian Reproduction Test	48-96	Different species can be used. Number of animals varies according to the species used.
OECD 210	Fish, Early-Life Stage Toxicity Test	360	It lasts until all control fish are free feeding.
OECD 212/ EU C. 15	Fish, Short-term Toxicity Test on Embryo and Sac-Fry Stages	180	Continues to free-feeding larval stage.
OECD 215/ EU C. 14	Fish, Juvenile Growth Test	number not stated	5 dose-level, control, number/group chosen according test design and statistical power.
OECD 305/ EU C.13	Bioconcentration: Flow-through Fish Test	108 - 144	2 dose-level, control, if necessary solvent control; sampling phase 4x5, depuration phase 4x4 fish.
OECD 402/ EU B.3	Acute Dermal Toxicity	15 (5X3)	10 animals with limit test since two sex are required using this approach. Usually rat, rodent of guinea pigs. If other species are used, it should be justified. No controls required by the guideline. Only one sex is used.

[37] http://www.oecd.org/document/40/0,3343,en_2649_34377_37051368_1_1_1_1,00.html

OECD 403/ EU B.2	Acute Inhalation toxicity	30	No controls required by the guideline. 10 animals with limit test. Rat as preferred species.
OECD 404/ EU B.4	Acute dermal Irritation/corrosion	1-3	Rabbit as preferred species.
OECD 405/ EU B.5	Acute eye irritation/corrosion	1 (2 for satellite* group if needed)	Rabbit as preferred species.
OECD 406/ EU B. 6	Skin sensitization	15-20	Guinea pigs as preferred species.
OECD 407/ EU B.7	Repeated dose 28-day oral toxicity study in rodent	40-60 (including satellite* group)	20 animals with limit test. Rat as preferred species. *An updated draft is in preparation, considering parameters for endocrine effects.*
OECD 408/ EU B.26	Repeated dose 90-day oral toxicity study in rodent	80-100 (including satellite* group)	Rat as preferred species.
OECD 409/ EU B. 27	Repeated dose 90-day oral toxicity in non-rodent	32-48 (including satellite* group)	16 animals with limit test. Dogs, mini-pigs or swine are used.
OECD 410/ EU B.9	Repeated dose dermal toxicity: 21-28 day study	40 -50 (including satellite* group)	20 animals with limit test. Rat, rabbit or guinea pig as preferred species.
OECD 411/ EU B. 28	Sub-chronic dermal toxicity: 90-day study	80 - 100 (including satellite* group)	40 animals with limit test. Rat, rabbit or guinea pig as preferred species.
OECD 412/ EU B. 8	Repeated dose inhalation toxicity: 28-day or 14-day study	40-50 (including satellite* group)	20 animals with limit test. Rat as preferred species.
OECD 413/ EU B. 29	Sub-chronic inhalation toxicity: 90-day study	80 – 100 (including satellite* group)	No limit test foreseen. Rat as preferred species.
OECD 414	Prenatal Developmental Toxicity Study	64-80	40 animals with limit test. Rat or mouse as preferred species.
OECD 415/ EU B. 34	One generation Reproduction Toxicity Study	160 + about 1200 offspring (80 litters, about15 pups/litter)	40 animals + about 600 offspring (40 litters, about 15 pups/litter) with limit test. Rat and mouse as preferred species.
OECD 416/ EU B. 35	Two generation reproductive toxicity	160 + about 2400 offspring (80 litters, about15 pups/litter x	160 adult animals in parent generation, 160 adult animals in F1 generation and offspring.

	study	2 generations)	Rat as preferred species.
OECD 417/ EU B. 36	Toxicokinetics	8-24	Rodent and non-rodent can be used. Two animals per dose if using non-rodent species.
OECD 418/ EU B.37	Delayed neurotox of organophosphorus substances following acute exposure	36 (hens)	Limit test can be used.
OECD 419/ EU B.38	Delayed neurotox of organophosphorus substances: 28-day repeated dose	48 (hens)	24 animals with limit test.
OECD 420/ EU B. 1 bis	Acute oral toxicity – Fixed Dose Method	5-7	Up to 5 animals with limit test (OECD GD 24). Rat as preferred species.
OECD 421	Reproduction/developmental toxicity screening test	80	40 animals with limit test. Rat as preferred species.
OECD 422	Combined repeated dose toxicity with the reproduction/developmental toxicity screening test	80 – 100 (including satellite* group)	40 animals with limit test. Rat as preferred species.
OECD 423/ EU B. 1 tris	Acute oral toxicity - Acute Toxic Classic Method	Average 7 (only one sex)	Up to 6 animals with limit test (OECD GD 24). Rat as preferred species.
OECD 424/ EU B.43	Neurotoxicity study in rodent	80	40 animals with limit test. Combined with chronic study: • 50 animal per dose = 200 • Combined with 90-day study: 30 animals per dose = 120 Rat as preferred species.
OECD 425	Acute oral toxicity - Up and Down Procedure	6-9	5 animals with limit test (OECD GD 24). One sex is used. Rat as preferred species.
OECD 429/ EU B.42	Skin sensitization: Local Lympho Node Assay (LLNA)	20	Mouse as preferred species. Female as preferred sex.
OECD 451/ EU B.32	Carcinogenicity study	400 per species	Two species are recommended = 800 animals. Rat and mouse as preferred species, but also non-rodent species may be used.

OECD 452/ EU B.30	Chronic toxicity study	80 (if rodent) 32 (if non-rodent)	Rat as preferred species. Dogs and primate as preferred species.
OECD 453/ EU B.33	Combined chronic toxicity/carcinogenicity study	400	60 animals for satellite group. Rat as preferred species.
OECD 474/ EU B.12	Mammalian erythrocyte micronucleus test	25-50	25 animals in case only one sex is used. 20 (or 10) animals if limit test is used. Mouse or rat as preferred species.
OECD 475/ EU B.11	Mammalian bone marrow chromosomal aberration test	25-50	25 animals in case only one sex is used. 20 (or 10) with limit test. Rodents are used in the test.
OECD 478/ EU B.22	Genetic toxicology: rodent dominant lethal test	Male to produce between 30-50 pregnant female	Rat or mouse as preferred species.
OECD 483/ EU B.23	Mammalian spermatogonial chromosome aberration test	25 male	15 animals with limit test. Hamster and mouse as preferred species.
OECD 486/ EU B.39	Unscheduled DNA Synthesis with mammalian liver cells *in vivo*	8-10	Allow only 1-2 animals for CTRs (positive and negative). 5 animals if limit test is used. Rat as preferred species.
OECD 505	Residues in livestock	In ruminants or monogastrics: 10-13 In hens: 30-34	In addition to CTR for the experiments, CTR animals should also used to determine adverse effects on egg production or milk yield.

The animals used in the dose-range finding experiments were not considered in the table since they are usually not indicated in the guidelines.

*Satellite group of animals are sometimes recommended in the test guidelines to observe effects after the completion of the experiment.