IPCS

INTERNATIONAL PROGRAMME ON CHEMICAL SAFETY

WHO ILO UNEP

Environmental Health Criteria 240

Principles and Methods for the Risk Assessment of Chemicals in Food

Chapter 4
HAZARD IDENTIFICATION AND CHARACTERIZATION: TOXICOLOGICAL AND HUMAN STUDIES

A joint publication of the Food and Agriculture Organization of the United Nations and the World Health Organization

Food and Agriculture Organization of the United Nations

World Health Organization

This report contains the collective views of an international group of experts and does not necessarily represent the decisions or the stated policy of the United Nations Environment Programme, the International Labour Organization or the World Health Organization.

Environmental Health Criteria 240

PRINCIPLES AND METHODS FOR THE RISK ASSESSMENT OF CHEMICALS IN FOOD

A joint publication of the Food and Agriculture Organization of the United Nations and the World Health Organization

Published under the joint sponsorship of the United Nations Environment Programme, the International Labour Organization and the World Health Organization, and produced within the framework of the Inter-Organization Programme for the Sound Management of Chemicals.

Food and Agriculture Organization of the United Nations

World Health Organization

The **International Programme on Chemical Safety (IPCS)**, established in 1980, is a joint venture of the United Nations Environment Programme (UNEP), the International Labour Organization (ILO) and the World Health Organization (WHO). The overall objectives of the IPCS are to establish the scientific basis for assessment of the risk to human health and the environment from exposure to chemicals, through international peer review processes, as a prerequisite for the promotion of chemical safety, and to provide technical assistance in strengthening national capacities for the sound management of chemicals.

The **Inter-Organization Programme for the Sound Management of Chemicals (IOMC)** was established in 1995 by UNEP, ILO, the Food and Agriculture Organization of the United Nations, WHO, the United Nations Industrial Development Organization, the United Nations Institute for Training and Research and the Organisation for Economic Co-operation and Development (Participating Organizations), following recommendations made by the 1992 UN Conference on Environment and Development to strengthen cooperation and increase coordination in the field of chemical safety. The purpose of the IOMC is to promote coordination of the policies and activities pursued by the Participating Organizations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

WHO Library Cataloguing-in-Publication Data

Principles and methods for the risk assessment of chemicals in food.

(Environmental health criteria ; 240)

1. Risk assessment. 2. Hazard assessment. 3. Exposure assessment. 4. Dose-response assessment. 5. Chemicals. 6. Food safety. 7. Food additives. 8. Contaminants. 9. Pesticide residues. 10. Veterinary drug residues. I.World Health Organization. II.Food and Agriculture Organization of the United Nations.

ISBN 978 92 4 157240 8 (NLM classification: WA 712)
ISSN 0250-863X

This document was technically and linguistically edited by Marla Sheffer, Ottawa, Canada.

Printed by Wissenchaftliche Verlagsgesellschaft mbH, Stuttgart, Germany.

4. HAZARD IDENTIFICATION AND CHARACTERIZATION: TOXICOLOGICAL AND HUMAN STUDIES

For acronyms and abbreviations used in the text, the reader may refer to the list of acronyms and abbreviations at the front of this monograph. Definitions of select terms may be found in the glossary at the end of the monograph.

4.1 Introduction

Toxicological studies may be broadly divided into in vitro studies, using cultured organisms or cells or tissue preparations from laboratory animals or humans, and in vivo studies in laboratory animals or humans. Such studies serve a number of purposes, including:

- identification of potential adverse effects;
- definition of the exposure conditions necessary to produce the effects;
- assessment of dose–response relationships for the adverse effects, including definition of dose levels that do not produce the effects; and
- interpretation of experimental data for risk assessment purposes, such as information on the mode of action and its relevance to humans and metabolism and toxicokinetic data that allow extrapolation of the data from laboratory animals to humans and to population subgroups.

A number of factors can influence the selection of appropriate methods for the toxicological testing of substances in food. Not all substances in food can or need to be tested toxicologically to the same degree or subjected to the same range of toxicity tests. The following text lists important factors to consider in the selection of test methods.

4.1.1 Nature of substances to be evaluated

The nature of the substance and its uses and levels of use can all influence the extent of toxicity testing necessary for risk assessment:

- The selection of test methods is governed to an extent by the nature of the substances to be tested.

- Substances evaluated by the Joint Food and Agriculture Organization of the United Nations (FAO)/World Health Organization (WHO) Expert Committee on Food Additives (JECFA) and the Joint FAO/WHO Meeting on Pesticide Residues (JMPR) range from single chemicals ingested in small amounts, such as contaminants, flavours, pesticides and certain food additives, to complex substances that may comprise a substantial portion of the diet, such as major food ingredients and whole foods.

- Substances consumed in small amounts can readily be subjected to appropriate and relevant toxicity tests, in which high dose levels can be used to increase the sensitivity of hazard identification. The majority of the tests discussed in this chapter are most readily applicable to low molecular weight, single-chemical entities.

- For substances consumed in large amounts, standard toxicity studies, while applicable, need to be designed and interpreted with caution because of possible physiological or nutritional perturbations that may be induced in test animals.

- For substances consumed in large amounts, human studies can play a significant role in assessing the tolerability of such substances.

4.1.2 Knowledge requirements for substances to be tested and evaluated

Prior to embarking on any toxicological testing of substances found in or intended for use in food, data should be available in several key areas:

- For a substance added either directly or indirectly to foods, information should be available on its source, including data on its manufacture (including aspects of Good Manufacturing Practice [GMP]) and appropriate information on its purity and specifications as a food-grade material. It is important that the substance being tested and evaluated is representative of that added to or present in food (see chapter 3).

- Knowledge of potential interactions of the substance with components of the foodstuff during processing and storage is essential in some cases to ensure that the appropriate chemical species are being tested and evaluated.

- Chemical speciation is important to consider for contaminants, residues of pesticides, packaging materials and residues of veterinary drugs, in order to ensure that toxicological and other studies are related to the chemical form or species that occurs in food.

4.1.3 Role of structure–activity relationships and metabolic fate

Careful examination of the composition, structure and known or presumed metabolic fate of the test substance should be undertaken prior to toxicity testing of substances added to or found in food. Examination of substances for structural alerts for toxicity can provide valuable guidance in the design of appropriate safety tests.

The general approach to safety evaluation should begin with an evaluation of the molecular structure of the substance in question. Some substances used as food additives and a large number of flavours are known to be endogenous substances or known or predicted to be readily converted in vivo into endogenous substances. Other substances may be known or presumed to be readily converted to metabolic products that could be considered harmless under the intended conditions of use of the parent substance. This may limit the extent to which such substances need to be subjected to toxicological testing.

Substances with structural alerts for specific forms of toxicity, such as neurotoxicity in the case of organophosphorus compounds or genotoxicity in the case of certain epoxides, nitrosamines, etc., should be subjected to detailed toxicological investigation, paying particular attention to that specific toxicity alert. Literature sources of knowledge regarding structure–activity relationships should be fully consulted before designing and conducting toxicity tests, especially to determine the need for any special studies related to identified safety concerns.

For substances intended to be consumed in large amounts, knowledge of the structure and metabolic fate may provide guidance on the interpretation of certain toxicological or physiological end-points. Substances that undergo colonic fermentation or produce caecal or colonic enlargement when given in large amounts or substances that raise the osmotic pressure of the colon often produce a cascading series of physiological events culminating in toxicological responses that may not be relevant to exposures encountered under conditions of practical use. Examples are polyols, which can produce hyperplasia of the adrenal medulla and phaeochromocytomas indirectly associated with abnormal calcium homeostasis, and the fat replacer olestra, which can produce adverse effects in high-dose animal studies by interfering with the absorption of fat-soluble vitamins.

For substances consumed in large amounts, secondary effects may limit the usefulness of conventional toxicological tests in assessing their safety, leading to an increased need to conduct appropriate and relevant studies in humans.

For substances for which there is no prior available knowledge of metabolic fate and pharmacokinetics (see section 4.2), such studies should be conducted prior to initiating large-scale toxicological studies.

4.1.4 Integrating data on dietary exposure

The extent and nature of testing that are considered adequate for a toxicological evaluation of a substance that is present in food should be based not only on any data on structure–activity relationships and metabolic fate, but also on presumed or known exposure:

- Exposure assessment should consider the likely duration and pattern of exposure (acute, short-term, long-term, intermittent, etc.) and the nature of the population that is likely to be exposed (e.g. the whole population or specific subgroups), as well as the potential for changes in exposure over time.

- Toxicological valuation of substances present in the diet at very low levels, such as flavouring agents (see chapter 9, section 9.1.2), may be based on data for structural analogues or more general thresholds of toxicological concern (TTCs) (chapter 9, section 9.1.1).

- TTCs (FAO/WHO, 1995, 1997, 2000b; Munro et al., 1996; Kroes et al., 2004), which define human exposure thresholds for different structure-based chemical classes, may be used to provide guidance on the degree of testing required (see also chapter 9, section 9.1.1).

4.1.5 General approach to toxicity testing

Several internationally recognized organizations, such as the Organisation for Economic Co-operation and Development (OECD), provide guidance for minimum standards for the design and conduct of toxicological studies. Hence, the following is a guide to general principles. All studies used in the risk assessment of a substance in food should be assessed for adequacy of design and conduct; for recent studies, this should include compliance with Good Laboratory Practice (GLP) (see chapter 3).

In making an assessment of the need for and extent of toxicity testing required for substances added to food, the following information needs to be considered in an integrated fashion: 1) structure–activity relationship, 2) metabolic fate and 3) exposure. The stepwise approach to assessing toxicity testing needs is illustrated in Figure 4.1.

4.1.5.1 *Role of in silico and in vitro studies*

It is generally accepted that animal testing should be reduced, refined or replaced as far as is practicable, and this has led to an increased use of alternative approaches. While recognizing the desirability of this, it is important that scientifically sound methods and approaches are used for the safety testing of food chemicals. Hence, although advances are being made in the development of in silico and in vitro approaches, at the present time these do not permit the replacement of animal testing for most end-points of concern.

In silico approaches encompass a wide range of methods, ranging from simple quantitative structure–activity relationships (QSAR) to sophisticated multiparametric simulation and even prediction based on quantum chemistry and other fundamental approaches.

At the present time, only a limited number of in silico and in vitro methods have been adopted by the OECD and other organizations involved in method approval. In a few instances, in vitro methods have been recognized as generally valid for risk assessment purposes, particularly in genotoxicity testing, but also for assessing some non-genotoxic end-points, such as corrosivity and phototoxicity. The use of in vitro methods for these purposes can provide robust data for risk assessment. Where non-standard methods are used as part of a data submission, evidence of their performance characteristics and validation should be provided.

In silico methods are a practical means of comparing the sequence of proteins and peptides with those of known allergens to determine whether there are epitopes in common, although the reliability of this approach is not high. In vitro methods are useful in determining

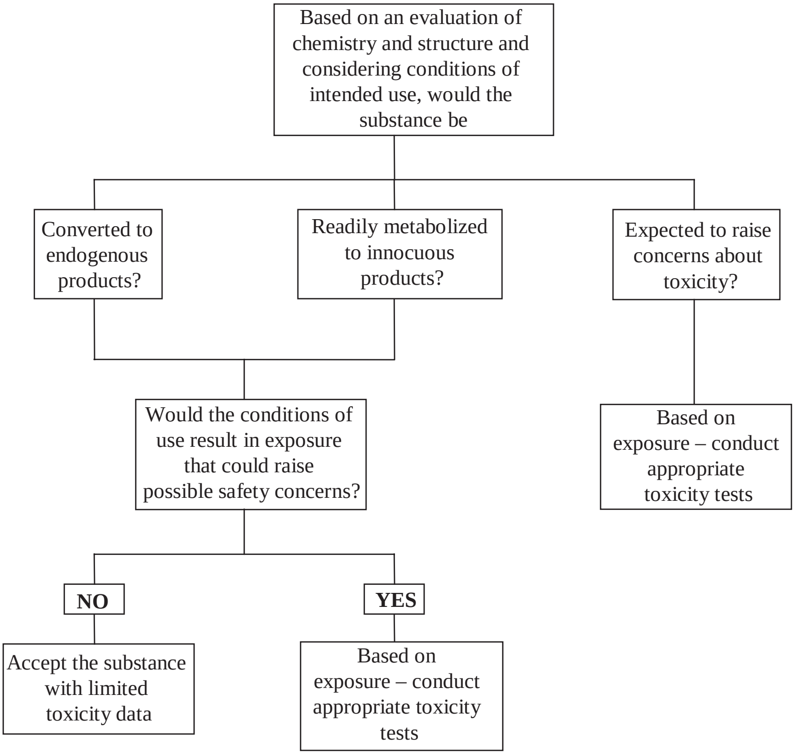


Fig. 4.1. A stepwise approach to assessing toxicity testing needs

the stability of proteins and peptides in digestive juices, such as gastric acid.

Mechanisms of toxicity are often investigated using in silico and in vitro methods. The results of such studies should be incorporated into a weight of evidence consideration of toxicity. In addition, such studies can provide insight into the relevance to humans of findings in experimental animals.

Also, in silico and in vitro methods are being used increasingly to characterize the metabolism of chemicals. Often, these data provide an invaluable bridge between laboratory animals and humans. Data derived from in silico and, even more so, from in vitro methods provide the basis for many physiologically based toxicokinetic (PBTK) models. Information that may be obtained in this way includes kinetic parameters for metabolism of the chemical, blood–tissue partition coefficients and plasma protein binding. Data can be obtained for both laboratory species and humans.

4.1.5.2 *Digestion and impact on gut flora*

Many substances in food have the potential to affect the gut flora, but some effects occur in experimental animals only when fed very high doses—for example, with poorly absorbed substances, such as polyols and modified starches. For such substances, effects in humans are extremely unlikely if the maximum human exposure is only a small fraction of the doses used in laboratory animal studies.

During the testing for systemic toxicity, experimental animals should be monitored routinely for possible direct and indirect effects on the gastrointestinal tract, by assessment of behaviour and clinical signs, biochemistry (serum and urine), gross morphology and histopathology. Where there are indications from toxicity tests of an effect on the gastrointestinal tract (e.g. caecal enlargement, diarrhoea), the reasons for this should be investigated.

Specific tests on the gut microflora should be carried out when there is an obvious potential for an effect on the gut flora, such as from an antibiotic. In testing for effects on the gut flora, several aspects should be considered, such as alteration of barrier effect and emergence of antimicrobial resistance. The choice of test system should be informed by the end-point of concern. Due consideration needs to be given to the nature of the microflora to be tested and the conditions under which the test will be conducted.

Where there is concern for an effect of the microflora on the substance—for example, in digestion or the production of microflora-specific metabolites—ex vivo studies could be undertaken using an appropriate selection of microflora of laboratory animal or human origin (see section 4.12).

4.1.5.3 *Absorption, distribution, metabolism and excretion (ADME)*

Studies on the fate and behaviour of substances in food are important in the design and interpretation of toxicity studies and in extrapolation to humans (IPCS, 1986a; Lipscomb & Ohanian, 2007). Interspecies and intraspecies differences in the kinetics of a substance are often a major contributory factor to interspecies and interindividual variation in response. Hence, a detailed understanding of the kinetics of the substance may enable some of the default uncertainty factors to be replaced with a chemical-specific adjustment factor (CSAF) (see IPCS [2005] and also chapter 5 for further discussion of uncertainty factors and CSAFs). ADME is described in section 4.2.

4.1.5.4 Considerations in the selection of appropriate in vivo studies and relevant species (models)

Although no experimental species is an ideal substitute for humans, there is extensive evidence that studies in test animals generally provide an effective means for evaluating the potential toxicity of substances in food, provided that the data are interpreted critically. Studies in experimental animals allow evaluation of toxicity to all mammalian organs and tissues and to physiological and metabolic processes and integrative functions. An important pragmatic factor influencing the choice of species and strain is the availability of historical control data; the absence of such data can severely limit the interpretation of equivocal findings.

The species selected should reflect the underlying biology of the end-point of concern and be of relevance to human biology. Hence, for studies of effects on fertility or development, animals of the appropriate life stage and reproductive capacity need to be selected, whereas animals of the appropriate sex (and often both sexes) would be used for potential effects on endocrine systems. However, not all such issues are resolved. For example, it is debatable as to which is the appropriate life stage in experimental animals for certain life stages in humans (e.g. children aged 1–3 years).

In selecting an animal model, its potential relevance to humans needs to be considered. There may be strain-specific or species-specific differences in metabolism or response such that findings for certain types of substance will not be relevant. For example, the CF-1 mouse is not a good animal model for investigating substances that show P-glycoprotein-dependent limits to their absorption.

The species and strain selected should be susceptible to the type of toxic effect being investigated. For example, some species or strains are known to be less susceptible to developmental toxicity than others.

Although test species and humans have many common pathways of foreign compound metabolism, it is unlikely that a species will be found that exhibits exactly the same metabolic profile for a substance as humans. Ideally, the species used in toxicity studies should produce all of the metabolites formed in humans. If human-specific metabolites are identified, it might be necessary to conduct toxicity studies with the metabolites themselves.

4.1.5.5 *Types of animal studies and their role in safety assessment*

Studies should be such that the toxicity of the substance can be assessed for all known or predicted exposure scenarios, for all relevant subgroups of the population and for all potential effects. As discussed above (section 4.1.4), the extent of testing necessary for regulatory purposes is related to the extent of human exposure.

Most end-points are adequately addressed by current study designs, such as the OECD testing guidelines (http://masetto.sourceoecd.org/vl=2781582/cl=14/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/v1n4/contp1-1.htm), but there are some specific types of toxicity or circumstances of exposure where there may be a need for modification of or even novel study designs. An example is the assessment of acute toxicity other than lethality, for which there is currently no approved study protocol. The exact choice of studies will depend on considerations of likely human exposure duration, the population to be exposed and any prior information on the substance.

It is not always necessary to test the substance specifically to cover all situations. It may be possible to adopt conservative assumptions, using a non-optimal study. For example, in the case of acute risk, if the predicted human exposures are well below the health-based guidance value, such as an acute reference dose (ARfD; see chapter 5, section 5.2.9) derived using data from a 90-day study, further refinement of the risk assessment would not be necessary. Conversely, should exposure assessment indicate a possible risk, a specific study of acute toxicity could be undertaken to help refine the risk assessment.

The lethality of the substance should be determined, but only up to a limit dose. This has been set at 2 or 5 g/kg body weight. Any non-lethal effects should be reported, as these may provide evidence for mechanism of lethality or of non-lethal acute toxicity.

In both short- and long-term studies, a wide range of end-points is investigated, including clinical signs, body and organ weights, clinical chemistry and haematology, urinalysis, and gross and histopathological examination of organs. These may be supplemented by validated biomarkers for specific effects.

The effects of the substance when administered short term should be assessed; this usually involves studies for about 10% of lifespan (e.g. 90 days in rat, 1 year in dog), although valuable data may be derived from extensive studies of shorter duration in rats or dogs. The need for two species, one non-rodent, should be considered.

Long-term studies for chronic toxicity and carcinogenicity should be conducted; these are usually of 2 years' duration in rodents, which is more or less equivalent to "lifetime" exposure. Such an extended duration may increase the sensitivity to detect cancer at the expense of a reduced sensitivity for other effects because of masking by age-related changes, although data obtained from interim results at 1 year could avoid this complication in evaluating toxicity.

The genotoxicity of the substance should be evaluated using a range of appropriate in vitro tests for mutation (bacteria), chromosomal damage and changes in chromosome number. Positive results should be confirmed in an in vivo genotoxicity study. In the absence of evidence to the contrary, a substance that is an in vivo genotoxin would be presumed to be a genotoxic carcinogen.

The relevance to humans of any tumorigenic response observed on administration of the substance to experimental animals should be assessed using a structured framework (Boobis et al., 2006).

The need for two species for the cancer bioassay, or indeed the need for a bioassay at all, should be considered. Alternative strategies might include a tiered approach involving genotoxicity testing, investigation of precursor effects for non-genotoxic carcinogenicity in short-term studies and the use of genetically modified animals (Gulezian et al., 2000).

The effects of the substance on reproductive performance of both males and females should be determined, if appropriate. The duration of exposure of the animals, relative to life stage, needs to be considered. For most substances, it will be necessary to consider the effects on embryonic and fetal development by treating pregnant dams. The need for two species for developmental testing should be considered.

The potential accumulation of the chemical also needs to be taken into account in the design and interpretation of such toxicity studies (e.g. the body burden of dioxins accumulates over a period of weeks of treatment).

Although studies such as those mentioned above should detect functional and structural effects on most tissues and organs, there are some systems for which additional testing may be required as appropriate. These include nutritional effects, neurobehavioural effects and neurotoxicity, both in adults and during development, and immunotoxicity. Appropriate further testing should be undertaken where there is reason to suspect such an effect, based on structure, prior knowledge or alerts from the results of more conventional tests.

Specific studies on mechanism of toxicity or mode of action, particularly for end-points that may be used in establishing reference values, such as health-based guidance values, may provide useful data.

For all study designs, careful consideration needs to be given to:

- dose spacing and number of study groups;
- maximum dose utilized;
- number of animals in each group;
- choice of controls and whether there is a need for a positive control group;
- dosing regimen;
- confirmation of dose administered compared with nominal dose;
- dose ingested (e.g. palatability, wastage of food); and
- incidental disease, such as infection.

Increasingly, the utility of studies of precursor effects, long used to help in the risk assessment of non-genotoxic carcinogens, needs to be considered. Often, measurements reflecting such precursor effects are being developed as biomarkers. High-volume profiling techniques (e.g. metabonomics) are now being utilized in the search for novel biomarkers (USNRC, 2004).

When biomarkers have been used in toxicity studies, consideration should be given to their interpretation. The relevance of a biomarker to toxicological effects needs to be assessed critically. Biomarkers are of particular value in studies of mechanism and mode of action—for example, on the interspecies relevance of a mode of action. Biomarkers need to be adequately characterized and assessed for fitness for purpose (IPCS, 2001c; Gundert-Remy et al., 2005). This is especially true for data derived from studies using "omic" techniques (e.g. transcriptomics, proteomics, metabonomics). In addition to their application in biomarker discovery and development, these technologies are particularly useful in mechanistic toxicology (Heinje et al., 2005; Gatzidou et al., 2007). However, use of such data in risk assessment provides appreciable challenges, both in bioinformatics and in biological interpretation. The changes observed do not necessarily reflect an adverse effect, but may simply be a result of homeostatic regulation or adaptation. A number of these issues were discussed at an International Programme on Chemical Safety (IPCS) workshop in 2003 (IPCS, 2003).

The methods for statistical analysis should be addressed with care. The numbers of animals used per dose group will affect the power of the study, so both type I (false positive) and type II (false negative) errors need to be considered. Paired or two-sample comparisons are often undertaken, and the statistical test should apply a correction when multiple comparisons of non-independent data are analysed. A trend analysis may be helpful for dose-dependent effects. The power of the study to identify a measurable effect needs to be considered when large numbers of end-points are compared in a small number of animal groups. If isolated significant findings are identified, such as in a single clinical chemistry parameter, particular attention should be given to biological consistency with other observations in the database.

The study design should be adequate to determine the reference point selected for hazard characterization, such as the no-observed-adverse-effect level (NOAEL), benchmark dose (BMD) or other points of departure (see chapter 5). This includes adequacy of dose range and spacing, numbers of animals, variation within groups and nature of end-point measured.

4.1.5.6 Role of human studies

In general, data from humans are preferable to data from experimental animals, as they will have been obtained in the species of interest (see section 4.11). However, there are ethical and practical difficulties in obtaining such information. Administration to humans would be considered unethical if the safety of the substance is unknown and there has been no prior exposure of humans. In observational studies, there can be difficulties in obtaining adequate information on the extent of exposure.

Information from humans can arise in a number of different ways. These include:

- controlled studies in volunteers from whom informed consent has been obtained;
- studies of incidentally exposed subjects through epidemiological assessment;
- surveillance of occupationally exposed individuals;
- case-studies of subjects who have accidentally or deliberately consumed the substance (usually acutely);
- supervised trials of those substances where the level of human intake precludes the normal application of large uncertainty (safety) factors to data from animal studies (e.g. novel foods); and
- clinical trials on substances that also have potential use in human medicine.

Where the effect observed in animals is mild, acute and readily reversible, it may be possible to investigate this in healthy volunteers. Data obtained from such studies should be considered in risk assessment when the study is of a suitable design.

Surveillance-type studies, even when the data are inadequate for risk assessment, can provide a very useful reality check on the results obtained in experimental animals, often enabling a lower bound for any effect in humans to be established (using conservative assumptions for exposure assessment). Post-marketing surveillance data can be useful in supporting tolerability in humans, but should not be used as a justification for reduced premarketing safety assessment.

When the reference point used for hazard characterization, such as the NOAEL, cannot be derived from human data, it may be possible to compare kinetic data from animals with in vivo human data obtained at low doses or to incorporate in vitro human data into a PBTK model. Such information can be invaluable for interspecies comparison and for interpreting the results of studies in experimental animals.

Human tissues or preparations may also be studied in vitro; such information can provide useful insights into the relevance of effects for humans and interspecies extrapolation.

The design of studies in humans needs to consider:

- choice of doses;
- duration of administration (usually acute);
- number of subjects;
- sex of volunteers; and
- how representative the subjects are of the potentially exposed population; important variables include age, genetics, concurrent disease/drug treatment, diet and lifestyle factors, such as alcohol use and smoking.

In using human data, the adequacy of study design in addressing all possible subgroups in the population needs to be considered. For example, toxicokinetic studies in adult male volunteers may not be representative of females or the very young. Uncertainties in the interpretation and use of data from studies in humans can be allowed for by the application of appropriate uncertainty or adjustment factors (see chapter 5).

4.2 Absorption, distribution, metabolism and excretion (including residues of toxicological concern)

4.2.1 Introduction

The relationship between the external, or administered, dose of a substance and biological responses can be divided into two aspects:

- *toxicokinetics*, which relates to the delivery of the chemical to and its removal from the site of action as the parent substance and/or any active metabolites; and
- *toxicodynamics*, which relates to the interaction between the chemical and/or any active metabolites at the site of action and the final outcome or toxicological response.

Knowledge of the biological disposition of a chemical (i.e. its ADME) is a key part of any hazard characterization and risk assessment (Lipscomb & Ohanian, 2007; Renwick, 2008). Such information can be important for two main aspects of risk characterization:

- the design of appropriate animal studies for identifying and characterizing the hazards associated with exposure to the chemical; and
- the interpretation of the resulting data in relation to the mechanism or mode of toxicity, consideration of interspecies scaling and consideration of potential human variability.

Historically, the ADME of substances were studied by following the biological fate of the radiolabelled substance (usually ^{3}H-labelled or ^{14}C-labelled) using nonspecific techniques to measure total radioactivity, combined with separation methods, such as chromatography, to identify the radiolabelled constituents in the biological sample. In recent years, basic ADME studies have been supplemented by the generation of toxicokinetic data in which the concentrations of the chemical or its circulating active metabolites are measured in plasma and body tissues and used to provide a mathematical description of the concentration–time course of internal exposure (Renwick, 2008).

The term *toxicokinetics* describes the movement of a substance around the body and therefore relates to its absorption from the site of administration, its distribution from the general circulation into, and out of, body tissues and its elimination, usually by metabolism and excretion. It is clear from this that toxicokinetics should cover both radiolabelled ADME studies and plasma concentration–time curves. Some texts maintain a largely artificial distinction between metabolism and toxicokinetics, probably related to the nature of the studies used to develop the data.

The principles of toxicokinetic studies were outlined in Environmental Health Criteria (EHC) 57 (IPCS, 1986a); such studies basically provide a biochemical, physiological and mathematical description of the fate of the chemical in the body. In EHC 70 (IPCS, 1987), such information is under the heading "The use of metabolic and pharmacokinetic studies in safety assessment", whereas in EHC 104 (IPCS, 1990), it is under "Absorption, distribution, metabolism, and excretion". The term "pharmacokinetics" is sometimes used, because many of the mathematical approaches and models were developed for studies on therapeutic drugs in humans. In consequence, toxicokinetic studies are most readily applicable to single-chemical entities, whether an additive, pesticide, veterinary drug or contaminant. Limited data may be produced for mixtures, by the use of nonspecific techniques that detect all constituents in a mixture or chemical-specific analysis of principal components. Simple studies on digestibility and caloric value may be all that is practicable for novel foods or macroingredients (see chapter 9, section 9.2).

Guidance on the design of toxicokinetic studies has been developed for pharmaceutical agents by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) for both single-dose studies (ICH, 1994a) and repeated-dose investigations (ICH, 1994b). The guidance is broadly applicable to studies on single-chemical entities in food, such as additives and residues of pesticides and veterinary drugs, except that the possible impact of the food matrix on the rate and extent of absorption is of major potential importance.

The different components of ADME are outlined below, followed by discussion on the value of such data in the design and interpretation of toxicological studies.

4.2.2 Absorption

Absorption is the process by which the substance is transferred from the site of administration into the circulation. For chemicals in food, absorption usually refers to passage across the gut wall into the circulation, although for some chemicals, uptake may be only as far as the epithelium of the gastrointestinal tract. Absorption may be as the parent compound or as metabolites formed within the lumen or the wall of the gastrointestinal tract. Because the term *absorption* does not define the nature of the absorbed material, it can give rise to confusion; for example, a substance might be completely absorbed from the gut, but with none of the parent compound detectable in the blood or tissues. To allow for this possibility, the pharmacokinetic term *bioavailability* is used to describe the fraction or percentage of the administered dose that enters the general circulation as the parent compound (Duffus & Worth, 2006). The term bioavailability is one of the most misused toxicokinetic terms (see Duffus & Worth [2006] for alternative and less specific definitions).

The main routes by which humans are exposed to chemicals are via ingestion in food or drinking-water, inhalation and across skin, with the last two being of relevance to occupational exposure to pesticides. These data may be useful for route-to-route extrapolation (see section 4.2.9).

The most important process involved in the transfer of foreign chemicals from the site of administration into the general circulation is passive diffusion down a concentration gradient. For each of the main routes of administration, the substance has to cross cell membranes before it enters the general circulation. In consequence, low molecular weight, lipid-soluble molecules are absorbed more rapidly and to a greater extent than highly water-soluble or larger molecules. Highly lipid-soluble substances, such as paraffin waxes, β-carotene and polyhalogenated dibenzodioxins, show incomplete absorption from the gut because they do not form a molecular solution in the gut lumen. Diffusion across the gastrointestinal wall is usually rapid for lipid-soluble molecules, because of the large surface area of the small intestine, but there may be a delay because of physiological processes such as gastric emptying. Diffusion of volatile substances across the airways may be extremely rapid, especially if the substance is delivered to the finer airways and alveoli. Absorption across the dermis is usually extremely slow and limited to lipid-soluble molecules only.

Although active transport processes are important in the absorption of nutrients from the gastrointestinal tract, they are highly specific to the normal nutrient substrate of the carrier protein; very few foreign chemicals are substrates for any of the physiological transporters in the gastrointestinal tract. An exception to this generalization is the efflux transporter known as P-glycoprotein, which transports a wide range of low molecular weight organic foreign molecules from the cytosol of enterocytes into the gut lumen. This efflux transporter may limit the absorption of some foreign compounds and can be a source of non-linear kinetics at high dietary concentrations (see below).

Information on absorption may relate to the rate at which the chemical is transferred into the general circulation or to the extent to which the administered dose enters the circulation or is excreted in urine, either as the administered substance or as its metabolites:

- The *rate of absorption* can be determined by serial measurements of the concentrations of the substance, or its metabolites, in plasma or their excretion in urine, as part of a toxicokinetic study. The absorption rate constant can be determined from the increase in plasma concentrations following the administration of a single dose by the appropriate route. The rate of absorption from the gastrointestinal tract and lungs is usually rapid and first order (i.e. the rate of absorption is proportional to the concentration available for absorption). The absorption rate is most likely to be important in relation to acute toxic effects and the establishment of short-term guidance values such as the ARfD. The rate of absorption across the skin tends to be slow and may result in low, but relatively constant, plasma concentrations.

- The *extent of absorption* is important for both acute and chronic toxicity. The extent of absorption may be estimated in two ways. The extent of total absorption following the administration of a radioactive dose can be estimated from the urinary excretion of the radiolabel after oral and intravenous administration. (The use of an intravenous dose allows correction for any compound in the general circulation that may be eliminated by other routes, such as biliary excretion or exhalation. Such information can also be obtained by bile duct cannulation and trapping of expired air.) Such data usually relate to the combined excretion of the administered substance and its metabolites in urine and would not indicate the extent of any metabolism that may occur prior to the substance reaching the general circulation (i.e. first-pass or presystemic metabolism). The extent of absorption as the parent compound (i.e. bioavailability) may also be determined from chemical-specific measurements of

> the compound in either the general circulation or urine following both oral and intravenous administration. (The use of an intravenous dose is essential, as it provides reference data corresponding to 100% "absorption" into the general circulation.)

The term bioavailability has a strict meaning and definition in pharmacokinetic terms, and its nonspecific use in other contexts can lead to confusion and misunderstanding. For food additives, contaminants and pesticide residues, the term is used in the toxicokinetic sense given above. For veterinary drug residues in food, it is used to reflect the fraction that can be released from the food matrix and is available for absorption, but this is only one of the factors that can determine the true bioavailability of the residue to the general circulation. Confusion can also arise when the calculated bioavailability is compared with the results from studies measuring the urinary excretion of radioactivity following an oral dose; for example, 100% of a radioactive dose may be eliminated in the urine, but the bioavailability would be only 10% if the substance undergoes 90% first-pass metabolism in the gut or liver prior to entering the general circulation.

The extent of absorption is of particular importance when the substance undergoes extensive first-pass metabolism or is only poorly absorbed from the gastrointestinal tract or site of administration, such that the bioavailability and the extent of absorption, as the parent compound plus metabolites, are low. Under such circumstances, the absorption process may be the source of wide differences between species or between different human individuals, adding greater uncertainty to the hazard characterization process. The bioavailability of a chemical can be affected considerably by the experimental conditions (e.g. diet versus gavage) and the vehicle used for gavage doses. Saturation of presystemic metabolism in the gut or liver at high oral doses results in a non-linear relationship between internal concentrations of the parent compound and the external dose.

4.2.3 Distribution

Distribution is the process by which the substance or its metabolites present in the general circulation move around the body and partition into and out of different body tissues.

Transfer from the general circulation into tissues is primarily by passive diffusion of the chemical down a concentration gradient. In consequence, tissue levels increase as the plasma concentrations rise during the absorption of the substance, and tissue concentrations fall when the plasma concentration decreases during the elimination of the substance from the body. Transfer from the general circulation into tissue cells requires that the substance cross the cell membrane, and again this occurs more rapidly for lipid-soluble molecules than for highly polar or larger molecules.

The entry of molecules into some organs, especially the brain, is largely limited to lipid-soluble molecules, because there are tight junctions between adjacent endothelial cells that prevent water-soluble molecules from leaving the lumen of the blood vessels. The small size of membrane pores in the endothelial cell membrane and the presence of active transporters, including P-glycoprotein, also contribute to the so-called “blood–brain barrier”. Active transporters in endothelial cells supplying the brain are important in the delivery of essential nutrients, such as glucose and amino acids, but, again, they are not available to the vast majority of non-nutrient chemicals.

The vasculature of certain organs, such as the liver, kidneys and brain, contains transporters that can either actively take up the chemical from the circulation or transport chemicals from the tissues back into blood. Tissue efflux transporters, such as P-glycoprotein and multidrug resistance associated protein (MRP), have low specificity and can be induced by chronic exposure to some substrates, which can affect tissue distribution on repeated administration. Membrane transporters can show species differences, sex differences and genetic polymorphisms. The toxicity of the pesticide abamectin shows wide differences between strains of mice, which can be related to the lower activity of P-glycoprotein in the gut wall and blood–brain barrier in the more sensitive strains (FAO/WHO, 1998).

As for absorption, distribution may be thought of in terms of the rate of the process and its extent—i.e. what proportion of the body burden of the substance moves out of the general circulation into body tissues:

- The *rate of distribution* is largely dependent on the rate of perfusion of those organs that show the highest affinity for the substance.

For example, if the substance is very lipid soluble, there will be a much higher concentration in adipose tissue than in the plasma, and therefore the rate at which the substance can enter adipose tissue is limited by the low perfusion rate of this tissue. The rate of distribution is usually determined by toxicokinetic measurements following an intravenous bolus dose.

- The *extent of distribution* is determined by the relative affinity of the circulation and of the organs of the body. Substances may dissolve in lipoproteins or cell membranes present in the general circulation, as well as intracellular and extracellular membranes within the tissues. In addition, many substances show reversible binding to plasma and tissue proteins. In consequence, the ratio of the concentration of the substance in the tissue to that in the plasma depends on the overall affinity of the tissue compared with plasma and may be extremely high in some organ systems; for example, lipid-soluble substances may show very high adipose tissue to plasma ratios.

The extent of distribution may be measured both using nonspecific radiochemical methods and from chemical-specific analyses. The former will provide information on the pattern of distribution of the parent compound plus its metabolites, but may also represent material that is covalently bound to tissue proteins, ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) (which is really an elimination process in relation to the parent compound). Consideration needs to be given to the position and chemical stability of the radioisotope within the molecule, as misleading data on tissue distribution could be obtained if the label were labile and entered general intermediary metabolism—for example, as tritiated water or a ^{14}C-labelled methyl residue. Chemical-specific analysis of the concentrations of parent compound in plasma and tissues can be used to indicate the pattern of distribution. Data from the plasma concentration–time curve following a single intravenous bolus dose can be analysed to determine the apparent volume of distribution, which reflects the ratio between the total body burden and the plasma concentration; this parameter can also be calculated from studies in humans. For highly lipid-soluble substances, such as polyhalogenated dibenzodioxins, the relationship between the total body burden and the concentrations present in adipose and other tissues depends on body composition and the percentage of body fat, which can vary between species and also between individuals (USNRC, 2006).

4.2.4 Metabolism

Metabolism (biotransformation) is the process by which the administered substance is changed structurally into molecules that are eliminated from the body.

Although metabolism is often thought of as representing a detoxification process, in many cases target organ toxicity can arise from the actions of a metabolite rather than those of the parent compound. In some cases, the metabolite may be so unstable that it interacts covalently with tissue proteins, RNA or DNA to produce cellular changes that are part of the mode of action of the toxic effect. In such cases, metabolism of the substance becomes an important part of the mode of action and may be a major source of species differences and human variability in sensitivity to the chemical.

It is important that toxicokinetic measurements used for hazard characterization relate to the active chemical entity in the circulation or tissue. Depending on the biological activity of the parent compound and its metabolites, toxicokinetic measurements based on the parent compound may not provide an adequate basis for consideration of species differences or human variability.

PBTK models (see below) can incorporate data on enzyme kinetics as part of the overall elimination process (Krishnan & Andersen, 2007). Some PBTK models also include local target organ metabolism, thereby providing a particularly powerful method for predicting the target organ dose of the active chemical entity in the experimental animals and predicting equivalent target organ doses in humans.

Although some food additives are metabolized by the enzymes of normal intermediary metabolism, the majority of additives, pesticides and veterinary drugs are low molecular weight, “foreign” organic molecules, and these are metabolized by a variety of phase I and phase II “drug-metabolizing” enzymes that are present largely in the liver. Phase I metabolism involves the oxidation, reduction or hydrolysis of the molecule with the introduction of groups suitable for subsequent phase II or conjugation reactions. Phase II reactions involve the conjugation of the foreign compound, or its phase I metabolite, with a molecule such as glucuronic acid or sulfate; this serves to mask potential active functional groups and generally leads to an increase in water solubility (Kemper et al., 2007).

Both phase I and phase II metabolic reactions usually lead to a decrease in toxicity and the generation of excretable products; however, they may also lead to the generation of reactive chemical species that are important in the toxicity of the molecule. In consequence, studies of metabolism should aim to define the processes involved in the elimination of the parent compound and any toxicity associated with that molecule, as well as the generation of any active chemical products of the substance and their subsequent detoxification and elimination from the body.

Consideration should be given to factors that might affect metabolism during the conduct of toxicity tests. These include strain and species differences, sex differences, route dependency, dose dependency (e.g. saturation, competing pathways with different kinetic parameters), time dependency (e.g. induction, inhibition) and concurrent pathology. The extent to which such differences can be extrapolated to humans should be evaluated; for example, many sex differences in metabolism observed in rats do not occur in humans. The enzymes involved in the metabolism of foreign compounds represent the most important source of interspecies differences and human variability in the biodisposition of the compound and, for many cases, in the generation of toxic effects.

At low substrate concentrations, the rate of metabolism is proportional to the substrate concentration, which means that toxicokinetic parameters, such as clearance and half-life (see below), are constant and independent of dose level. However, the amounts of metabolizing enzymes in the body are limited, and saturation of metabolism can occur at high dose levels; saturation of metabolism results in slower elimination at higher doses and a disproportionately increased body burden with increase in dose level during repeated dosing. Saturation of metabolism is not always a feature of toxicity studies, because adverse effects are often found at doses that do not saturate metabolism; however, saturation that occurs over the dose range used for toxicity studies complicates analysis of the dose–response data and their extrapolation to humans.

Metabolism is only one possible route of elimination from the body, and the measured rate of elimination from the body—for example, the plasma half-life—is the sum of all elimination processes.

4.2.5 Excretion

Excretion describes the processes involved in the elimination of the substance or its metabolites from the general circulation into a biological waste product, such as urine, faeces or exhaled air.

The urine is the major route of elimination of low molecular weight foreign compounds from the body. However, it is efficient only for low molecular weight, highly water-soluble molecules, because lipid-soluble molecules will be reabsorbed from the renal tubule and re-enter the general circulation. It is for this reason that low molecular weight, lipid-soluble molecules tend to be retained in the body and undergo metabolism prior to their excretion. The rate of renal excretion of a compound may be very high if it is a substrate for the various anionic or cationic carriers that transport molecules from the general circulation into the lumen of the renal tubule, but may be very slow for compounds that are highly bound to plasma proteins. There are a number of different transporters for organic anions (organic anion transporters, or OAT, transporters for acids), organic cations (organic cation transporters, or OCT, transporters for bases), peptide transporters and nonspecific transporters (members of the MRP family). These may occur on either the basolateral or apical membranes of the renal tubule or both, are important in extracting chemicals from blood and transferring them into the tubule lumen, and show species and sex differences (Lee & Kim, 2004). In addition, compounds filtered at the glomerulus may undergo pH-dependent passive reabsorption from the renal tubule back into the general circulation.

Another important route of elimination is via the bile, where the molecule is incorporated into the micellar constituents of bile and passes into the lumen of the gastrointestinal tract. Biliary excretion can also involve a number of efflux transporters, such as P-glycoprotein and MRP. Although the excretion effectively removes the compound from the general circulation, it is possible that the metabolites eliminated in bile may be further metabolized within the lumen of the gastrointestinal tract and reabsorbed. For example, the glucuronic acid conjugate of a compound may be formed in the liver, eliminated in bile and hydrolysed back to the original compound in the gut lumen; the compound is then absorbed from the lower bowel to re-enter the general circulation. Such a process is known as enterohepatic circulation.

Compounds eliminated in the exhaled air are usually of low molecular weight and volatile or are fragments of larger administered substances that possess these characteristics.

4.2.6 Overall elimination from the body

The overall rate of elimination of a chemical from the body, which can be measured from the decrease in plasma concentration with time, reflects the sum of all the processes contributing to the elimination of that chemical—i.e. metabolism plus renal excretion plus biliary excretion plus exhalation plus any other minor routes of elimination.

Because physiological and metabolic processes are first order with respect to substrate at low concentrations, decreases in plasma concentrations with time are usually exponential in nature and can be defined by measurement of the appropriate elimination rate constant or its associated half-life. The rate of elimination and half-life are important parameters, as they indicate the duration of exposure of the body and its tissues to the substance, and they also indicate the potential for accumulation on repeated dosing.

Again, it is important to recognize the difference that may be obtained from measurements based on total radioactivity (parent compound plus metabolites) and chemical-specific assays that will measure separately the parent compound and characterized metabolites. A major advantage of nonspecific methods such as the use of radioisotopically labelled substrates is the ability to measure all metabolic products, including those that have not been characterized. However, such information could be misleading if the measured half-life reflected that of an inactive and non-toxic metabolite and therefore was not related to the body burden or the accumulation of the toxic moiety. The same criticism would apply if a chemical-specific method were applied to an inactive moiety. In cases where the active chemical entity is produced within the target organ and does not enter the circulation, the plasma toxicokinetics should relate to the circulating precursor molecule (usually the parent compound).

4.2.7 The role of toxicokinetic studies in the design of animal toxicity tests

ADME and toxicokinetic studies are important in selection of the appropriate test species and the dosing regimen. There are major species differences in the routes and rates of elimination of test substances in different animal species compared with humans. Quantitative differences between the species used in toxicity studies and humans are an almost inevitable part of hazard characterization.

Although it is frequently suggested that the animal species used in toxicity studies should be as metabolically similar to humans as possible, in reality only a few species are used in toxicity tests. This is because of the need for background knowledge of the animal's histopathology and physiology combined with practical aspects, such as size, housing conditions and longevity. In consequence, despite known differences compared with humans in the rates and extents of metabolism and excretion, most studies are performed in a relatively small number of test species. Under these circumstances, knowledge of the qualitative and quantitative nature of any differences between the test species and humans can be a very important part of hazard characterization.

Although the primary aims of dose selection are to identify hazards and to define their dose–response characteristics, toxicokinetic information can help to inform this process. The biological processes outlined above are essentially first order at low concentrations, but the exaggerated dosages used in animal toxicity studies for the identification and characterization of hazards may lead to saturation of transporters or metabolic enzyme systems, such that the relationship between dose and target organ exposure to the parent compound or its metabolites is not a simple linear relationship. Saturation of metabolism may lead to lower than predicted concentrations of the metabolites formed by the metabolic pathway that is saturated, but higher than predicted concentrations of the parent compound and other metabolites. The toxicological consequences of this may be a non-linear dose–response relationship with exaggerated toxicity at high, saturating doses, if the parent compound is the active toxicant, but reduced toxicity at high doses, if the product of the saturated enzyme is the primary toxicant. Specifically designed toxicokinetic studies can provide the key to interpreting dose–response relationships derived from toxicity studies.

4.2.8 The role of toxicokinetic studies in the interpretation of data from animal toxicity studies

Toxicokinetic studies are designed to produce information on the profile of exposure to the active chemical entity at the site of toxicity under the conditions that produce the toxicity and that are the basis for determining the NOAEL and hazard characterization. Important toxicokinetic data relate to:

- the internal dose in animals based on plasma, serum or blood concentrations of the parent compound or its active metabolites; the most commonly made measurements are the area under the concentration–time curve (AUC), the observed peak concentration (C_{max}) and the time of the peak concentration (T_{max});
- the relationship between the external dose given to animals and the internal dose (as indicated by the AUC for plasma or tissue);
- the relationship between the plasma or blood concentrations (AUC or C_{max}) and those at the site of toxicity; and
- information on appropriate plasma or blood concentrations after the administration of tracer doses to human volunteers in order to allow extrapolation of animal data to humans.

Data on the AUC and C_{max} of the parent compound in blood or plasma derived from specifically designed, single-dose toxicokinetic studies (ICH, 1994a) can be used to calculate related toxicokinetic parameters that describe the basic handling of the substance in the body. These parameters can then be used to predict the fate of the substance on repeated dosage and assist in interspecies extrapolation (Renwick, 2008). Important toxicokinetic parameters are:

- *Clearance (CL)*: the volume of blood or plasma cleared of the substance per unit time; units are volume per unit time (e.g. ml/min or ml/min per kilogram body weight); value is dependent on the in vivo functional capacity of the organs of elimination, which may be limited by organ blood flow or tissue activity; calculated as [AUC/intravenous dose].

- *Apparent volume of distribution (V)*: the volume of blood or plasma in which the body burden appears to be dissolved; units are volume (e.g. ml or ml/kg or l/kg); value is dependent on the extent of distribution from the general circulation into tissues, which is affected by protein binding, the lipid solubility of the compound and body composition; calculated as [intravenous bolus dose/C_{max}], but other more robust methods are normally used in practice (Renwick, 2008).

- *Elimination half-life* ($t_{½}$): the time taken for the post-peak blood or plasma concentration to halve; units are time (e.g. min or h); value is dependent on CL and *V*, which are independent physiologically related variables; calculated from regression analysis of the concentration–time course data or as [0.693*V*/CL].

- *Bioavailability (*F*)*: the fraction (or percentage) of the administered dose that reaches the general circulation as the parent compound; a unitless fraction; for oral doses, the value is dependent on the extent of transfer from the gut lumen and any presystemic metabolism in the gut lumen, gut wall and liver; calculated as [AUC_{oral} × $dose_{iv}$ / AUC_{iv} × $dose_{oral}$] or [AUC_{oral}/AUC_{iv}] when the same dose levels are given by each route (oral and intravenous, or iv).

Each of the above parameters is independent of concentration at doses that do not saturate the enzyme systems or transporters involved in the biological fate of the compound. Non-linear kinetics may also arise from physicochemical non-linearity, such as the saturation of solubilization at the site of administration. Dependent on the nature of the plasma or blood concentration–time curve, a compartmental model containing one, two or more exponential terms may be fitted to the data.

Quantification of systemic exposure or body burden in the test species during the performance of toxicity studies provides important information that can assist in the interpretation of similarities and differences in toxicity across species, dose groups and sexes (ICH, 1994a). Suitable data may sometimes be obtained from all animals on a toxicity study or from representative subgroups, but because of the invasive nature of toxicokinetic methods, data are usually obtained from specially established satellite groups or from separate studies.

ADME studies based on the elimination of radioactive compound and metabolite after a single oral dose may be useful in defining the extent of species differences and of saturation of metabolic pathways in the biodisposition of the compound in the test species. When comparable data are available from studies in humans, these can be used to define the adequacy of the test species as a model for humans, providing that the biological consequences of metabolism (i.e. detoxification or bioactivation) have been characterized. In some cases, data are available for small numbers of human subjects given a single oral dose of the radiolabelled substance, and such information can be very informative.

Of greater potential value are data relating to the circulating concentrations of the parent compound and any active metabolites in the test species under the experimental conditions giving rise to the hazard that will be the basis for hazard and risk characterization. Suitable toxicokinetic data from studies in experimental animals and humans can reduce the uncertainties associated with interspecies extrapolation and also give insights into the potential human interindividual variability.

When the toxicity database on a substance is to be used to estimate a health-based guidance value, such as an acceptable daily intake (ADI), the most relevant toxicokinetic data are for the test species under the experimental conditions giving the NOAEL for the critical effect and matching information for humans at the projected ADI or health-based guidance value. Although there are ethical considerations with respect to the intentional administration of non-therapeutic agents to humans, it is difficult to envisage objections to intentional exposures to doses of food additives or pesticides that would represent the ADI for unintentional exposure in the absence of any such study in humans.

In vitro data can provide extremely important information relating to the enzymes involved in the metabolic detoxification or activation of the substance. Definition of the enzyme kinetics of the major pathways in organs taken from the test animal species and from humans can be particularly valuable in defining species differences and in the development of PBTK models that characterize species differences. Unlike the basic toxicokinetic parameters given above, PBTK models can provide data on the concentrations in potential target organs and describe how they change with time and with repeated dosage. In some cases, such models can be extended to include local tissue bioactivation and detoxification processes within the target organ for toxicity and therefore provide insights that are not possible from in vivo pharmacokinetic measurements. In principle, PBTK models could be used to predict human variability in target organ doses, providing there were sufficient data on human variability in the key parts of the PBTK model, such as organ blood flows and enzyme kinetics.

In addition to the development of PBTK models, in vitro studies using livers with characterized expression patterns for different isoenzymes can be useful in identifying the isoenzymes responsible for different metabolic processes; similar information can also be obtained from in vitro enzyme expression systems. Such information may be particularly valuable in predicting the likely human variability in metabolism of the substance.

A major uncertainty associated with most forms of hazard characterization arises from the relatively limited number of data available from studies in humans and the inadequacy of such data to define the extent of human variability in biodisposition. Information on human variability is rarely available from studies using radioactive substrates; more extensive information may be available in some cases where chemical-specific assays have been used to describe the toxicokinetics following administration of low doses of the unlabelled substance.

Knowledge and understanding of the major pathways involved in the detoxification and any bioactivation of the substance can be used to predict likely human variability in the biodisposition of the substance based on known human variability for substrates that are metabolized by the relevant pathways. For example, a substrate metabolized extensively by an enzyme exhibiting genetic polymorphism would show considerably more interindividual variability within the human population than would a substrate eliminated primarily unchanged via renal excretion. Such potential human variability in toxicokinetics needs to be considered as part of hazard characterization and assessment of the adequacy of the default uncertainty factors.

Parameters, such as bioavailability, clearance and half-life, derived from a single-dose toxicokinetic study can be used to predict the concentrations in plasma or blood following chronic administration, providing that repeated dosage does not alter the biovailability, clearance or distribution. The body burden during chronic administration is called the "steady-state body burden". The term "steady state" relates to the condition during repeated dosing in which the daily dose of a substance is eliminated from the body within 24 h (i.e. there is no overall change in the average body burden of the substance). However, this term should not be confused with a constant unvarying plasma concentration and body burden. For substances that are rapidly absorbed and eliminated from the body, there will be significant peak and trough concentrations between each dose. Peaks and troughs are most apparent when a substance with a short half-life is given as a single daily bolus gavage dose; in contrast, when such a substance is incorporated into the diet, the plasma and tissue concentrations of the substance will reflect the diurnal pattern of food intake. For substances with long half-lives, such as the dioxins and other chlorinated hydrocarbons, there will be significant accumulation during repeated dosage. The daily pattern of dose input will represent a small fraction of the total body burden or plasma concentrations at steady state, and there will be little diurnal variation, so that the "steady-state" condition will actually be represented by relatively constant plasma and tissue levels.

Problems of accumulation on repeated dosing and saturation of elimination are particularly pertinent to high-dose animal toxicity studies, and information on these areas can be obtained readily from suitably designed in vivo toxicokinetic studies.

During repeated dosing, the average or steady-state plasma concentration is determined by the rate of dose administration and the systemic clearance and bioavailability of the substance, parameters that are readily determined from a single oral dose. Therefore, single-dose toxicokinetic studies can be used to predict the average steady-state plasma concentration and body burden. Similarly, single-dose tissue distribution data can be used to predict steady-state tissue concentrations based on the plasma concentration at steady state and the single-dose tissue to plasma ratios.

Inherent in the use of single-dose data for predictions about steady-state conditions is the assumption that repeated dosing does not alter either the bioavailability or the clearance of a substance. Although this is a reasonable assumption in the majority of cases, the bioavailability and clearance can be altered by prior treatment for substances that are either inducers or inhibitors of their own metabolism. Under these circumstances, the single-dose data would either overpredict or underpredict, respectively, the steady-state plasma and tissue concentrations of the parent compound. In addition, substances that produce adverse effects on the liver or kidneys may affect the elimination of the substance itself during repeated administration at doses that give rise to such toxic effects. Comparison of the plasma toxicokinetics of a substance following a single oral dose given as gavage with the concentration–time profile for a dose interval at steady state (e.g. over a 24 h period) can give useful insights in relation to the possible influence of repeated dosage on both the absorption and elimination of the substance.

Single-dose toxicokinetic studies in experimental animals can be important for route-to-route extrapolation (see section 4.2.9). Data following treatment with gavage doses, incorporation of the compound into the diet and other routes of administration that are relevant to the hazard characterization can be used in the interpretation of hazard characterization data that were generated using routes or vehicles that are not of direct relevance to human exposure.

It is important that the life stage investigated in toxicokinetic studies is the same as that which becomes the focus for hazard and risk characterization. Absorption and elimination processes vary during the life of both experimental animals and humans; they are immature in the neonatal period, but then increase rapidly to adult levels, followed by a slow decline as the organism ages. In consequence, an apparently constant dosage regimen expressed in milligrams per kilogram body weight may be associated with elevated plasma and tissue concentrations during the later phases of a chronic bioassay. At the period when toxicokinetic processes are most immature (i.e. the neonate), the principal route of exposure is via maternal milk, and this may be of particular significance for neonatal exposure to lipid-soluble substances. Transfer of chemicals into milk may be an important measurement component of the exposure profile of animals during reproductive toxicity and two-generation carcinogenicity studies.

Both health-based guidance values and the starting points for their determination, such as the NOAEL (see chapter 5), are expressed on a body weight basis (e.g. mg/kg body weight per day), with an uncertainty factor used to allow for possible species differences and human variability in both toxicokinetics and toxicodynamics. The clearance of foreign compounds is usually greater in rodent species than in humans on a body weight basis, and this difference in toxicokinetics is an important reason for the application of an interspecies uncertainty factor. Many physiological and metabolic characteristics relate more closely to body surface area or body weight$^{0.7}$ (Rodricks et al., 2007). The use of surface area for interspecies scaling to convert the NOAEL into an ADI would reduce the need for an interspecies uncertainty factor. Such an approach would be most valid for compounds that are metabolized by normal intermediary metabolism, but would be less valid for compounds eliminated by phase I and phase II foreign compound metabolizing enzymes, because these show wide species differences that do not scale closely with body surface area. In contrast, the use of body weight$^{1.0}$ is more conservative than the use of body weight$^{0.7}$ when considering the kinetics in children compared with adults, because children show greater elimination capacity on a simple body weight basis, and therefore their internal dose would be lower than in an adult given the same external dose expressed as milligrams per kilogram body weight.

4.2.9 Route-to-route extrapolation

The target site dose is the ultimate determinant of risk. Substances that do not establish an internal dose by a given route would not be presumed to produce internal toxicity by that route. Conversely, substances that cause internal toxicity by one route of exposure would be assumed to do so by any other route that also produces a comparable internal dose of the active chemical entity at the target tissue. The differences in biological processes between different routes of exposure (oral, inhalation, dermal, intravenous) can be great. In oral studies, even the mode of administration (gavage versus diet versus drinking-water) may be an issue for extrapolation within the same route.

If the route for the kinetic studies in either animals or humans varies from that on which the critical effect level is based, then route-to-route extrapolation may be necessary, and the data will need to be assessed critically on a case-by-case basis (Pepelko, 1987), including for use for the development of a CSAF. Toxicokinetics in general, and PBTK modelling in particular, are useful for quantifying route-to-route extrapolations, including using a combination of existing data and modelling approaches.

4.3 General systemic toxicity

4.3.1 Introduction

Tests of general systemic toxicity are conducted to identify target organs for toxicity and to confirm or mitigate the need for additional or more specific testing. Principles that are common to tests for general systemic toxicity, utilizing repeated-dose protocols, are described in this section. To a large extent, the designs of toxicity studies have been standardized, and common parameters are evaluated at different time points in studies of different durations. Standardized toxicity testing guidelines have been produced by the OECD (see http://masetto.sourceoecd.org/vl=2781582/cl=14/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/v1n4/contp1-1.htm) for:

- Repeated Dose 28-Day Oral Toxicity Study in Rodents (Test Guideline No. 407; OECD, 1995a) (updated for endocrine effects, adopted in 2008; OECD, 2008);
- Repeated Dose 90-Day Oral Toxicity Study in Rodents (Test Guideline No. 408; OECD, 1998a);
- Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents (Test Guideline No. 409; OECD, 1998b);
- Chronic Toxicity Studies (Test Guideline No. 452; OECD, 1981b); and
- Combined Chronic Toxicity/Carcinogenicity Studies (Test Guideline No. 453; OECD, 1981c).

Additional information is available in United States Environmental Protection Agency (USEPA) test guidelines (USEPA, 1998d,e,f, 2000; see also http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/index.html), in the United States Food and Drug Administration (USFDA) Redbook 2000 (USFDA, 2000) and in Jacobson-Kram & Keller (2006).

Tests of general systemic toxicity assess the effects of a test substance on a wide range of end-points indicative of toxicity, including observational, functional, biochemical and pathological end-points.

The goal of such tests is to determine which organs are affected by the test substance and how they are affected. Testing is done in a manner that best relates to human exposure scenarios; for substances present in foods, administration of the substance in repeated-dose animal studies is usually via the diet, by gavage or via drinking-water.

Reproductive or developmental toxicity, neurotoxicity and immunotoxicity are not assessed adequately in tests of general systemic toxicity. There is more information on tests for these forms of toxicity in sections 4.7, 4.8 and 4.9 (reproductive and developmental toxicity, neurotoxicity and immunotoxicity, respectively).

4.3.2 Tests for general systemic toxicity

Tests for general systemic toxicity are multidose studies of various durations. Ideally, the dose levels are selected such that toxic effects, but not death or severe suffering, are produced at the highest dose level, with lower dose levels producing graded responses and no adverse effects observed at the lowest dose level (NOAEL). Dose selection may be based on prior knowledge, but often a range-finding study may be necessary to define the doses to be used in the toxicity studies. Data from studies of shorter duration are normally used in the selection of dose levels for long-term or chronic studies. All studies should include a control group of animals; the handling of controls should be identical to that of the treated animals, including the administration of the dosing vehicle if relevant.

Whereas conventional acute toxicity studies (section 4.4) are conducted to determine a single maximally tolerated or lethal dose, tests for general systemic toxicity are conducted using repeated dosing over various periods of time, from days to years. In general, studies are conducted for 14–28 days, 13 weeks, 52 weeks or longer. Two-year carcinogenicity studies in rats are often combined with a 1-year study of toxicity by including satellite groups for toxicological evaluations. The terms subacute (14–28 days), subchronic (13 weeks) and chronic (52 weeks) are used to describe tests of general systemic toxicity, but these designations are not precisely defined; tests of shorter or longer duration (e.g. 7 days, 26 weeks or 2 years) are also common. The terms used are less important than understanding that the objective is to test for a defined proportion of an animal's lifespan.

4.3.3 Testing strategies

Studies of variable duration are typically conducted in sequence, with shorter-duration studies conducted before studies of longer duration. In this way, information gained early on in testing can be used to determine appropriate methods and doses or to otherwise optimize study designs for subsequent tests of longer duration or to evaluate specific end-points (e.g. immunotoxicity or neurotoxicity studies).

The type and amount of data needed to evaluate various substances should be determined on a case-by-case basis, so testing strategies will vary from substance to substance. Knowledge of the anticipated human exposure to and chemical structure of the substance will help in the design of an appropriate testing strategy.

4.3.4 Study design and data interpretation

4.3.4.1 Good Laboratory Practice

Non-clinical laboratory studies should be conducted according to the principles of GLP (see http://www.oecd.org/document/63/0,3343,en_2649_34381_2346175_1_1_1_1,00.html) and related national regulations, or similar guidelines. These cover the care, maintenance and housing of experimental animals as well as other general study considerations, such as resources, protocols and written procedures, characterization of test items and test systems, documentation and quality assurance. The use of GLP helps to ensure that studies are conducted appropriately and that the results can be used with confidence for risk assessment purposes. Studies not conducted to GLP or similar standards can provide valuable data (e.g. related to mode of action) and should not be ruled out for consideration when setting health-based guidance values.

4.3.4.2 Test substance

The test substance should be thoroughly characterized with respect to chemical identity, purity, stability and other properties, such as pH or solubility. For commercial substances, such as additives, pesticides and veterinary drugs, the substance tested should be the (intended) article of commerce. If the article of commerce is not the test substance, its relationship to the test substance must be accurately described.

The effect of a vehicle or other formulation aids on the test substance should also be considered; for example, they may affect the rate or extent of absorption from the gastrointestinal tract. Use of a single lot of test substance throughout a study will help to minimize inconsistent results due to differences in composition or levels of contaminants between batches, but relevant stability data on the test substance are then necessary to ensure consistency of the material dosed throughout the study.

4.3.4.3 *Species, number and sex*

General systemic toxicity studies are typically conducted in two species, a rodent and a non-rodent species or two rodent species, to maximize the opportunity to find an effect (hazard identification). The animals most often tested are rats and dogs, but other species may be used. Pigs, for example, may be the animal of choice for testing a fatty substance, because the metabolism of fat in pigs most closely approximates fat metabolism in humans. When other species are used, existing protocols may need to be modified to account for the unique characteristics of the selected test species. It is essential that all protocol modifications are reported so that the results can be properly interpreted.

Both sexes should be tested. Equal numbers of males and females of each species and strain should be tested to allow for an evaluation of potential hormonal influences, differences in metabolism or other sex differences. The animal's sensitivity in relation to the nature of the toxicity of the test substance needs to be considered in both designing and interpreting a study.

Longevity has become an issue for some strains of rats, with rates of survival so low that data collection from and interpretation of long-term studies are compromised. The anticipated survival of the animals should help influence the number of animals entered into a study so that there are enough animals available at termination to provide meaningful study results. In general, more animals are tested as the duration of the study increases. For a 13-week study, a minimum of 20 rodents per sex per group or at least 4 dogs per sex per group are common recommendations. Fewer animals may be included in a range-finding study, whereas more animals may be included if interim necropsies are planned.

Animals should be randomly assigned to control and treated groups to help minimize bias and assure comparability of pertinent variables across groups. As an example, mean body weights and body weight ranges should not differ substantially across groups at the start of an experiment if group data are to be evaluated. In some situations, additional control groups are useful—for example, when dietary imbalances are suspected (e.g. the highest dose causes significant caloric dilution).

4.3.4.4 Dose selection

The dose selection should take into account the anticipated human exposure, the frequency of exposure and the duration of exposure. Dose selection for toxicity studies should also be based on information known about the test substance and any prior results of toxicity tests. In general, responses require higher doses in studies of shorter duration than in long-term studies; in shorter studies, higher doses may be tolerated.

Three to five dose levels of the test substance and a concurrent control group are ordinarily sufficient to be able to relate toxicity to level of exposure. As a primary aim of any study is to define the quantitative relationship between exposure and effect (i.e. the dose–response; see chapter 5), more doses instead of fewer are generally desired. At a minimum, three dose levels of the test substance and a concurrent control group should be used in tests of general systemic toxicity. The dose range selected should allow for the expression of toxicity at the highest dose (e.g. 10% reduction in body weight) and no toxicity at the lowest dose tested; intermediate toxicity would be expected at intermediate doses (e.g. 5% reduction in body weight). For essentially non-toxic substances, the top dose studied may be set by an accepted limit dose, such as 5% addition to the diet. Other factors that need to be considered include the potential human exposure and the possibility of non-linear kinetics at high doses, which can complicate data interpretation and extrapolation to humans.

4.3.4.5 Administration of the test substance

Differences in toxicity related to route of administration are common, and therefore the route of administration of the test substance should approximate that of normal human exposure. For risk assessment of chemicals in food, studies in which the test substance is administered orally are the most useful. However, in some instances (e.g. contaminants), most of the available data may be from routes other than the oral route; for resource and animal welfare reasons, it is important to utilize such data where possible. Toxicokinetic data can be used to correct for route-dependent differences in systemic exposure in cases where the available data were derived using a route different from that by which humans are exposed.

For food chemicals (e.g. food additives, residues of pesticides and veterinary drugs), the test substance is often added to the diet. The diet selected must meet the nutritional requirements of the test species. Control and test diets should ordinarily be isocaloric and nutritionally equivalent; the percentage of test substance in the diet and use of a vehicle are relevant issues to address in this regard. Subtle differences in the diet have the potential to result in nutritional imbalances or underfeeding or overfeeding, thereby confounding study results and their interpretation. Pair-feeding can be useful if effects on feed and nutrient intake are suspected—for example, if palatability is an issue. Caloric restriction, intentional or otherwise, can have profound effects on toxicity; for example, it reduces the background tumour burden in animals and thus has the potential to increase the ability of a study to detect a test substance–related increase in incidence. Administration by encapsulation (common in dog studies) or oral intubation (gavage) may be used if the diet does not provide satisfactory delivery; however, such bolus administration is often associated with higher peak blood levels than would occur by dietary administration of the same daily dose. Delivery in drinking-water may be appropriate for a substance used in a beverage; however, measurement of water intake may be inaccurate if, for example, the animals play with water spouts. Addition of microencapsulated test substance into the diet has proved useful for administration of volatile substances, which would otherwise be lost from the diet.

4.3.5 Observations and measurements

Standardized protocols for tests of general systemic toxicity define a range of end-points and indicators of toxicity. These include, but are not limited to, mortality, cage-side observations, haematology, blood chemistry, gross pathology, histopathology and functional assessments.

4.3.5.1 *Mortality*

Except for lifetime studies, mortality greater than 10% in any treatment or control group is a cause for concern. High mortality in high-dose groups may be an indication of poor dose selection. High rates of mortality increase the chances for autolysis of tissues and organs, possibly resulting in incomplete data collection. High mortality may also be indicative of infection or other problems not associated with the test substance that could compromise study results and interpretation.

4.3.5.2 *Observations of test animals*

Routine cage-side observations are made on all animals at least once or twice a day throughout the study to assess general signs of pharmacological or toxicological effects and to detect morbidity and mortality. Expanded sets of observations, including functional evaluations performed inside or outside of the cage, are commonly incorporated in tests of general systemic toxicity. Such observations provide a general indication of the overall state of health of the animal, and they may identify the need to conduct additional testing with either standard or modified experimental designs (e.g. ataxia or seizures indicate central nervous system toxicity and call for a comprehensive neurotoxicity assessment).

4.3.5.3 *Body weight and feed intake data*

Test animals and controls are weighed on a regular basis (usually weekly for 13 weeks, then monthly thereafter), and food intake is assessed during the conduct of a study. Reductions in body weight or decrements in body weight gain are sensitive indicators of toxicity; in some cases, however, diet palatability rather than toxicity may be the reason for changes in feed intake and body weight. Failure to monitor feed intake or to regularly measure body weight seriously compromises the interpretation of toxicity studies on food chemicals.

4.3.5.4 *Ophthalmology*

Eye examinations in all animals are typically conducted at the start and end of a study. Anatomical differences in eye structure among various species have to be factored in to the interpretation of any findings. Although ophthalmology rarely reveals changes, it was a key investigation in the evaluation of the toxicity of the food and feed colour canthaxanthin (FAO/WHO, 1995).

4.3.5.5 *Haematology*

Blood is sampled in either fasting or non-fasting animals at variable time periods throughout the study, usually at the start and at the end of the study or, in a chronic study, at other time intervals in between. Measurements include haematocrit, haemoglobin concentration, erythrocyte count, total and differential leukocyte counts, mean corpuscular haemoglobin, mean corpuscular volume and mean corpuscular haemoglobin concentration. Clotting time, prothrombin time, thromboplastin time and platelet count are measured to assess clotting potential. Reticulocyte counts and changes in bone marrow cytology are also appropriate measures to include in assessing injury to the haematopoietic system.

The interpretation of results may be difficult as a result of turnover of cell types in the bone marrow or lymphoid tissue. Other sources of variability in the data may come from stress or nutritional factors and age of the animals, to name but a few. In addition, adaptation or tolerance may alter the responses observed over time. Because of their variability, interpretation of the toxicological significance of haematological changes requires careful consideration of consistency of effect, dose–response and comparison with historical control ranges.

4.3.5.6 *Clinical chemistry*

Clinical chemistry tests in general include measurements of electrolyte balance, carbohydrate metabolism, and liver and kidney function. Serum enzyme levels indicative of hepatocellular function that are typically evaluated include alanine aminotransferase (ALT, previously known as serum glutamate–pyruvate transaminase, or SGPT), aspartate aminotransferase (AST, previously known as serum glutamate–oxaloacetate transaminase, or SGOT), sorbitol dehydrogenase and glutamate dehydrogenase. Assessment of hepatobiliary function may include measurements of serum alkaline phosphatase, bilirubin (total), gamma-glutamyl transpeptidase (GGT), 5′-nucleotidase and total bile acids. Markers of cellular function or change include albumin, calcium, chloride, cholesterol (total), cholinesterase, creatinine, globulin (calculated), glucose (in fasted animals), phosphorus, potassium, protein (total), sodium, triglycerides (fasting) and urea nitrogen. Other tests for acid/base balance, hormones, lipids, methaemoglobin or proteins may be indicated, depending on the nature of the test substance.

Changes in serum enzyme levels are commonly associated with target organ toxicity, because enzymes are released from injured cells. Thus, changes in clinical chemistry parameters may signal renal, cardiac or hepatic toxicity. They may be particularly useful for interpretation of study results where there are changes in organ weight, such as liver or kidney, but no overt histopathological changes, as alterations in clinical chemistry parameters associated with organ function can be the first indication of toxicity. A number of enzyme changes are associated with cardiotoxicity, for example, including increases in AST, lactate dehydrogenase and creatinine kinase. Changes in plasma lipids may indicate liver toxicity, whereas changes in blood glucose suggest the possibility of renal toxicity. Concentrations of electrolytes vary with food intake and hydration status, so they are not very sensitive indicators of toxicity.

Clinical chemistry data are subject to a number of sources of variability. Temperature and humidity are two environmental factors that could influence results. Attributes of the test animals, such as sex and age, and study conditions, such as time of sampling and extent of handling, may cause variability in the data recorded. Thus, as with haematological changes, interpretation of changes in clinical chemistry parameters requires careful consideration of consistency of effect, dose–response and comparison with historical control ranges.

Measurement of the test substance in blood samples can provide important information on systemic exposure. Absorption and presystemic metabolism are important factors in determining how much of the test substance reaches the systemic circulation. Toxicokinetics, which defines the movement of a substance around the body and delivery to its site of action, is addressed in section 4.2. Toxicokinetic data from short-term studies can provide useful information for the design of long-term studies, especially in relation to dose selection.

4.3.5.7 *Urinalyses*

Urinalyses consist of determining the volume of urine produced, specific gravity, pH, glucose and protein. In addition, microscopic evaluation for sediment and presence of blood or blood cells is typically done. These analyses are usually conducted during the last week of the study. Analysis of urine, and faeces if indicated, may provide important information relating to changes in normal excretory functions caused by the test substance.

4.3.5.8 Necropsy

Gross necropsy, including examination of external surfaces, orifices, cranial, thoracic and abdominal cavities, carcass and all organs, is typically conducted on all animals. Necropsy should be performed soon after an animal is killed or found dead, or steps need to be taken so that interpretation of the data is not compromised by loss of tissues due to autolysis. Tissue specimens should be taken from the animals and placed in appropriate fixatives during necropsy for subsequent histopathological examination.

4.3.5.9 Organ weight

Organs that are typically weighed include the adrenals, brain, epididymides, heart, kidneys, liver, lung, spleen, testes, thyroid/parathyroid, thymus, ovaries and uterus. Data are often expressed as absolute weights and relative to the animal's body weight. Ratios of organ weight to brain weight may be more reliable indicators of organ-directed toxicity than are ratios of organ weight to body weight; this is because brain weight is rarely affected nonspecifically by toxicity, whereas body weight is more variable and may change as a result of toxicity. Organ weight changes may be indicators of possible morphological or functional changes.

4.3.5.10 Histological examination

In rodents, gross lesions and all scheduled tissues from the animals in the control group and high-dose group should be microscopically examined. When effects are observed, histological examination is extended to other dose groups until a dose level is examined at which no effects are observed. Any animals found dead or terminated early in the study must also be examined histologically. If a small number of animals are tested (e.g. in studies using dogs), histological examinations are normally performed on the controls and all treated groups.

The appropriateness of the fixation and staining techniques for various types of tissues may influence the ability to interpret study results. For example, artefacts such as vacuoles may be produced inadvertently and confused with manifestations of toxicity if fixation is done incorrectly. Ineffective visualization of tissue components and inclusions could result if routine stains (e.g. haematoxylin and eosin) are used when special stains (e.g. silver staining) are required. Properly conducted histological examination is usually the most powerful means of assessing toxicity. As with other toxicological end-points, adaptation or tolerance may alter the responses observed over time. Thus, minor changes observed in short-term studies may no longer be evident in the terminal kills in chronic studies. More commonly, changes observed in short-term studies may become more severe in chronic studies. In addition, normal age-related pathological changes may mask the toxic effects of a chemical in chronic studies.

4.3.5.11 Neurotoxicity and immunotoxicity

Tests of general systemic toxicity commonly incorporate some end-points that are useful for an initial evaluation of the neurotoxic and immunotoxic potential of the test substance. These assessments can be used to define additional testing requirements. The incorporation of additional end-points, however, should not compromise the original purpose of the study. More information on neurotoxicity and immunotoxicity can be found in sections 4.8 and 4.9, respectively.

4.3.5.12 Reversibility

Additional animals are sometimes included in short-term general systemic toxicity studies to determine if effects that might have been observed in earlier studies are reversible. Studying reversibility can assist in deciding whether a change is a physiological or adaptational effect, rather than a toxic effect. The relevance of the reversibility of a toxic effect will depend on the pattern of human exposure. For example, if exposure to a particular chemical in the diet could be more or less daily, then reversibility does not lessen the potential risk.

4.3.5.13 Other considerations

The comparison of data from treated groups with data from concurrent controls is the most important part of the analysis. However, comparison with data from historical controls may be necessary to understand the significance of a finding. Historical control data should be from the same strain of animals, preferably from the same test facility and relatively concurrent (e.g. over 5 years centred on the study of interest).

Statistical analyses are essential for evaluating data from rodent studies. For dogs, the data collected for each animal may be evaluated individually, with each dog serving as its own control (to the extent possible). There are limitations in interpreting results of studies conducted in dogs when too few animals are entered into the studies.

Dose–response relationships should be analysed to determine if the effect is significantly related to treatment and also to provide the information necessary for risk characterization (see chapters 5 and 7). Risk characterization frequently focuses on data from long-term, general systemic toxicity studies, as these often show the greatest effects at the lowest doses.

Studies of general systemic toxicity with durations of a year or less are not adequate to determine the carcinogenic potential of a test substance. However, in rodents, it is possible to conduct combined chronic toxicity/carcinogenicity studies, which are usually 18 months (mice) or 2 years (rats) in duration. As with indicators of immunotoxic or neurotoxic potential, indications of carcinogenic potential obtained from a shorter-duration toxicity study may be a signal that appropriately designed and conducted carcinogenicity tests may be needed (see section 4.6).

Conclusions from tests of general systemic toxicity should be made taking into account everything that is known about the test substance and test conditions. Data on intermediate or precursor effects identified in short-term studies can be useful both for dose selection in long-term studies and also in assessing the possible mode of action.

4.4 Acute toxicity

4.4.1 Introduction

Acute toxicity describes the responses of an organism that are observed within a short time of exposure to, or administration of, a chemical, either as a single exposure or dose or (less commonly) as multiple exposures or doses received over a period of 24 h or less. The nature of the toxicity ascertained normally involves severe adverse reactions or death. Formal acute toxicity tests in animals usually record such reactions for a period of 14 days after the administration of the chemical. In relation to most chemicals in food, acute toxicity tests are not generally useful for hazard identification or risk assessment, because human exposures usually are considerably lower and continue for much longer than the exposures that give rise to acute toxicity. Moreover, other types of toxicity usually occur at doses well below those that are acutely toxic, and it is these other toxicities that are normally pivotal to the risk assessment. However, in certain circumstances, such as the sporadic presence of high residues of an acutely toxic pesticide or a microbial contaminant, there is the potential for acute effects, and acute toxicity needs to be assessed.

JECFA and JMPR routinely consider the toxicity of chemicals in food and establish ADIs or tolerable daily intakes (TDIs), usually on the basis of data from repeated-dose studies, such as chronic toxicity or multigeneration studies. Some substances (e.g. certain metals, mycotoxins, marine biotoxins, veterinary drug residues, pesticide residues or low-digestible carbohydrates, such as polyol sweeteners) could give rise to acute health effects in relation to short periods of intake. JECFA has included in its evaluations an assessment of acute effects (e.g. for inorganic tin) and, where appropriate, the possibility of acute effects in sensitive individuals. JMPR has also set ARfDs for some pesticides and now routinely considers the need to set an ARfD for all pesticides it evaluates.

The appropriateness, or otherwise, of using doses and end-points from subchronic and chronic studies to establish ARfDs needs to be carefully considered. Particular weight should be given to observations and investigations at the beginning of repeated-dose studies. In the absence of information to the contrary, all toxic effects seen in repeated-dose studies should be evaluated for their relevance in establishing an ARfD.

The guidance prepared by JMPR on the setting of ARfDs is outlined in chapter 5 (section 5.2.9). It offers a stepwise approach for setting ARfDs for agricultural pesticides, but the principles are also applicable to other chemical residues in food and drinking-water. In particular, the detailed guidance (Solecki et al., 2005) discusses some toxicological end-points that may be particularly relevant as key acute toxicity alerts. JMPR has also proposed a protocol for a single-dose study, described below.

4.4.2 Guidance for a single-dose study

Currently available data sets usually do not allow accurate evaluation of the acute toxicity of compounds. JMPR has therefore developed a protocol for a single-dose study, with the aim of enabling more accurate derivation of ARfDs. The protocol describes a targeted study suitable for substances with a well-defined toxicity profile but an inadequate database for derivation of an ARfD. Such a single-dose study should not be regarded as routinely required, but rather as a higher-tier study that is necessary only when refinement of the acute risk assessment is required. For example, if a compound has negligible residues, such that dietary intake calculations indicate an adequate margin of safety even when measured against a conservative ARfD derived from a repeated-dose study, then it should be considered unnecessary to perform a single-dose study.

A specific study designed to enable an accurate ARfD to be set should be undertaken only once the toxicological profile of an active substance is reasonably well documented and understood (i.e. at least the short-term toxicity has been evaluated in rats and dogs). The most sensitive species and relevant toxicological end-points for an active substance should be known, enabling a focused study to be designed to investigate the end-points. A flexible approach is necessary, depending on the species and the observed or expected effects with a given substance. Only the minimum number of animals necessary for a thorough safety assessment should be used, while ensuring the minimum amount of distress in the animals in the test.

The principle of the study is to administer the test substance orally as a single dose at several dose levels to groups of experimental animals. A control group is also included. The animals are followed closely for signs of toxicity, with termination of subgroups at one of two time periods (within 24 h and up to 14 days post-treatment). Dose levels and study design will be influenced by the quantitative and qualitative outcome of the repeated-dose studies and findings in existing high-dose acute studies and will be supported by relevant data on toxicokinetics.

The aim of the single-dose study is to identify the most appropriate NOAEL or lowest-observed-adverse-effect level (LOAEL) to derive an ARfD, to provide further information on the dose–response curve, time to peak effects and reversibility for the acute toxic effects, and to provide a flexible approach for an adequate characterization of relevant acute effects. The single-dose study does not aim to identify any lethal doses or provide data on mortality or morbidity after acute exposure to a chemical. The information should be considered with a view to possible refinement of safety factors used in the derivation of the ARfD.

Section 4.5

Genotoxicity

Second edition

(2020)

CONTENTS

This text updates section 4.5 of Chapter 4, Hazard Identification and Characterization: Toxicological and Human Studies, of Environmental Health Criteria 240 (EHC 240), which was originally published in 2009. It was developed through an expert consultation and further advanced following comments received through a public consultation in December 2019.

For abbreviations used in the text, the reader may refer to the list of abbreviations at the front of this section. Definitions of select terms may be found in the glossary in Annex 1 of EHC 240 (http://www.inchem.org/documents/ehc/ehc/ehc240_annex1.pdf).

List of abbreviations

ACD	Advanced Chemistry Development, Inc.
ADI	acceptable daily intake
AOP	adverse outcome pathway
ARfD	acute reference dose
ATSDR	Agency for Toxic Substances and Disease Registry (USA)
BMD	benchmark dose
CAS	Chemical Abstracts Service
CCRIS	Chemical Carcinogenesis Research Information System
CEBS	Chemical Effects in Biological Systems
Cefic	European Chemical Industry Council
CHL	Chinese hamster lung
CHO	Chinese hamster ovary
CTD	Comparative Toxicogenomics Database
DDI	DNA damage–inducing
DNA	deoxyribonucleic acid
ECHA	European Chemicals Agency
EFSA	European Food Safety Authority
EHC	Environmental Health Criteria
EU	European Union
EURL ECVAM	European Union Reference Laboratory for alternatives to animal testing
FAO	Food and Agriculture Organization of the United Nations
GENE-TOX	Genetic Toxicology Data Bank
GLP	Good Laboratory Practice
gpt	glutamic–pyruvic transaminase
HBGV	health-based guidance value
Hprt/HPRT	hypoxanthine–guanine phosphoribosyl transferase
HTRF	Homogeneous Time-Resolved Fluorescence
IATA	Integrated Approaches to Testing and Assessment
ICH	International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
INCHEM	Internationally Peer Reviewed Chemical Safety Information
IPCS	International Programme on Chemical Safety
IRIS	Integrated Risk Information System (USA)
ISS	Istituto Superiore di Sanità (Italy)

ISSMIC	Istituto Superiore di Sanità database on in vivo mutagenicity (micronucleus test)
ISSSTY	Istituto Superiore di Sanità database on in vitro mutagenicity in *Salmonella typhimurium* (Ames test)
JECDB	Japanese Existing Chemical Data Base
JECFA	Joint FAO/WHO Expert Committee on Food Additives
JMPR	Joint FAO/WHO Meeting on Pesticide Residues
LRI	Long-range Research Initiative
MAK	maximum workplace concentration
MN	micronucleus/micronuclei
MOA	mode of action
MOE	margin of exposure
NGS	next-generation DNA sequencing
NIHS	National Institute of Health Sciences (Japan)
NOAEL	no-observed-adverse-effect level
NOGEL	no-observed-genotoxic-effect level
NTP	National Toxicology Program (USA)
OECD	Organisation for Economic Co-operation and Development
PAH	polycyclic aromatic hydrocarbon
PCR	polymerase chain reaction
Pig-a	phosphatidylinositol glycan complementation group A
qPCR	quantitative polymerase chain reaction
QSAR	quantitative structure–activity relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RNA	ribonucleic acid
RT-qPCR	reverse transcription quantitative polymerase chain reaction
S9	9000 × *g* supernatant fraction from rat liver homogenate
SAR	structure–activity relationship
SciRAP	Science in Risk Assessment and Policy
SYRCLE	Systematic Review Centre for Laboratory Animal Experimentation
Td	threshold dose
TDI	tolerable daily intake
T.E.S.T.	Toxicity Estimation Software Tool
TG	test guideline; thioguanine
TIMES	tissue metabolism simulator

Tk/TK	thymidine kinase
ToxRTool	Toxicological data Reliability Assessment Tool
TTC	threshold of toxicological concern
UDS	unscheduled DNA synthesis
USA	United States of America
USEPA	United States Environmental Protection Agency
USFDA	United States Food and Drug Administration
WHO	World Health Organization
WOE	weight of evidence
Xprt/XPRT	xanthine–guanine phosphoribosyl transferase

List of contributors

Dr Virunya Bhat
PAHO/WHO Collaborating Centre on Food Safety, Water Quality and Indoor Environment, NSF International, Ann Arbor, Michigan, United States of America (USA)

Emeritus Professor Alan R. Boobis (*co-lead author*)
National Heart & Lung Institute, Faculty of Medicine, Imperial College London, London, United Kingdom

Dr Riccardo Crebelli
Istituto Superiore di Sanità, Rome, Italy

Dr Nathalie Delrue
Test Guidelines Programme, Environment, Health and Safety Division, Environment Directorate, Organisation for Economic Co-operation and Development, Paris, France

Professor David Eastmond (*co-lead author*)
Department of Molecular, Cell, and Systems Biology, University of California, Riverside, California, USA

Dr Susan Page Felter
Mason, Ohio, USA

Dr Rainer Guertler
Federal Institute for Risk Assessment (BfR), Berlin, Germany

Professor Andrea Hartwig
Karlsruhe, Germany

Dr Frank Le Curieux
European Chemicals Agency, Helsinki, Finland

Professor Angelo Moretto
Department of Biomedical and Clinical Sciences, University of Milan, and International Centre for Pesticides and Health Risk Prevention, Luigi Sacco Hospital, Milan, Italy

Professor Pasquale Mosesso
Department of Ecological and Biological Sciences, Università degli Studi della Tuscia, Viterbo, Italy

Dr Utz Mueller
Australian Pesticides and Veterinary Medicines Authority, Kingston, Australian Capital Territory, Australia

Dr Takehiko Nohmi
Biological Safety Research Center, National Institute of Health Sciences, Kamiyoga, Setagaya-ku, Tokyo, Japan

Dr Grace Patlewicz
National Center for Computational Toxicology, United States Environmental Protection Agency, Durham, North Carolina, USA

Professor David H. Phillips
King's College London, London, United Kingdom

Dr Andrea Richarz
Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra, Italy

Dr Raymond R. Tice
National Institutes of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Dr Paul A. White
Genetic Toxicology Group, Environmental Health Sciences & Research Bureau, Environmental & Radiation Health Sciences Directorate, Healthy Environments & Consumer Safety Branch, Health Canada, Ottawa, Ontario, Canada

Dr Kristine L. Witt
National Institutes of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

4.5 Genotoxicity

4.5.1 Introduction

The study of toxic effects on the inherited genetic material in cells originated with the experiments of Muller (1927), who observed "artificial transmutation of the gene" by ionizing radiation in the fruit fly, *Drosophila melanogaster*. Chemically induced mutation also has a long history, with the first scientific publication, using Muller's fruit fly model, describing mutations arising from exposure to sulfur mustard (Auerbach, Robson & Carr, 1947). A key event stimulating the development and validation of genetic toxicity tests occurred in 1966, when geneticists recommended at a conference sponsored by the United States National Institutes of Health that food additives, drugs and chemicals with widespread human exposure be routinely tested for mutagenicity (see next paragraph for definitions) (Zeiger, 2004).

The term "mutation" refers to permanent changes in the structure or amount of the genetic material of an organism that can lead to heritable changes in its function; these changes include gene mutations as well as structural and numerical chromosomal alterations. The term "mutagen" refers to a chemical that induces heritable genetic changes, most commonly through interaction with DNA,[1] and "mutagenicity" refers to the process of inducing a mutation. The broader terms "genotoxicity" and "genetic toxicity", which are synonymous, include mutagenicity, but also include DNA damage, which may be reversed by DNA repair processes or other known cellular processes or result in cell death and may not result in permanent alterations in the structure or information content of the surviving cell or its progeny (OECD, 2017a). When reference is made to genotoxicity testing, often what is meant is mutagenicity testing. More properly, genotoxicity testing also includes tests that measure the capability of substances to damage DNA or cellular components regulating the fidelity of the genome – such as the spindle apparatus, topoisomerases, DNA repair systems and DNA polymerases – and encompasses tests of a broad range of adverse effects on genetic components of the cell. Although such information can be of value in interpreting the results of mutagenicity tests, it should be considered supplementary data when assessing mutagenic potential. Therefore,

[1] Pro-mutagens are mutagens requiring metabolic activation for mutagenesis.

the broader term "genotoxicant" is used to refer to a chemical that induces adverse effects on genetic components via any of a variety of mechanisms, including mutation, but does not necessarily connote the ability to cause heritable changes. The purpose of mutagenicity testing is to identify substances that can cause genetic alterations in somatic or germ cells, and this information is used in regulatory decision-making (OECD, 2017a).

The overview presented in this section focuses on the identification of mutagens and on the use of such information in assessing the role of DNA-reactive gene mutation in the adverse effects of chemicals, consistent with the World Health Organization (WHO)/International Programme on Chemical Safety (IPCS) harmonized scheme for mutagenicity testing (Eastmond et al., 2009).

National and international regulatory agencies historically have used genotoxicity information as part of a weight-of-evidence (WOE) approach to evaluate potential human carcinogenicity and its corresponding mode of action (MOA; discussed further in section 4.5.4.4). A conclusion on the genotoxic potential of a chemical – and, more specifically, on a mutagenic MOA for carcinogenicity – can be made on the basis of the results of only a few specific types of study, if properly conducted and well reported.

Information on mutagenicity is also of value in assessing the risk of other adverse effects, particularly developmental effects occurring through mutation of germ cells or genotoxicity occurring in somatic cells during embryogenesis and fetal development (Meier et al., 2017).

A chemical could be acknowledged as having genotoxic potential but low concern for a mutagenic MOA in its carcinogenicity or other adverse effects because of mitigating factors, such as toxicokinetics (e.g. phenol and hydroquinone; UKCOM, 2010) or overwhelming toxicity (e.g. dichlorvos; FAO/WHO, 2011).

Some regulatory agencies, such as those within the USA, Canada, the United Kingdom and the European Union (EU), consider heritable mutation a regulatory end-point. Mutations in germ cells may be inherited by future generations and may contribute to genetic disease. Germline (or germ cell) or somatic cell mutations are implicated in the etiology of some disease states, such as cancer, sickle cell anaemia and neurological diseases (Youssoufian &

Pyeritz, 2002; Erickson, 2003, 2010; Lupski, 2013; D'Gama et al., 2015). Inherited mutations linked to human diseases are compiled in the Human Gene Mutation Database (HGMD, 2017).

Testing for mutagenicity should utilize internationally recognized protocols, where they exist. For example, mutagenicity (gene mutation and structural and numerical chromosomal alterations) is one of six basic testing areas that have been adopted by the Organisation for Economic Co-operation and Development (OECD, 2011) as the minimum required to screen high-production-volume chemicals in commerce for toxicity.

Safety assessments of chemical substances with regard to mutagenicity are generally based on a combination of tests to assess three major end-points of genetic damage associated with human disease:

1) gene mutation (i.e. point mutations or deletions/insertions that affect single or blocks of genes);
2) clastogenicity (i.e. structural chromosome changes); and
3) aneuploidy (i.e. the occurrence of one or more extra or missing chromosomes, leading to an unbalanced chromosome complement).

Existing evaluation schemes tend to focus on single chemical entities with existing data. However, there are scenarios that do not involve single chemicals, such as enzyme preparations used in food production that are mixtures including proteins and one or more low-molecular-weight chemicals, or that involve chemicals, such as minor plant and animal metabolites of pesticides or veterinary drugs, that lack empirical data. Special considerations related to these scenarios, including the evaluation of the mutagenicity of food extracts obtained from natural sources, which are often complex botanical mixtures that may not be fully characterized, are also discussed in this section.

4.5.1.1 *Risk analysis context and problem formulation*

The identification of compounds to which exposure may lead to cancer (or other adverse effect) via a mutagenic MOA affects how these compounds are handled within regulatory paradigms. A distinction is often made between substances that require regulatory approval before use (e.g. pesticides, veterinary drugs, food additives) and those to which exposure is unavoidable (e.g. contaminants, natural constituents of the diet). In practice, this distinction affects the nature of information provided to risk managers. For substances intentionally added to or used in food that require regulatory approval, key outputs of the hazard characterization are health-based guidance values (HBGVs) (e.g. acceptable daily intake [ADI], tolerable daily intake [TDI], acute reference dose [ARfD]). Intrinsic to the establishment of such a value is that there is negligible concern when exposure is below the HBGV, and implicit in this is that there are biological and population thresholds for the adverse effect. Mutagenicity, particularly gene mutation, is often assumed to lack a threshold, in part due to uncertainty related to human exposure levels and the assumption that even one molecule of a DNA-reactive mutagen could theoretically induce heritable changes leading to an adverse effect. Consequently, for substances considered to act through a mutagenic MOA, it may not be possible to establish with confidence an HBGV below which concern is considered negligible; under such circumstances, in the context of the work of the Joint FAO/WHO Expert Committee on Food Additives (JECFA) and the Joint FAO/WHO Meeting on Pesticide Residues (JMPR), it is generally understood that it would be inappropriate to establish an HBGV. Nevertheless, risk managers may still require an indication of the degree of health concern, and this should be reflected in the problem formulation, which is a key component of risk analysis that involves consideration of the risk management scope and goals in relation to relevant exposure scenarios, available resources, urgency of the assessment and the level of uncertainty that is acceptable (Meek et al., 2014). In practice, in the international context in which JECFA and JMPR work, rather than a detailed problem formulation, the general question to be addressed is whether the compound poses a significant mutagenic hazard and, if so, whether there is a concern at estimated dietary exposures.

Most currently approved (e.g. by OECD) tests for mutagenicity, both in vitro and in vivo, are designed to identify a mutagenic hazard and in general are used for a simple yes/no answer for risk management purposes (see section 4.5.2). Such a dichotomous approach is useful for managing substances intentionally permitted in food, such as food additives, pesticides and veterinary drugs, for which regulatory approval is often required. Qualitative, semiquantitative and non-testing approaches useful for managing data-poor substances, such as unavoidable contaminants and plant and animal metabolites, include:

- in silico approaches, such as (quantitative)structure–activity relationship [(Q)SAR] models (see section 4.5.5.1);
- the threshold of toxicological concern (TTC) approach (see section 4.5.5.2); and
- grouping and read-across approaches (see section 4.5.5.3).

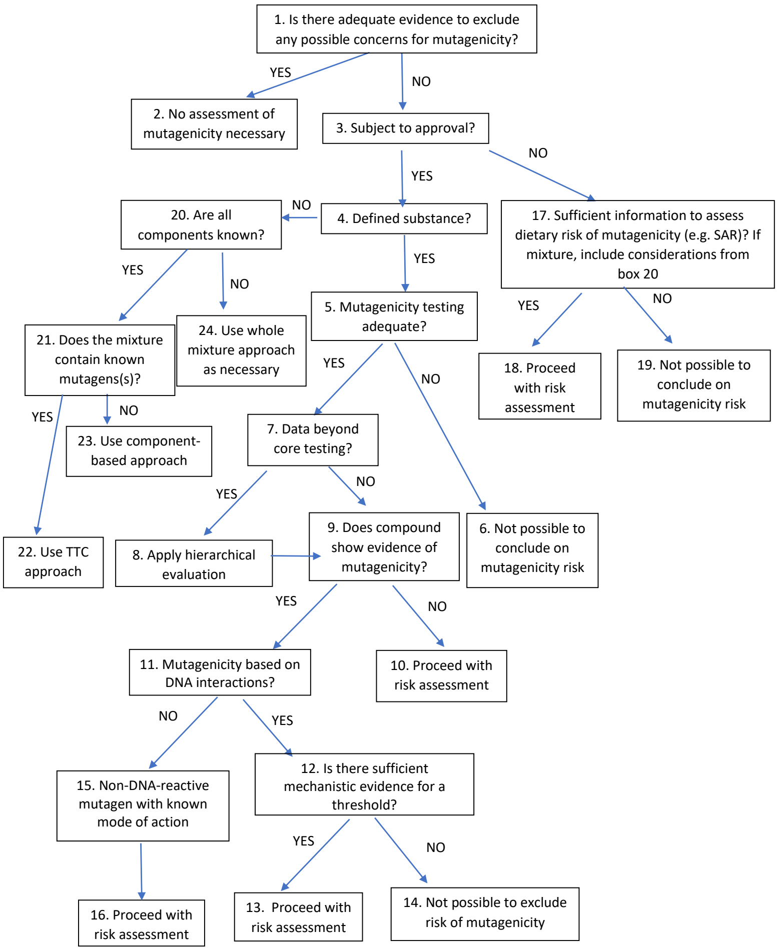
Quantitative dose–response approaches for genotoxicity may also be appropriate for unavoidable contaminants (see section 4.5.7.4). However, as this is a deviation from current practice, the acceptability of such approaches should be indicated in the problem formulation (see, for example, MacGregor et al., 2015a,b; UKCOM, 2018).

JECFA and JMPR do not set data requirements for their food additive, veterinary drug and pesticide residue evaluations, although there is a minimum data set expected in order to conduct an assessment. In the case of mutagenicity, the nature of and guidance to interpret the information are described in this section. In general, JECFA and JMPR evaluate the available data, most often generated in support of regulatory submissions elsewhere. Data requirements set by a regulatory agency for a chemical evaluation can vary substantially, depending on the chemical's use and potential for human exposure.

4.5.1.2 Decision-tree for assessing the mutagenicity of substances that can be found in food

Fig. 4.1 is a decision-tree illustrating issues to be considered in assessing the mutagenic potential of different types of substances that can be found in food. Subsequent subsections will describe the process of identifying relevant and reliable mutagenicity data and, depending on the regulatory jurisdiction, determining whether the data and WOE are adequate to conclude on mutagenic potential. If a substance is shown to possess mutagenic potential, the process of discerning the likelihood of a mutagenic MOA for carcinogenicity and other adverse effects is also discussed, in conjunction with repeated-dose toxicity or carcinogenicity data, if available.

Fig. 4.1. Decision-tree illustrating issues to be considered in assessing the mutagenic potential of different types of substances that can be found in food



1. Is there adequate evidence to exclude any possible concerns for mutagenicity?

While it may be rare to exclude possible concerns for mutagenicity a priori, occasionally the nature of the substance or its production process may provide sufficient assurance that substance-specific mutagenicity data are not necessary. One example is a natural constituent of the diet produced by a fully controlled process (e.g. invertase derived from *Saccharomyces cerevisiae* fermentation; FAO/WHO, 2002). [See section 4.5.6.4.]

2. No assessment of mutagenicity necessary

If the answer to the question in box 1 is YES, no further consideration of mutagenic potential is necessary, and risk assessment of non-genotoxic (non-mutagenic) effects can proceed. [See other sections of chapter 4.]

3. Subject to approval?

If concerns about potential mutagenicity cannot be excluded a priori (i.e. the answer to the question in box 1 is NO), does the substance require regulatory approval in Member States prior to uses that could knowingly result in its presence in food (i.e. pesticides, veterinary drugs and food additives, including flavouring agents)? Excluded are contaminants and natural constituents of the diet (e.g. mycotoxins), for which there are different considerations for tolerated concentration limits. [See section 4.5.1.1.]

4. Defined substance?

If the answer to the question in box 3 is YES, does the substance comprise a single chemical or a small number (e.g. stereoisomers) of chemicals of known structure? In other words, is it chemically defined? If not, the substance is considered a mixture. Included in this group are single substances of unknown structure. Note that a critical consideration is the purity of the substance. Expert judgement is needed to decide whether, based on analytical or other relevant data, a substance that nominally is a single chemical is so impure that it should be considered a mixture with uncharacterized constituents (e.g. <90% purity). [See section 4.5.6.1.]

5. Mutagenicity testing adequate?

For substances subject to regulatory approval in some jurisdictions and where the answer to the question in box 4 is YES, are the available data adequate to conclude whether the substance is likely to pose a mutagenic risk in vivo at dietary levels of exposure? [See sections 4.5.2 and 4.5.4.4.]

6. Not possible to conclude on mutagenicity risk

If mutagenic potential has not been adequately tested (i.e. the answer to the question in box 5 is NO), it is not possible to conclude on the likelihood of mutagenic risk in vivo at dietary levels of exposure. As such, it may be inappropriate to establish HBGVs that encompass potential mutagenicity. The main data gaps precluding a conclusion on mutagenic potential should be clearly articulated. [See section 4.5.4.5.]

7. Data beyond core testing?

For some compounds, particularly newer ones, mutagenicity testing may be adequate (i.e. the answer to the question in box 5 is YES) based on available data from a small range of relevant and reliable "standard" mutagenicity tests. [See section 4.5.4.2.] However, for others, particularly those in use for some time or about which there are specific concerns (e.g. bisphenol A; EFSA, 2015), the available data may be much more extensive, including a variety of test systems with a range of quality (i.e. in design, conduct or reporting), and the results may be contradictory. It should be noted if the genotoxicity database is considered to fall into this category. [See section 4.5.3.]

8. Apply hierarchical evaluation

When the genotoxicity database is complex or contradictory (i.e. the answer to the question in box 7 is YES), a WOE approach that considers factors such as the results of in vivo versus in vitro testing, the relevance of the test or end-point to humans and the relevance of the route of exposure and dose is used to weight the studies. [See sections 4.5.4.1 and 4.5.4.2.]

9. Does compound show evidence of mutagenicity?

Regardless of how extensive the database is (i.e. the answer to the question in box 7 is NO or after application of the hierarchical evaluation in box 8), a WOE conclusion should be reached on whether the substance shows evidence of mutagenicity for relevant end-points. For example, as defined by the OECD, an isolated positive result at high, cytotoxic concentrations in vitro, without evidence of mutagenicity in numerous guideline studies conducted to an appropriate standard, is insufficient to conclude that, overall, there is concern for mutagenicity. As the objective is not a hazard classification, reaching a conclusion requires expert judgement, which should be clearly explained and can often be the most difficult aspect of the assessment. [See sections 4.5.3, 4.5.4.1 and 4.5.4.2.]

10. Proceed with risk assessment

If the WOE does not suggest mutagenicity (i.e. the answer to the question in box 9 is NO), no further consideration of the mutagenic potential of the substance is necessary, and risk assessment of non-genotoxic (non-mutagenic) effects can proceed. [See other sections of chapter 4.]

11. Mutagenicity based on DNA interactions?

If there is evidence of mutagenicity (i.e. the answer to the question in box 9 is YES), the nature of the mutagenicity should be determined – specifically, whether the mutagenicity is based on the parent compound or a metabolite interacting with DNA, thereby resulting in heritable DNA changes. This evidence should come primarily from appropriate tests for gene mutation, clastogenicity and aneuploidy, and supporting evidence may include a variety of non-standard tests, such as DNA reactivity/adduct formation. [See section 4.5.2.]

12. Is there sufficient mechanistic evidence for a threshold?

For a mutagenic chemical (i.e. the answer to the question in box 11 is YES), the relevance of the dose/concentration used in testing to the estimated dietary exposure should be considered. For the majority of mutagens, there may be little or no evidence for an effect threshold. Hence, in the absence of such evidence, it is assumed that even high-dose effects are relevant for assessing mutagenic potential in humans. For a few substances, however, there may be clear mechanistic evidence in vitro and in vivo for a biological threshold. Hence, in theory, it may be possible to discount effects seen only at doses that are irrelevant to conceivable human dietary exposure (or even a multiple of that exposure) (e.g. dichlorvos; FAO/WHO, 2011). [See also section 4.5.7.4.]

13. If there is sufficient mechanistic evidence for a threshold for mutagenicity, proceed with risk assessment

If it is concluded that a biological threshold exists for the mutagenicity observed experimentally (i.e. the answer to the question in box 12 is YES) and, after allowing for interspecies and intraspecies differences, the estimated human dietary exposure is clearly well below this, risk assessment based on the critical effect(s) can proceed. [See other sections of chapter 4.]

14. Not possible to exclude risk of mutagenicity

If it is concluded that the mutagenicity observed experimentally is, or might be, relevant, considering conceivable human dietary exposure levels (i.e. the answer to the question in box 12 is NO), it will ordinarily be inappropriate to establish an HBGV. [See section 4.5.4.5.]

15. Non-DNA-reactive mutagen with known mode of action

For mutagenic compounds in which a DNA-reactive MOA can be excluded (i.e. the answer to the question in box 11 is NO), the nature of the mutagenicity, its molecular mechanism and the dose–response relationship should be characterized. For some mechanisms, there is evidence for a biological threshold – for example, aneuploidy due to spindle disruption or mutagenicity secondary to inflammation that generates reactive oxygen species. [See section 4.5.4.4.]

16. Proceed with risk assessment

The output of the mutagenic hazard characterization (i.e. output from the question in box 15) can be used in the risk assessment, as appropriate. For example, if mutagenicity is considered to exhibit a threshold, the "normal" approach to establishing HBGVs and to risk characterization can be applied. In many cases, this would mean that the critical effect was other than mutagenicity, as it occurred at lower exposure levels. In some cases, it might not be possible to conclude that mutagenicity exhibits a threshold, in which case a margin of exposure (MOE) approach may be appropriate. In either case, a concluding statement regarding the potential risk of mutagenicity in vivo at dietary levels of exposure should be provided. [See section 4.5.4.5.]

17. Sufficient information to assess dietary risk of mutagenicity (e.g. SAR)?

For substances not subject to regulatory approval (i.e. the answer to the question in box 3 is NO) that have unavoidable dietary exposure, such as contaminants or natural dietary constituents (e.g. mycotoxins), it should be assessed whether there is sufficient information to reach a conclusion about potential mutagenicity. When existing empirical mutagenicity data are insufficient to reach a conclusion, additional information from the substance, from related analogues (i.e. read-across) or from in silico approaches, such as (Q)SARs, should also be considered in an overall WOE for the mutagenic potential of the substance. [See sections 4.5.5.1 and 4.5.5.3.]

18. Proceed with risk assessment

Where sufficient information is available to conclude on the mutagenic potential of the substance (i.e. the answer to the question in box 17 is YES), a risk assessment can proceed. This may justify establishing an HBGV, such as a TDI, or the use of an MOE approach. Where exposures are likely to be very low and the compound is a potential mutagen, the TTC approach can be used. If exposure is below the mutagenicity TTC value (0.0025 µg/kg body weight per day for chemicals with structural alerts for DNA reactivity), there is low concern for effects on human health. [See section 4.5.5.2 and other sections of chapter 4.]

19. Not possible to conclude on mutagenicity risk

When it is not possible to conclude on potential mutagenicity (i.e. the answer to the question in box 17 is NO), advice should be provided on the assumption that the substance might be a mutagen. Hence, the TTC for such compounds (0.0025 µg/kg body weight per day) could be used, recognizing the considerable uncertainty in such an assessment and that the risk may be appreciably overestimated. Alternatively, it may be concluded that it is not possible to provide any advice on potential human risk without additional data.

20. Are all components known?

For substances that are not composed of a single defined chemical or a small number of defined chemical entities (i.e. the answer to the question in box 4 is NO), are all of the components of the mixture known? If all of the components are known and have established chemical structures and concentrations, the mixture is considered "simple", whereas if a significant fraction of components are of unknown structure or concentration, the mixture is considered "complex". [See section 4.5.6.1.]

Although there is no explicit question in the decision-tree as to whether mixtures are subject to approval, a number of the considerations for defined substances will also apply to mixtures. That is, for those mixtures subject to approval, consideration will need to be given to the adequacy of mutagenicity testing (of the components or of the mixture as a whole). For those that are not, a WOE approach using information on direct testing, read-across and (Q)SAR can be applied, to the extent possible.

21. Does the mixture contain known mutagen(s)?

Where all of the components in a "simple" mixture above a minimum level of concern (as determined by expert judgement) are known (i.e. the answer to the question in box 20 is YES), each component should be assessed for its mutagenicity, on the basis of prior knowledge. Are one or more known mutagens present? If so, these should be assessed before considering the potential mutagenicity of other components.

22. Use TTC approach

For mutagenic substances known to be present in a defined mixture (i.e. the answer to the question in box 21 is YES), the TTC approach can be applied. If estimated human exposure is below the mutagenicity (DNA-reactive gene mutation) TTC, there is low concern for mutagenicity in exposed individuals from these substances, and the remaining components can then be assessed individually, as described under the component-based approach in box 23. If the estimated exposure exceeds the mutagenicity (DNA-reactive gene mutation) TTC, additional information will be needed to determine if there is concern for possible mutagenicity in exposed individuals. [See section 4.5.6.3.]

23. Use component-based approach

For a "simple" mixture in which none of the components is known to be mutagenic (i.e. the answer to the question in box 21 is NO), each component should be assessed for potential mutagenicity, as described for defined chemicals. [See section 4.5.6.1.]

24. Use whole mixture approach as necessary

For a "complex" mixture in which a significant fraction of the mixture is unknown (i.e. the answer to the question in box 20 is NO), extracts, subfractions or the whole mixture should be tested for mutagenicity, depending on the nature of the mixture, the information available and the mixture's intended use. [See section 4.5.6.]

4.5.2 Tests for genotoxicity

More than 100 different in vitro and in vivo genotoxicity test methods exist. Given the high degree of overlap, a much smaller number of methods, most of which have OECD test guidelines (TGs), although some are in an earlier stage of development, are commonly used (Table 4.1) and can be grouped according to the test system (e.g.

Table 4.1. Examples of assays for genotoxicity

Gene mutation	Chromosomal damage	DNA damage/repair
In vitro assays		
Bacterial tests [see section 4.5.2.1] • Reversion to a specific nutrient independence in *Salmonella typhimurium* and *Escherichia coli* (OECD TG 471) *Mammalian tests [see section 4.5.2.2]* • Forward mutation at the *TK*/*Tk* gene (OECD TG 490) in cell lines such as mouse lymphoma L5178Y and human TK6 • Forward mutation at the *Hprt/HPRT* gene (OECD TG 476) in primary cells or cell lines such as mouse lymphoma (L5178Y), Chinese hamster ovary (CHO), Chinese hamster lung (V79), human TK6 and human lymphocytes	• Sister chromatid exchange (OECD TG 479)[a] • Chromosomal aberrations (OECD TG 473) in CHO, CHL or V79 cell lines and human cells (lymphocytes and TK6) [see section 4.5.2.4(a)] • MN (resulting from clastogenicity and aneuploidy) (OECD TG 487) in CHO, CHL or V79 cell lines and human cells (lymphocytes and TK6) [see section 4.5.2.4(b)] • Chromosomal aberrations (OECD TG 490) in mouse lymphoma L5178Y and human TK6 cells [see section 4.5.2.4(c)]	• UDS in primary cultures (often hepatocytes; OECD TG 482)[a] • DNA strand breakage and alkali-labile sites monitored by single-cell gel electrophoresis (comet assay) or by sucrose gradient, filter elution or alkaline unwinding, in cell cultures [see section 4.5.2.6] • Upregulation or stabilization of DNA damage responses (e.g. p53, ATAD5, pH2AX) • DNA adduct measurement in cell cultures

Gene mutation	Chromosomal damage	DNA damage/repair
In vivo assays		
Somatic cell assays [see section 4.5.2.3(a)] • Transgenic rodent assays: *gpt*, Spi⁻ (*gpt* delta mouse or rat), *lacZ* plasmid, bacteriophage or *cII* (Muta™Mouse) or *lacI* or *cII* (Big Blue® mouse or rat) (OECD TG 488) • *Pig-a* gene mutation assay (mouse, rat, human) *Germ cell assays [see section 4.5.2.3(b)]* • Specific locus test (mouse) • Dominant lethal assay (rodents) (OECD TG 478) • Transgenic rodent assays: *gpt,* Spi⁻ (*gpt* delta mouse or rat), *lacZ* or *cII* (Muta™Mouse) or *lacI* or *cII* (Big Blue® mouse or rat) (OECD TG 488)	*Somatic cell assays* • Sister chromatid exchange (OECD TG 482)[a] in bone marrow (rodent) • Chromosomal aberrations (OECD TG 475) [see section 4.5.2.5(a)] • MN (resulting from clastogenicity and aneuploidy) (OECD TG 474) in erythrocytes (rodent) [see section 4.5.2.5(b)] *Germ cell assays* • Chromosomal aberrations (OECD TG 483) (rodent) [see section 4.5.2.5(a)] • Dominant lethal mutations (OECD TG 478) (rodent)	• Strand breakage and alkali-labile sites monitored by single-cell gel electrophoresis (comet assay) in nuclear DNA in various tissues (OECD TG 489) [see section 4.5.2.7(a)] • DNA adduct measurement [see section 4.5.2.7(b)] • UDS (liver; OECD TG 486) [see section 4.5.2.7(c)]

CHL: Chinese hamster lung; CHO: Chinese hamster ovary; DNA: deoxyribonucleic acid; *gpt*: glutamic–pyruvic transaminase; *Hprt*: hypoxanthine–guanine phosphoribosyl transferase; MN: micronuclei; OECD: Organisation for Economic Co-operation and Development; TG: Test Guideline; *Tk*: thymidine kinase; UDS: unscheduled DNA synthesis

[a] OECD TGs for these assays were deleted in 2014; legacy data may be used in a comprehensive assessment of genotoxicity, but new tests of this nature should not be conducted.

in vitro or in vivo) and the genetic end-point assessed for genetic damage:

- Gene mutations:
  - gene mutation in bacteria;
  - gene mutation in mammalian cell lines; and
  - gene mutation in rodents in vivo using constitutive or transfected genes;

- Clastogenicity and aneuploidy:
  - chromosomal aberrations in cultured mammalian cells (to assess structural chromosome changes);
  - micronucleus (MN) induction in cultured mammalian cells (to assess structural and numerical chromosome changes);
  - chromosomal aberration in vivo in mammalian haematopoietic cells (to assess structural chromosome changes); and
  - MN induction in vivo in mammalian haematopoietic cells (to assess structural and numerical chromosome changes);

- DNA damage/repair:
  - DNA damage in vitro (e.g. formation of DNA adducts, DNA strand breaks/alkali-labile sites);
  - end-points related to damage/repair (e.g. unscheduled DNA synthesis [UDS]; gamma-H2AX);
  - DNA damage in vivo (e.g. DNA binding, DNA strand breaks/alkali-labile sites, UDS in liver cells).

Complete consistency among the results of different classes of assays is generally not expected, as the assays measure different end-points. In addition to the commonly used tests in Table 4.1, there are numerous methods with more limited validation, such as those in which yeast, moulds and insects (e.g. *Drosophila*) are used as test organisms.

Identification of germ cell mutagens is difficult, and studies in rodents to identify these agents historically required large numbers of animals. In contrast, identification of somatic cell mutagens can be accomplished in vitro or with fewer animals in vivo. To date, all identified germ cell mutagens are also somatic cell mutagens. Thus, in risk assessment, a default assumption is that a somatic cell mutagen may also be a germ cell mutagen. Regulatory decisions declaring that such hazards exist would not ordinarily have different consequences, unless there are demonstrated differences in potency between the doses causing somatic versus germ cell mutagenicity, which, for example, may result in differential advice to pregnant women and the general population. For the majority of known germ/somatic cell mutagens, if the individual is protected from the genotoxic and carcinogenic effects of a substance, then that individual would also be protected from the heritable genetic effects. Although national regulatory authorities might take a different view, this is the practical viewpoint of JMPR and JECFA at this time, as information on developmental and reproductive toxicity is often available (particularly for chemicals subject to authorization in Member States).

The following text provides a brief description of the main tests for genotoxicity. For full details of test design and data interpretation, and for information on less commonly used tests, the reader is referred to the respective OECD TG (available at https://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788).

4.5.2.1 Bacterial mutagenicity

As one of the original mutagenicity assays (Ames, Lee & Durston, 1973) to be required for regulatory submissions, the bacterial reverse mutation assay (OECD TG 471) remains the most frequently conducted of all current assays. The test uses several strains of *Salmonella typhimurium* that carry different mutations in various genes of the histidine operon, in which form it is widely referred to as the "Ames test", and some strains of *Escherichia coli*, which carry the AT base pair mutation at a critical site in the *trpE* gene. Among these strains, multiple modes of mutation induction (e.g. base substitution or frameshift mutation) can be detected. When these auxotrophic bacterial strains are grown on a minimal agar containing only a trace of the required amino acid (histidine or tryptophan, respectively), only those bacteria that revert by mutation to amino acid independence will grow to form visible colonies. Metabolic activation is provided by exogenous mammalian enzymes – for example, liver post-mitochondrial (S9) fraction from rats induced with Aroclor 1254 or phenobarbital/5,6-benzoflavone.

4.5.2.2 In vitro mammalian cell mutagenicity

Currently, two in vitro assays for the induction of mammalian cell gene mutation have formal OECD TGs, as described below.

(a) Forward gene mutation tests using the *Tk* gene

The mammalian cell TK gene mutation assay (OECD TG 490) detects mutagenic and clastogenic events at the thymidine kinase (*Tk*) locus of L5178Y mouse lymphoma $Tk^{+/-}$ cells (Lloyd & Kidd, 2012). Although less frequently used, the human lymphoblastoid cell line TK6 is also used for evaluating mutations induced at the *TK* locus. Exogenous S9 provides metabolic activation. Cells that remain $Tk^{+/-}$ after chemical exposure die in the presence of the lethal nucleoside analogue trifluorothymidine, which becomes incorporated into DNA during cell replication, but the lethal analogue cannot be incorporated into the DNA of mutated $Tk^{-/-}$ (and $Tk^{-/0}$) cells, which survive and form colonies; large colonies often result from gene mutation (point mutations or base deletions that do not affect the rate of cell doubling), whereas small colonies often result from chromosomal mutation (chromosomal rearrangements or translocations that result in slow growth and extended cell doubling times). Similarly, $TK^{-\backslash-}$ (and $TK^{-/0}$) mutants in TK6 cells can be selected with trifluorothymidine, and early-appearing and late-appearing colonies often indicate gene mutation and chromosome mutation, respectively.

(b) Forward gene mutation tests using the *Hprt* and *Xprt* genes

OECD TG 476 describes a test method that measures mutations at the hypoxanthine–guanine phosphoribosyl transferase (*Hprt*) gene on the X chromosome of mammalian cells or at a transgene of xanthine–guanine phosphoribosyl transferase (*Xprt*) on a somatic chromosome. Male cells possess a single copy of the *Hprt* gene, and one copy of the gene is inactivated in female cells, resulting in one functional allele. Mutation of the single copy makes the cells unable to incorporate lethal 6-thioguanine (6-TG) into their DNA; therefore, mutant cells will survive when cultured in the presence of 6-TG, whereas $Hprt^{+}$ cells will incorporate 6-TG into their DNA during replicative synthesis and die (Dewangan et al., 2018). A number of different cell lines can be used for the HPRT assay (e.g. Chinese hamster ovary [CHO], Chinese hamster lung [V79], mouse lymphoma L5178Y, human TK6), whereas CHO-derived AS52 cells containing the glutamic–pyruvic transaminase (*gpt*) transgene (and having the *Hprt* gene deleted) are used for the XPRT test (OECD TG 476), either directly or in the presence of S9-mix for metabolic activation, or with the use of genetically modified cell lines that stably express metabolic enzymes.

Thus, the TK and HPRT/XPRT assays measure mutant frequencies at the named genes in mammalian cells following chemical exposure, but each genetic target detects a different spectrum of mutational events. Mutant frequency is measured by counting mutant colonies arising on plates with selective media. The mouse lymphoma TK assay (OECD TG 490) is used rather than the HPRT/XPRT assay (OECD TG 476) when an investigator wants to detect a broader range of mutagenic events.

4.5.2.3 *In vivo mammalian cell mutagenicity*

(a) Somatic cell assays

***Transgenic rodent assays*.** The OECD TG 488 assays employ transgenic mice or rats harbouring lambda phage (or plasmid) DNA carrying reporter genes in all cells (Nohmi, Suzuki & Masumura, 2000; Thybaud et al., 2003; Nohmi, Masumura & Toyodo-Hokaiwado, 2017). After chemical treatment, the transgenes are rescued from the DNA as phage particles by in vitro packaging reactions and introduced into *E. coli* cells to detect mutations fixed in vivo as bacterial colonies or phage plaques. These assays are advantageous for further evaluation of rodent carcinogens because gene mutations can be detected in almost any organ or tissue, aiding evaluation of the target organs for carcinogenesis, and because of the ability to distinguish DNA-reactive genotoxic carcinogens from DNA-non-reactive (or non-genotoxic) carcinogens. Transgenic rodent assays – such as the *gpt*, *lacI*, *lacZ* and *cII* assays that detect point mutations (base substitution or frameshift) and the Spi^{-} and *lacZ* plasmid methods that detect deletion mutations – can be integrated into 28-day repeated-dose toxicity studies with other genotoxicity assays, such as the in vivo MN assay (see section 4.5.2.5(b)), *Pig-a* assay (see below) or comet assay (see section 4.5.2.7(a)). DNA sequencing of mutants can be useful to examine chemical MOA by comparing the mutation spectrum with those of other known mutagens and to identify duplicate mutants generated by clonal expansion of single mutants.

Pig-a assay in rats or mice (or humans). This assay uses the constitutive phosphatidylinositol glycan complementation group A (*Pig-a*) gene as a reporter for mutation (Miura et al., 2008a,b; Gollapudi et al., 2015). Mutations in the *Pig-a* gene result in the loss of glycosylphosphatidylinositol-anchored proteins in the cell surface, and thus the mutant cells fail to express surface markers such as the

CD59 or CD24 antigens and be labelled by antibodies targeting these antigens. The absence of these cell surface antigens, which is easily detected by flow cytometry, is a direct reporter of *Pig-a* mutation. The assay is rapid and low cost, requiring only a small volume of blood, and can be conveniently integrated into rodent 28-day repeated-dose toxicity studies along with other genotoxicity assays (Dertinger et al., 2011a; Khanal et al., 2018). This assay can be conducted in rats, mice and humans, because the *Pig-a* gene is conserved. Currently, detection of the *Pig-a* mutant phenotype is limited to erythrocytes (mature and immature) in peripheral blood (Kimoto et al., 2016), which necessitates similar considerations of target tissue exposure as those for the in vivo MN test (see section 4.5.2.5(b)). Other cell types are being investigated for suitability in this assay, such as T-lymphocytes. An OECD TG for this assay is under development (as of July 2020). An in vitro version of the *Pig-a* assay amenable to scoring by flow cytometry is described in section 4.5.7.2.

(b) Germ cell assays

Mouse specific locus test. The specific locus test for mutagenicity in germ cells is rarely used because of its cost and the large number of animals needed (Russell & Shelby, 1985). In a typical specific locus test, chemically exposed male mice are mated with unexposed females that are homozygous for recessive alleles at seven loci (Russell, 2004). If a mutation is induced in one of these loci of male germ cells, the offspring will express altered phenotypes for traits such as eye or coat colour. The interval between chemical treatment and conception is used to identify the stage in spermatogenesis when the mutation was induced. For example, mutations detected in offspring born 49 days after the last treatment are derived from exposed spermatogonial stem cells. About 30 chemicals have been examined by the specific locus test, and several chemicals (e.g. ethyl nitrosourea) were detected as mutagenic in spermatogonial stem cells (Shelby, 1996). Novel approaches, such as Trio analysis, in which direct comparison of DNA sequences is made between parents and offspring (Masumura et al., 2016a,b; Ton et al., 2018), the expanded simple tandem repeats assay (Yauk, 2004) or the transgenic rodent assays described below, have also shown some success in detecting germ cell mutations.

Rodent dominant lethal assay. The dominant lethal assay investigates whether a chemical induces mutations associated with embryo or fetal death. The mutations originate primarily from chromosomal aberrations in germ cells (OECD TG 478). Although the assay has advantages, such as in vivo metabolism, pharmacokinetics and DNA repair processes that contribute to the response, it requires a large number of animals. To conserve animals, this assay can be integrated with other bioassays, such as developmental, reproductive or somatic cell genotoxicity studies.

Transgenic rodent assays. The OECD TG 488 transgenic rodent assays can, with some modifications, also be applicable to the examination of germ cell mutagenesis (Douglas et al., 1995). The transgenes are rescued from male germ cells collected from the cauda epididymis and the vas deferens, where mature sperm are present. Female germ cells are usually precluded because there is no DNA synthesis in the oocyte in adult animals. Unlike somatic cell mutations, where cells are collected shortly after the last treatment of test chemical, sperm cells are collected 49 days (mice) or 70 days (rats) after the last treatment, because those periods are necessary for spermatogonial stem cells to mature into sperm and for the cells to reach the vas deferens and cauda epididymis (Marchetti et al., 2018). Mutations are induced during the proliferation phase of spermatogenesis. A recent evaluation indicates that treatment for 28 days followed by a 28-day expression period allows mutagenic and non-mutagenic chemicals to be distinguished in both rats and mice (Marchetti et al., 2018).

4.5.2.4 *In vitro chromosomal damage assays*

(a) Chromosomal aberration assay

The in vitro chromosomal aberration assay (OECD TG 473) assesses chemical-induced structural chromosomal damage in cultured mammalian cells (e.g. CHO cells, human lymphocytes), but is time-consuming, requires skilled and experienced scorers and does not accurately measure aneuploidy (i.e. changes in chromosome number). In the early years of conducting this assay, excessive cytotoxicity affecting data interpretation was a major confounding factor in many laboratories. As a result, updated guidelines have been established identifying acceptable cytotoxicity levels (OECD, 2016a) and have improved the reliability of the test.

(b) Micronucleus (MN) assay

The in vitro chromosomal aberration assay has gradually been replaced by the in vitro MN assay (OECD TG 487), which is less expensive, faster, less subjective and amenable to automation using flow cytometry or high-content screening; automation allows a far greater number of cells to be scored, thus increasing the statistical power of the assay (Bryce et al., 2010, 2011; Avlasevich et al., 2011). Another feature of the MN assay is its capability to detect both clastogenic and aneugenic events.

Both the in vitro chromosomal aberration assay (see section 4.5.2.4(a) above) and the in vitro MN assay must be conducted under strict conditions limiting cytotoxicity to acceptable levels (defined in the OECD TGs). When these in vitro tests for chromosomal damage are conducted with appropriate bioactivation, more compounds are detected as active for chromosomal damage than in the in vivo tests, leading to suggestions that they produce many positives of limited or questionable relevance. The increased sensitivity may involve factors such as enhanced exposure of cells in culture compared with target cells in vivo, higher achievable concentrations of the test article in cultures and cytotoxicity-related DNA damage. Positive results in the in vitro assay are typically followed by an in vivo test for chromosomal damage (e.g. an in vivo rodent MN assay; see section 4.5.2.5(b)) to evaluate potential in vivo mutagenicity (Kirkland et al., 2007).

(c) TK assay in mammalian cells

The TK assay in mouse lymphoma or TK6 (human) cells (OECD TG 490), described above in section 4.5.2.2(a) for its ability to detect changes in the nucleotide sequence in the *Tk/TK* gene (gene mutations), is also used as an assay for chromosomal damage. Compared with the other chromosomal damage assays, it has a much lower background and much wider dynamic range, which can make it easier in practice to differentiate a modest increase in damage from background. Some regulatory agencies, such as the United States Food and Drug Administration (USFDA, 2007), prefer this assay to other mammalian cell assays for evaluating the mutagenicity of food additives.

4.5.2.5 *In vivo chromosomal damage assays*

(a) Chromosomal aberration assay

The in vivo chromosomal aberration assay (OECD TG 475) detects structural chromosomal aberrations induced by chemical exposure in target tissues of rodents (e.g. rats, mice), most commonly the bone marrow, because of its high proliferative capacity. However, mitogen-stimulated peripheral blood lymphocytes in whole blood or as an isolated population from rodents have also been used (e.g. Au et al., 1991; Kligerman et al., 1993). The test provides an accurate assessment of induced chromosomal damage, but, like the in vitro chromosomal aberration assay (OECD TG 473; see section 4.5.2.4(a)), is labour-intensive, requiring skilled and experienced scorers, and, as commonly performed, does not accurately measure aneuploidy, a core mutagenicity end-point.

A modified version of this assay can also be performed in mammalian spermatogonial cells (OECD TG 483). The germ cell test measures chromosome- and chromatid-type structural chromosomal aberrations in dividing spermatogonial cells, but, as normally performed, is not suitable for the detection of aneuploidy. The assay is used to identify chemicals capable of inducing heritable mutations in male germ cells.

(b) Micronucleus (MN) assay

The in vivo MN test (OECD TG 474) is the most commonly used in vivo assay for chromosomal damage, as it can capture numerical and structural chromosomal changes, is not technically exacting and can be manually scored. It also lends itself to automation (flow cytometry), which speeds up data acquisition and increases the statistical power of the assay, as more cells can be readily counted (Torous et al., 2000; Dertinger et al., 2006, 2011b; MacGregor et al., 2006; Kissling et al., 2007). The standard assay evaluates MN formation in newly formed bone marrow erythrocytes of mice and rats. Modified versions of the assay can also be used in other tissues, such as the liver, spleen and colon (Morita, MacGregor & Hayashi, 2011). In most species, except mice, the spleen sequesters and destroys micronucleated erythrocytes entering the circulation, limiting the use of this assay in peripheral blood. However, this potential limitation has been overcome in a new flow cytometry version of the MN assay, which employs fluorescent dyes to identify cell surface markers (transferrin receptors) specific to immature erythrocyte populations. This ability to distinguish erythrocytes by maturation stage allows the peripheral blood MN assay to be conducted in mice, rats and a variety of other species. MN are formed primarily by direct DNA damage, although formation through indirect mechanisms resulting from cytotoxicity and hypothermia can also occur. Positive results in in vivo chromosomal damage assays correlate with rodent (and human) carcinogenicity (Witt et al., 2000). However, the standard in vivo MN assay is limited to assessing events occurring in the rapidly dividing pro-erythrocyte population in the bone marrow, so negative results should be supported by evidence that this target cell population was adequately exposed to the putative reactive parent compound or metabolite (see subsection on "Relevance" in 4.5.4.1(b)).

4.5.2.6 In vitro DNA damage/repair assays

In vitro DNA damage/repair assays have historically assessed DNA damage and repair by measuring unscheduled DNA synthesis (UDS) in cultured mammalian cells (OECD TG 482); however, based on the observation that certain OECD TGs, including OECD TG 482, are rarely used in various legislative jurisdictions and have been superseded by more sensitive tests, OECD TG 482 has been deleted by the OECD. Although information from such assays can still contribute to a WOE assessment of mutagenicity, testing of chemicals using these assays is not now recommended by the OECD (2017a). JECFA and JMPR would expect information on new substances to be based on the most up-to-date tests.

The in vitro comet assay is another approach to measuring DNA damage in vitro, although a validated OECD TG does not currently exist. Future, extended applications of the in vitro comet assay are described in section 4.5.7.2.

4.5.2.7 In vivo DNA damage/repair assays

(a) Comet (single-cell gel electrophoresis) assay

The comet assay (OECD TG 489) detects DNA damage in the form of breaks that may occur endogenously through the normal action of enzymes involved in maintaining DNA integrity, such as DNA repair processes, or may be induced by exposure to DNA-damaging agents, either directly or indirectly (through the action of DNA repair processes on chemical-induced damage). The assay detects overt double-strand and single-strand breaks as well as alkali-labile lesions (e.g. oxidized bases, alkylations, bulky adducts, crosslinks that can be converted to single-strand breaks under alkaline [pH > 13] conditions) that are visualized following electrophoresis. Furthermore, DNA strand break assays such as alkaline elution or alkaline unwinding in combination with specific DNA repair enzymes may be used to quantify specific DNA lesions, such as 8-oxoguanine. Some types of DNA breaks can be rapidly repaired, so tissues should be harvested shortly (usually 2–6 hours) after the last dose of chemical has been administered.

The comet assay is increasingly employed as a second in vivo assay to accompany the in vivo MN assay (see section 4.5.2.5(b)), as the comet assay is not limited to a rapidly dividing cell population and can be conducted with cells from virtually any tissue. For example, site-of-contact tissues can be assessed for DNA damage that depends on route of administration. There is another important distinction between in vivo chromosomal damage assays (e.g. the MN assay) and the comet assay: MN are biomarkers of chromosomal damage, which is associated with a number of adverse health outcomes in humans, and positive results correlate well with cancer in rodents and an elevated risk of cancer in humans (positive predictivity is high, but sensitivity is low). The comet assay, in contrast, is an indicator test for genotoxicity, as there are multiple fates of the DNA damage detected in this assay: accurate repair of the damage, cell death due to inability to repair, or incorrect repair, which may lead to mutation or chromosomal damage (i.e. permanent, viable, heritable change). Hence, there may be no heritable consequences of a positive finding in this assay.

The standard comet assay has a low capability of detecting some types of DNA damage (e.g. oxidative damage, crosslinks, bulky adducts). When the type of damage can be predicted, suitable modifications can be made to the assay protocol to enable the detection of such lesions. This makes the assay much more sensitive and provides additional mechanistic information. Some organs may exhibit relatively high backgrounds and variability in DNA fragmentation, and experimental conditions need to be refined for these tissues (OECD, 2014a). It should also be noted that OECD TG 489 was updated in 2016 (OECD, 2016b) to improve the reliability and robustness of this assay.

(b) DNA adduct assays

The detection and characterization of DNA adducts can provide mechanistic information on the MOA of mutagenic agents. Numerous methods can be employed, with varying degrees of specificity, and thus the choice of method should be considered on a case-by-case basis (Phillips et al., 2000; Brown, 2012). A broadly applicable and nonspecific, but highly sensitive, method is the ^{32}P-postlabelling assay (e.g. Phillips, 1997; Jones, 2012). This involves labelling of adducted nucleosides from digested DNA with ^{32}P and their quantification following chromatographic separation. A number of physical detection methods may be suitable for agents with the physicochemical properties necessary for the detection method used (e.g. fluorescence or electrochemical detection, coupled with high-performance liquid chromatography). Immunological methods have been used where antisera have been raised against carcinogen-modified DNA or against a specific adduct. Mass spectrometry has the ultimate ability to characterize and identify DNA adducts. Where it is possible to investigate radiolabelled compounds (usually with ^{14}C), accelerator mass spectrometry offers the highest sensitivity in detection, but does not provide structural information. As with the comet assay (see section 4.5.2.7(a)), there can be different fates of adducted DNA, not all of which lead to heritable changes in the cell.

(c) Unscheduled DNA synthesis (UDS) assay in mammalian liver

The UDS assay (OECD TG 486) is an indicator test that measures the synthesis of DNA outside of normal S-phase synthesis and reflects the repair of DNA damage (mainly bulky adducts repaired by nucleotide excision repair) induced by chemical or physical agents. Synthesis is commonly measured by the incorporation of tritiated thymidine into the DNA of liver cells obtained from treated and untreated rats. Although the assay has a long history of use, concerns continue to be raised about it, particularly its sensitivity to detect mutagenic agents (Eastmond et al., 2009). As explained in ECHA (2017a):

> the UDS test can detect some substances that induce in vivo gene mutation because this assay is sensitive to some (but not all) DNA repair mechanisms. However not all gene mutagens are positive in the UDS test and it is thus useful only for some classes of substances. A positive result in the UDS assay can indicate exposure of the liver DNA and induction of DNA damage by the substance under investigation but it is not sufficient information to conclude on the induction of gene mutation by

the substance. A negative result in a UDS assay alone is not a proof that a substance does not induce gene mutation.

4.5.3 Identification of relevant studies

As the assessment of mutagenicity is preferably based on all available data, an appropriate literature search should be performed. WHO (2017) guidance on systematic literature searches can be consulted for general aspects, such as selection of the database, inclusion and exclusion criteria (e.g. language(s)), documentation of search strategy and screening of the results.

Generally, information on the chemical of interest is obtained using a database such as ChemIDplus,[2] which enables combining the Chemical Abstracts Service (CAS) number, chemical names and literature search terms from databases such as PubMed. Structure searches should be performed with care and should consider stereochemistry, tautomerism, salt form and counterions, if applicable.

At a minimum, the following search terms should be used with the chemical identifier:

aneugen*	aneuploid*
"chromosom* aberration*"	clastogen*
"DNA adduct*"	"DNA damage*"
"DNA strand break*"	"gene mutation*"
"genetic damage*"	"genetic toxicity"
"genetic toxicology"	genotox*
micronucle*	mutagen*
mutation*	polyploid*

Search terms for specific tests may also be used (e.g. "in vivo comet assay*"). In addition, depending on the problem formulation, further non-pivotal assays could provide supporting information, such as:

[2] https://chem.nlm.nih.gov/chemidplus/.

“unscheduled DNA synthesis”	“DNA repair”
“sister chromatid exchange*”	“cell transformat*”

Search terms with an asterisk (*) cover all expansions of a term (e.g. mutagen* covers mutagens, mutagenicity, mutagenic, etc.). Quotation marks can be used to search for a specific term comprising two or more words (e.g. “DNA damage*”).

The main focus of the literature search is to identify the most relevant and reliable studies from those available. At a minimum, the identified data should assess gene mutations, structural chromosomal aberrations or aneuploidy. Lacking these data, the chemical is considered data poor. For data-poor chemicals with known chemical structures, read-across, structural alert, QSAR or TTC-based approaches can be considered for the evaluation and are discussed in section 4.5.5.

It may be appropriate to further limit the search, such as by language and time period, for chemicals with previous evaluations. Exclusion criteria, if applied, should be clearly described, and justification should be provided for excluded publications, for the purposes of transparency. For example, a publication lacking original data could be appropriately excluded.

Additional information sources include commercial and public databases with chemical-specific empirical data that may include associated mechanistic information or information on structurally related compounds. Some useful open-access databases are shown in Table 4.2.

For details of a testing scheme for the three mutagenicity end-points (i.e. gene mutation, clastogenicity and aneuploidy), reference should be made to the updated WHO/IPCS harmonized scheme for mutagenicity testing, described in Eastmond et al. (2009).

4.5.4 Interpretation of test results

Mutagenicity can be a hazard end-point of concern per se or a potential key event in the MOA for an adverse outcome such as carcinogenicity or developmental toxicity. Assessment of mutagenicity, both qualitatively and quantitatively, can therefore be of great value in interpreting the toxicological consequences of such adverse outcomes. Quantitatively, the potency of the response could inform the nature of the overall dose–response relationship and the implications for establishing HBGVs based on these or other effects. Qualitatively, it can add to the WOE for mutagenicity as a key event in an adverse outcome, in different species, tissues, life stages, etc.

4.5.4.1 *Presentation and categorization of results*

Criteria for the evaluation of the results of a genotoxicity test, similar to those described in the respective OECD guidelines, should be used to judge a study result as positive, negative or equivocal. In general, the result should be considered clearly positive if all three of the following criteria are fulfilled:

Table 4.2. Open-access sources of genotoxicity data (non-exhaustive list)

Database	Description
ATSDR	United States Agency for Toxic Substances and Disease Registry (ATSDR) chemical database with genotoxicity information https://www.atsdr.cdc.gov
CCRIS	Chemical Carcinogenesis Research Information System (CCRIS) database with summary carcinogenicity and genotoxicity results of studies conducted in 1985–2011 https://www.nlm.nih.gov/toxnet/Accessing_CCRIS_Content_from_PubChem.html
CTD	Comparative Toxicogenomics Database (CTD) with chemical–gene/protein interactions and gene–disease relationships http://ctdbase.org
ECHA	European Chemicals Agency (ECHA) database with summary carcinogenicity and genotoxicity study results https://echa.europa.eu
EFSA	European Food Safety Authority (EFSA) genotoxicity database for pesticide residues (290+ active substances and ~600 metabolites) https://data.europa.eu/euodp/data/dataset/database-pesticide-genotoxicity-endpoints
EURL ECVAM	Genotoxicity and Carcinogenicity Consolidated Database of Ames Positive Chemicals of the European Union Reference Laboratory for alternatives to animal testing (EURL ECVAM) http://data.jrc.ec.europa.eu/dataset/jrc-eurl-ecvam-genotoxicity-carcinogenicity-ames

Table 4.2 (continued)

Database	Description
GENE-TOX	Externally peer-reviewed data from the Genetic Toxicology Data Bank (GENE-TOX) from literature published in 1991–1998 https://www.nlm.nih.gov/toxnet/Accessing_GENETOX_Content_from_PubChem.html
IPCS INCHEM	International Programme on Chemical Safety (IPCS) database of summary documents including genotoxicity via Internationally Peer Reviewed Chemical Safety Information (INCHEM) http://www.inchem.org
IRIS	Integrated Risk Information System (IRIS) database from the United States Environmental Protection Agency (USEPA) with chemical risk assessments, including genotoxicity https://www.epa.gov/iris
ISSSTY, ISSMIC	In vitro *Salmonella typhimurium* mutagenicity (ISSSTY) and in vivo MN test results (ISSMIC) from Istituto Superiore di Sanità http://old.iss.it/meca/index.php?lang=1
Japanese NIHS Ames list	Japanese National Institute of Health Sciences (NIHS): Ames mutagenicity data for approximately 12 000 new chemicals, list of strongly positive chemicals http://www.nihs.go.jp/dgm/amesqsar.html
JECDB	Japanese Existing Chemical Data Base (JECDB) of high-production-volume chemicals, including genotoxicity studies http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp
MAK	Maximum workplace concentration (MAK) value documentations for chemical substances at the workplace, including data on genotoxicity and carcinogenicity https://onlinelibrary.wiley.com/doi/book/10.1002/3527600418
NTP-CEBS	Chemical Effects in Biological Systems (CEBS) database of United States National Toxicology Program (NTP) study results, including genotoxicity https://tools.niehs.nih.gov/cebs3/ui/
NTP-Tox21 Toolbox	Tox21 Toolbox, including the DrugMatrix toxicogenomics database and its companion ToxFX database of the United States NTP https://ntp.niehs.nih.gov/whatwestudy/tox21/toolbox/index.html
USEPA CompTox Chemicals Dashboard	Web-based dashboard integrating diverse data types with cheminformatics, with links to other sources, including genotoxicity data (e.g. USEPA IRIS, GENE-TOX, ECHA) https://comptox.epa.gov

Source: Modified from Amberg et al. (2016)

1) At least one of the test concentrations (or doses) results in a statistically significant increase compared with the concurrent negative control.

2) The increase is dose related when evaluated with an appropriate trend test.

3) Any of the results are outside the distribution of the historical negative control data (e.g. statistically based control limits).

In contrast, results are considered clearly negative if none of the three criteria is fulfilled, given a lack of major methodological deficiencies. Expert judgement or additional studies are recommended if only one or two criteria are fulfilled (i.e. the result is equivocal). Whereas these criteria could generally be applied to results from unpublished studies, which may or may not conform to an OECD TG, historical control data are rarely reported in published studies. In such cases, the reproducibility of the result should be considered when separate experiments were performed in the same study. The magnitude of the effect may also be considered. If a study result cannot be evaluated based on these three criteria, the limitations and potential uncertainties should be described.

The distinction between the terms “equivocal” and “inconclusive” by EFSA (2011) may be informative to assist in an evaluation. The term “equivocal” usually refers to a situation where not all the requirements for a clear positive or clear negative result have been met. In contrast, an “inconclusive” result is one where the lack of a clear result may have been a consequence of some limitation of the test. In this case, repeating the test under the correct conditions may produce a clear result. Similarly, the OECD (2017a) recommends that when, even after further investigations, the data set precludes a definitive positive or negative call, the test chemical response should be concluded to be equivocal (interpreted as equally likely to be positive or negative).

(a) Assessing whether results of an assay are positive, negative or equivocal for genotoxicity

Specific aspects that should be considered for the evaluation of positive and negative findings in mutagenicity/genotoxicity studies have been addressed by the European Chemicals Agency (ECHA, 2017a). These are recommended for use in JECFA and JMPR assessments, as described below.

Particular considerations when evaluating positive results include:

- testing conditions (e.g. pH, osmolality, precipitates) in in vitro mammalian cell assays and their relevance to in vivo conditions;
- factors such as the cell line, the maximum concentration tested, the measure of cytotoxicity and the metabolic activation system, which can influence specificity for in vitro mammalian cell assays;
- responses generated only at highly toxic doses or highly cytotoxic concentrations, which should be interpreted with caution (i.e. based on criteria defined in OECD TGs);
- the presence or absence of a dose (concentration)–response relationship; and
- the presence of known genotoxic impurities.

Particular considerations when evaluating negative results include:

- testing conditions (e.g. solubility of test agent, precipitates in the medium), degree of variability between replicates, high concurrent control value and widely dispersed historical control data;
- whether the doses or concentrations tested were adequately spaced and sufficiently high to elicit signs of (cyto)toxicity or reach the assay limit concentration;
- whether the test system was adequately sensitive (e.g. some in vitro assays are sensitive to point mutations and small but not large deletions);
- concerns about test substance stability or volatility;
- use of proper metabolic activation and vehicles – for example, some common diluents, such as dimethyl sulfoxide, methanol and ethanol, inhibit CYP2E1 (Busby, Ackermann & Crespi, 1999) and thus may interfere with bioactivation; and
- excessive cytotoxicity, particularly in bacterial mutation assays.

(b) Assessing data quality

Evaluation of data quality for hazard/risk assessment includes the evaluation of the adequacy, relevance and reliability of the data

(Klimisch, Andreae & Tillmann, 1997; OECD, 2005; ECHA, 2011). Relevance and reliability of study results and relevance of the test system, as they relate specifically to genotoxicity data, are described further below, as their combination helps define the adequacy of the genotoxicity database to support a conclusion on mutagenic potential for hazard/risk assessment purposes. Adequacy is discussed in section 4.5.4.3; weighting and integration of available information, which are pivotal to determining adequacy, are discussed in section 4.5.4.2. A genotoxicity database may also include specific mechanistic or MOA studies, particularly if the substance is carcinogenic or causes other relevant effects, such as developmental toxicity; these are discussed in sections 4.5.4.4 and 4.5.4.5.

Relevance of study results for a conclusion on mutagenicity. The relevance of available genotoxicity data should be evaluated based on whether the data inform one of the three mutagenicity end-points (i.e. gene mutation, clastogenicity and aneuploidy) or other genotoxic effects, with the former being more relevant and the latter considered supporting information. Some considerations that could have an impact on the relevance of the study results include the following (EFSA, 2011):

- *Purity of test substance*: Generally, test substances should have high purity, unless a substance of lower purity is more relevant to food and dietary exposures.

- *Uptake/bioavailability under testing conditions*: In certain cases, standard testing protocols (e.g. OECD TGs) may not ensure the bioavailability of test substances – for example, of poorly water-soluble substances or nanomaterials.

- *High cytotoxicity*: A positive result in mammalian cells in vitro is of limited or no relevance if observed only at highly cytotoxic concentrations.

- *Metabolism*: A negative result in an in vitro assay in which the exogenous metabolizing system does not adequately reflect metabolic pathways in vivo is of low relevance (e.g. azo-compounds, which require reduction for their activation; Suzuki et al., 2012).

- *Target tissue exposure*: A negative result from an in vivo study may have limited or no relevance if supporting information that the test substance reached the target tissue (e.g. cytotoxicity or reduced proliferation) is lacking and if there are no other data (e.g. plasma concentrations or toxicokinetics data) on which such an assumption could be based (ICH, 2011; Kirkland et al., 2019).

- *Problem formulation*: Problem formulation – that is, whether the assessment is being conducted as part of hazard classification or risk characterization – also needs to be taken into consideration here. For example, if the acceptable maximum oral dose does not give rise to significant exposure of the target tissue to either the parent compound or a bioactive metabolite, there will be no risk of mutagenicity in that tissue in vivo from dietary exposure (e.g. phenol, which undergoes efficient first-pass metabolism when administered orally; UKCOM, 2010).

- *Inconclusive results*: Inconclusive results are generally less relevant than clearly positive results; however, they may suggest mutagenic potential, which should be clarified by further testing, as recommended by OECD TGs. Some modification of the experimental conditions may be necessary when repeating the study – for example, to allow for the possible absence of enzymes of activation in the original test.

When the available data preclude an assessment of the potential to induce gene mutations, clastogenicity and aneuploidy, the outcome of the literature search may be described narratively, with the most notable limitations specified.

Reliability of study results for a conclusion on mutagenicity. Factors to be considered in assessing the reliability of a study include the following:

- Were the results with concurrent positive and negative controls, cell growth characteristics, etc., consistent with expectations based on published ranges (Lorge et al., 2016; Levy et al., 2019)?

- Was the highest dose/concentration adequate based on the upper concentration or cytotoxicity limit described in the relevant TGs?

- For mammalian cell assays limited by cytotoxicity, were data available from concentrations at both low and moderate levels of cytotoxicity, as described in the relevant TGs?

- When the initial test result was inconclusive due to a modest response near a limit dose/concentration, was the test repeated using appropriate protocol modifications (OECD, 2017a; Levy et al., 2019)?

- Was the test conducted under currently acceptable protocols? The OECD recommends consideration of results from any test conforming to the TG in effect at the time the test was conducted, but such data may be less reliable than those from studies conducted according to current guidelines. This applies equally to published studies.

Some approaches for evaluating reliability, although not specific to genotoxicity, include the Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE) Risk of Bias tool for animal studies (Hooijmans et al., 2014), the Toxicological data Reliability Assessment Tool (ToxRTool) (Schneider et al., 2009) and Science in Risk Assessment and Policy (SciRAP) (Molander et al., 2015; Beronius & Ågerstrand, 2017). Klimisch, Andreae & Tillmann (1997) provided a classification approach, including 1) Reliable without restriction, 2) Reliable with restrictions, 3) Not reliable and 4) Reliability not assignable. The resulting classifications are often referred to as “Klimisch scores”. The approaches described here may be particularly helpful when assessing unpublished studies based on secondary sources. However, the value of the information obtained from their use for primary study reports, including peer-reviewed literature, should be considered on a case-by-case basis, based on the problem formulation and given the resource-intensive nature of such approaches. The choice of whether to use a formal scoring system, and, if so, which one, should be decided on a case-by-case basis, and a clear explanation should be provided for the decisions made.

The type of document (e.g. published or unpublished study report) and TG or GLP conformance do not necessarily have an impact on reliability. Adequate data reporting is more relevant, recognizing that the quality of articles published in peer-reviewed journals is significantly higher than the quality of articles published in non-peer-reviewed journals. It is also recognized that for regulated substances, such as food additives or pesticides, appropriate data can be requested from the petitioner or producer; this is not possible for substances such as food contaminants, for which the evaluation is performed based on available data and assessment approaches such as read-across from similar chemicals and (Q)SAR.

Relevance of the test system. The relevance of the test system (high, limited or low) to conclusions on mutagenicity is based on the genetic end-point, with gene mutations, clastogenicity and aneuploidy considered of high relevance. The in vivo comet assay, which detects DNA damage, is also generally considered to be of high relevance as supporting information. Similarly, measurement of DNA adducts, as supporting information, may be considered of high (or lower) relevance, depending, for example, on the methodology used to assess their occurrence and on the types of adducts induced (e.g. bulky adduct). Other tests of limited or low(er) relevance may also provide useful supporting information. The available studies should be categorized according to the end-point assessed. For chemicals in food, results from oral in vivo genotoxicity studies are generally preferred to data obtained through exposure by non-oral routes, such as intraperitoneal, dermal or inhalation routes.

Presentation of results. If data to assess gene mutations, clastogenicity or aneuploidy are available, it is useful to tabulate the results grouped by end-point, as described in the JMPR *Guidance document for WHO monographers and reviewers* (WHO, 2015a), with columns on 1) Reliability/comments, 2) Relevance of the test system and 3) Relevance of the study result. Tables reporting in vivo studies should include the test system (e.g. bone marrow MN assay; 10 12-week-old male B6C3F1 mice per dose), route (e.g. oral gavage, feed, intraperitoneal), dose (in mg/kg body weight; if only the concentration in feed or drinking-water is reported), result (as reported by the study author(s)) and reference, as well as the three additional columns mentioned above.

The result should be presented as judged by the genotoxicity experts/reviewers, preferably as positive, negative, equivocal or inconclusive. Discordance between judgements of the genotoxicity experts/reviewers and those of the study authors should be described (e.g. in the Comments section of JECFA/JMPR evaluations).

Generally, the quality of a study result is based on its reliability and on the relevance of the test system. Conformance to Good

Laboratory Practice (GLP) can also provide confidence related to study protocol and standard operating procedure, but should not be a reason for exclusion a priori. Only the relevant and reliable studies should be tabulated, rather than an exhaustive list. Studies considered to have low relevance of both the test system and the study result should be omitted. The relevance of the study result is low if either the reliability is low (e.g. a Klimisch score of greater than 2) or the relevance of the test system is low (or both).

Any limitation that results in or contributes to a judgement of limited or insufficient reliability should be described in the "Reliability/comments" column. As an example of how studies might be scored and the factors to be considered, in the Klimisch, Andreae & Tillmann (1997) classification approach, a reliability score of 2 (Reliable with restrictions) indicates that although the results in general are scientifically acceptable, the study does not conform to a TG, and hence there will be some uncertainties in the methodology. A score of 3 (Not reliable) indicates that there were either methodological deficiencies or aspects of the study design that were not appropriate, such as inappropriate doses, lack of appropriate controls, inappropriate solvent/carrier, insufficient protocol details, inappropriate data analysis, unreported source and purity of chemical, use of a chemical mixture (unless target substance) and potential for bias (e.g. samples not analysed blind); and, for human studies, uncharacterized or mixed exposures, inappropriate sampling times, etc. A score of 4 (Not assignable) indicates a report that provides insufficient information for data assessment, such as a report with no original data or a conference abstract without subsequent full publication.

Conflicting results in more than one test with similar reliability should be judged for whether the differences might be attributable to different test conditions (e.g. concentrations, animal strains, cell lines, exogenous metabolizing systems). Without a plausible explanation, the data may be of limited use, and a further study may provide clarification.

Recommended templates for the reliability and relevance of a test system and study results are provided for in vitro studies (Table 4.3) and in vivo studies (Table 4.4).

Table 4.3. In vitro study table showing recommended columns for reliability and relevance

Test system	Concentrations	Result	Reference(s)	Klimisch reliability/ comments	Relevance of test system	Relevance of study result
				1	High	High
				2	High	Limited
				3	High	Low
				4	High	Low
				1	Limited	Limited
				2	Limited	Limited
				3	Limited	Low

Table 4.4. In vivo study table showing recommended columns for reliability and relevance

Test system	Route	Doses	Result	Reference	Klimisch reliability/ comments	Relevance of test system	Relevance of study result
					1	High	High
					2	High	Limited
					3	High	Low
					4	High	Low
					1	Limited	Limited
					2	Limited	Limited
					3	Limited	Low

A general footnote can be included to indicate that studies with low relevance of both the test system and the study result have been omitted. After the data are tabulated, the most notable data gaps, whether in vitro or in vivo, that have an impact on the evaluation should be discussed narratively.

4.5.4.2 *Weighting and integration of results*

In assessing mutagenicity specifically and the broader concept of genotoxicity in general, a WOE approach should be used, with considerations for elements such as relevance and reliability of study results and relevance of the test system, as described in section 4.5.4.1(b) above, reproducibility and consistency, significance and mechanism of the genetic alteration, phylogenetic relationship to humans, study type (i.e. in vivo or in vitro) and physiological relevance of the dose and route of administration with respect to human exposures (see below in this section and Eastmond, 2017 for additional details). In applying this guidance, reviewers should have flexibility in evaluating all relevant scientific information in order to apply best scientific judgement to reach conclusions about the significance of the genotoxicity results. The WOE approach should account for the key genetic end-points (i.e. gene mutations, clastogenicity and aneuploidy) and the appropriateness of in vivo follow-up for positive in vitro results.

Studies with the following characteristics are generally given the greatest weight in assessing human health risks, although all appropriate studies should be considered:

- highly relevant and reliable studies, as described in section 4.5.4.1(b); the studies should not be in draft form and should have sufficient detail for a thorough review;
- results that have been independently reproduced;
- studies measuring key end-points of mutagenicity (i.e. gene mutations, clastogenicity and aneuploidy);
- studies using accepted and validated models and protocols, with proper negative and positive controls within historical ranges, protections against bias (e.g. coding and blind scoring of slides, randomization of animals for treatment), chemical purity known and within an acceptable range, and proper statistical analyses;
- studies measuring genotoxicity in a known or suspected target organ;

- in vivo studies in humans, other mammals or other species known or likely to respond similarly to humans;
- human studies with well-characterized exposures and an absence of co-exposures or other potential confounders;
- studies conducted using an exposure route physiologically relevant to the problem formulation (i.e. oral, dermal or inhalation; studies by the oral route are preferred when evaluating chemicals present in the diet) and under other conditions (e.g. acceptable concentrations/doses, levels of toxicity and diluents, absence of co-exposures) within generally accepted guidelines;
- studies in which the damage has been well characterized or identified (e.g. specific DNA adducts derived from the chemical of interest have been identified); and
- studies involving bioactivation systems known or likely to mimic bioactivation in humans or those known to be involved in the bioactivation of similar compounds.

In contrast, little or no weight is given to DNA damage or other types of genotoxicity occurring through mechanisms for which there is sufficient evidence that these will not occur or are highly unlikely to occur in humans. For example, DNA damage occurring in the bladder of saccharin-treated rats secondary to urinary crystal formation (USNTP, 2011) and DNA damage occurring as a consequence of or secondary to toxicity, such as during the cytotoxic phase in male rat kidney cells following exposure to a chemical that binds to and induces α2u-globulin nephropathy (Swenberg, 1993), are weighted less in an evaluation (Eastmond, 2017). Although the comet assay can provide valuable information, positive results alone (i.e. with no positive results in assays for any of the mutagenic end-points) should be viewed with caution, given the fact that the assay detects only overt or alkali-induced DNA strand breaks and, in itself, is unable to establish the mechanism for the strand break (see also section 4.5.2.7(a) above).

In many cases, substances exhibit a positive result in more than one assay or test system. However, a single, clear positive mutagenicity result in a relevant and reliable study may, at times, be sufficient to conclude that a substance is mutagenic, without other evidence of genotoxicity. This will depend on expert judgement. Contrasting results for the same end-point in studies using comparable methodology should be evaluated on a case-by-case basis using the weighting considerations outlined above.

As indicated above, assessing study quality includes determining whether the study was conducted according to standard guidelines and protocols, such as those published by the OECD (see http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788). Guideline-compliant studies are generally considered relevant and reliable and weighted more in an evaluation. Conversely, deficiencies or other limitations with respect to the guidelines should be noted. The decisions on the relevance and acceptability of non-compliant or pre-guideline studies may require particular attention and expert judgement, particularly when guideline studies exist.

Another consideration is that, as noted above, the results should be reproducible. The strength of a finding is increased if the same result has been demonstrated in different laboratories. An observation made in a single laboratory – even if repeated on separate occasions – may be viewed with less confidence than one that has been reproduced in other laboratories.

Another consideration is whether a consistent pattern exists. The observed results should be plausible given the known mechanisms of toxicity or action of the agent. It is anticipated that a substance that is clastogenic in vivo would also be clastogenic in vitro and that an agent that is clastogenic in somatic cells in vivo would also be clastogenic in germ cells (with appropriate toxicokinetic or sex considerations, if applicable). Deviations from the expected pattern should be scrutinized with special care. Inferences with regard to mutagenicity in vitro versus in vivo have been limited owing to the few adequately validated in vivo mutagenicity tests. It is recognized that this situation has improved in recent years with the increased use of the transgenic and *Pig-a* mutation models.

An additional consideration is the purity of the substance used in the different studies. The amount of impurity present in the material tested should be compared with the amount specified in the technical material. This information should be used when assessing the relevance of the results from different studies. Where concern exists about the mutagenicity of an impurity, approaches described elsewhere in this document should be considered, including application of a TTC approach.

The WOE evaluation should also note whether evidence exists to support a biological threshold or alternative, non-mutagenic MOAs for the adverse effects observed, such as cancer or developmental toxicity (discussed in further detail below in section 4.5.4.4), and whether structural relationships to known mutagenic substances exist, to identify data gaps and uncertainties. The evaluation should ultimately enable a final conclusion on genotoxicity and, more specifically, mutagenicity (described further in section 4.5.4.3).

4.5.4.3 Adequacy of the genotoxicity database

After a critical review of relevant and reliable genotoxicity data has been completed, WHO (2015a) recommends that a conclusion on the genotoxic risk to humans be included based on standard phrases for defined scenarios. For example, when a compound "has been tested for genotoxicity in an adequate range of in vitro and in vivo assays" and "no evidence of genotoxicity is found", it is acceptable to conclude that the compound "is unlikely to be genotoxic". Recent examples are abamectin (WHO, 2016), tioxazafen (WHO, 2019) and pyriofenone (WHO, 2019). It is important to note that when JMPR and JECFA use the term genotoxic(ity), in most instances they are referring to mutagenic(ity), as defined in this section of EHC 240. Hence, it is recommended that the terms "genotoxic" and "genotoxicity" in the above standard phrases be changed to "mutagenic" and "mutagenicity", as appropriate.

In contrast, the database can be considered "inadequate" to allow a conclusion on genotoxicity after review of the available in vivo and in vitro genotoxicity data for the compound. For example, JECFA was unable to complete the evaluation of the copolymer food additive anionic methacrylate copolymer (FAO/WHO, 2018); although the copolymer itself was not a health concern, JECFA noted that there were insufficient data to conclude on the genotoxic potential of the residual monomer, methyl acrylate, and requested further studies to clarify its in vivo carcinogenic potential.

For chemicals of interest (e.g. residues or contaminants) that lack data from the minimum range of tests (i.e. an indication of their ability to induce gene mutations, clastogenicity and aneuploidy), it is necessary to evaluate their mutagenicity using (Q)SAR, read-across or TTC-based approaches (see section 4.5.5).

There is considerable flexibility in the description when positive or equivocal test results exist (WHO, 2015a). For example, when tested in an adequate range of in vitro and in vivo assays, the compound "gave a positive/equivocal response in the in vitro [names of end-point/assay], but it was negative in the in vivo [names of end-point(s)/assay(s)]". The data may also support a more specific conclusion, such as the compound is "unlikely to be genotoxic in vivo", followed by the primary rationale. For example, JMPR found no evidence of genotoxicity in numerous in vivo assays for acetochlor (96% purity), despite weak mutagenicity in vitro with less pure material (89.9% purity) and clastogenicity occurring at cytotoxic concentrations; recognizing the lack of a specific assay for gene mutations in vivo, JMPR concluded that, on the basis of the WOE, acetochlor was unlikely to be genotoxic in vivo (WHO, 2016). It is expected that positive results in vitro would be followed up by an appropriate in vivo assay for the respective end-point. As mentioned in section 4.5.2, the comet assay (OECD TG 489) and transgenic rodent assays (OECD TG 488) are being increasingly employed as a second in vivo assay to accompany the in vivo MN assay (OECD TG 474).

Exposure context, such as whether the observed mutagenicity would be expected to occur in humans exposed to low-level pesticide residues in food, should also be considered (Eastmond, 2017). It is useful to specify the exposure route that was considered in the overall evaluation, such as through the diet, by the dermal route or by inhalation, when concluding on mutagenic potential.

4.5.4.4 *Mutagenic mode of action and adverse outcomes*

The WOE conclusion on mutagenicity can be used to help interpret available data on specific adverse outcomes in humans or laboratory animals, particularly carcinogenicity and developmental toxicity. The default assumption in hazard and risk characterization has been that if the substance is mutagenic, then this is its MOA as a carcinogen. This policy decision has driven the manner in which mutagenic carcinogens are dealt with in national and international regulatory arenas and assumes that a single mutation in a single relevant gene (e.g. oncogene) could cause oncogenic transformation; therefore, it is reasoned, there can be no DNA damage threshold that is without consequence and, hence, no safe level of exposure to a mutagenic carcinogen. However, recent studies challenge this linear, non-threshold or "one-hit" theory of carcinogenesis, and experimental thresholds have been observed for some DNA-reactive mutagenic carcinogens (Kobets & Williams, 2019). For example, studies on chromosomal damage and gene mutations in mice repeatedly exposed to the mutagen ethyl methanesulfonate demonstrated a clear, practical threshold or no-observed-genotoxic-effect level (NOGEL) (Pozniak et al., 2009). Thus, even for DNA-reactive mutagens, non-linear, threshold-type dose–response curves can be seen. For all mutagens, there may be a level of exposure below which chemical-induced mutation levels cannot be distinguished from background (spontaneous) mutation levels, which are tightly monitored by endogenous systems designed to control cellular perturbations, including DNA damage, caused by exogenous and endogenous stressors. In reaching a conclusion on the nature of the dose–response relationship and its linearity or otherwise, all relevant information on toxicokinetics and toxicodynamics should be considered, as described by Dearfield et al. (2002, 2011, 2017). In most cases, however, the available evidence is insufficient to enable a conclusion on the existence of a threshold, and the risk assessment should proceed as if there is no threshold. This is because even should a threshold exist, there would be considerable uncertainty, potentially by orders of magnitude, as to the dose at which it occurs.

For substances that do not react with DNA, such as those that affect spindle function and organization, inducing aneuploidy, or chromosome integrity through topoisomerase inhibition, threshold-based mechanisms may be proposed. Other examples of mutagenic mechanisms that may be characterized by non-linear or threshold dose–response relationships include extremes of pH, ionic strength and osmolarity, inhibition of DNA synthesis, alterations in DNA repair, overloading of defence mechanisms (antioxidants or metal homeostasis), high cytotoxicity, metabolic overload and physiological perturbations (e.g. induction of erythropoiesis) (Dearfield et al., 2011; OECD, 2011). Nevertheless, some indirect interactions that may give rise to non-linear dose–response curves can occur at very low exposures, such as for arsenite carcinogenicity, where DNA repair inhibition has been reported to occur at very low, environmentally relevant concentrations (Hartwig, 2013).

Determining that a substance is mutagenic is not sufficient to conclude that it has a mutagenic MOA for an adverse outcome (Cimino, 2006). A WOE approach that applies various weights to different end-points or assays is recommended when evaluating whether a substance is likely to act via a mutagenic MOA. The level of evidence is specific to the end-point that the assay is evaluating and thus needs to be considered along with all available evidence to conclude on the overall likelihood of a mutagenic MOA. Expert judgement is necessary with respect to the data quality described in section 4.5.4.2 (i.e. relevance, reliability, adequacy). For example, some factors that provide more weight include the following:

- The substance is mutagenic in the target organ or system in which the adverse outcome was observed.
- The substance is DNA reactive, or there is significant conversion to a DNA-reactive intermediate that is confirmed to be associated with the adverse outcome.
- There is evidence of substantial covalent binding to DNA, preferably in vivo in the target tissue or system.
- The substance is a multiroute, multisite and multispecies carcinogen in animal bioassays, particularly if tumours arise in tissues that do not have high spontaneous incidences or are not hormonally sensitive.
- There is evidence that the substance acts as an initiator in a well-conducted rodent tumour initiation:promotion assay.
- Highly similar structural analogues produce the same, or a pathologically closely related, adverse outcome via a mutagenic MOA; the WOE is increased if the substance contains structural alerts for DNA mutagenicity and reactivity.

Factors that stimulate cell replication (e.g. classical tumour promoters in the case of carcinogenicity, which stimulate growth of initiated cells), epigenetic alterations (e.g. DNA/histone methylation) and non-mutagenic or indirectly mutagenic (i.e. non-DNA-reactive) events are important in certain adverse outcomes (e.g. cancer, developmental toxicity) in both experimental animals and humans. Indirectly mutagenic MOAs that are particularly relevant involve interactions with proteins (including enzymes) involved in maintaining genomic stability, such as inhibition of DNA repair processes, tumour suppressor functions, cell cycle regulation and apoptosis. Some of these mechanisms may lead indirectly to an increase in mutant frequency – for example, by an accumulation of

DNA lesions induced by endogenous processes or by exogenous DNA-reactive agents due to diminished repair. Also, accelerated cell cycle progression due to impaired cell cycle control may reduce the time for DNA repair and thus increase the risk of mutations during DNA replication. For some classes of compounds, such as some carcinogenic metal compounds, such interactions have been observed at particularly low concentrations and thus appear to be relevant under low-exposure conditions (e.g. Hartwig, 2013).

Epigenetic alterations refer to changes in gene expression without alterations in DNA sequences. They include alterations in DNA methylation patterns, in histone and chromatin modifications, in histone positioning and in non-coding RNAs. Disruption can lead to altered gene function, such as activation of proto-oncogenes or inactivation of tumour suppressor genes. Thus, epigenetic alterations can contribute to the initiation and progression of some adverse outcomes, such as cancer (for review, see Kanwal, Gupta & Gupta, 2015). Again, for carcinogenic metal compounds such as arsenic, nickel and chromium, epigenetic alterations appear to be a major mechanism contributing to carcinogenicity (e.g. Beyersmann & Hartwig, 2008; Chervona, Arita & Costa, 2012; Costa, 2019). From a risk assessment point of view, these MOAs are usually thought to exhibit a threshold, which, in principle, would, at low doses, protect against the respective adverse outcome. However, the no-observed-adverse-effect level (NOAEL) in humans is frequently unknown and may be very low, occurring sometimes even at background exposure levels of the general population, as is believed to be the case for arsenic (e.g. Langie et al., 2015). In general, however, such information would more inform the WOE than contribute directly to the risk assessment.

DNA-reactive, epigenetic and non-DNA-reactive mechanisms can cooperate in inducing an adverse outcome. Indeed, epigenetic changes often occur as a result of initial mutagenic events (see Nervi, Fazi & Grignani, 2008).

4.5.4.5 Integration of carcinogenicity and mutagenicity

JECFA and JMPR integrate information on mutagenicity and carcinogenicity, together with all other relevant data, to reach an overall conclusion on carcinogenic risk. Similar to the standard phrases for mutagenic potential mentioned in section 4.5.4.3, standard phrases with defined scenarios for chemicals with mutagenicity and carcinogenicity evaluations may include the following (adapted from WHO, 2015a, to reflect the updated guidance in this section of EHC 240). It should be noted that the wording for the conclusions on specific substances is taken from the respective meeting reports. It is anticipated that future conclusions of JMPR and JECFA will reflect the recommendations in this section of EHC 240:

[compound not carcinogenic or mutagenic]

In view of the lack of mutagenicity and the absence of carcinogenicity in mice and rats, it is concluded that [compound] is unlikely to pose a carcinogenic risk to humans.

For example, the evaluation of chlormequat by JMPR in 2017 (FAO/WHO, 2017a) noted that "In view of the lack of genotoxic potential and absence of carcinogenicity in mice and rats, the Meeting concluded that chlormequat is unlikely to pose a carcinogenic risk to humans."

or

[compound not carcinogenic or mutagenic in vivo with positive in vitro mutagenicity]

In view of the lack of mutagenicity in vivo and the absence of carcinogenicity in mice and rats, it is concluded that [compound] is unlikely to pose a carcinogenic risk to humans at levels occurring in the diet.

For example, the evaluation of flufenoxuron by JMPR in 2014 (WHO, 2015b) noted that "In view of the lack of genotoxicity in vivo and the absence of carcinogenicity in mice and rats at exposure levels that are relevant for human dietary risk assessment, the Meeting concluded that flufenoxuron is unlikely to pose a carcinogenic risk to humans from the diet."

or

[compound carcinogenic but not mutagenic]

In view of the lack of mutagenicity, the absence of carcinogenicity in [species] and the fact that only [tumours] were observed and that these were increased only in [sex] [species] at the highest dose tested, it is concluded that [compound] is unlikely to pose a carcinogenic risk to humans from the diet. *[There is considerable flexibility in wording here.]*

For example, the evaluation of ethiprole by JMPR in 2018 (WHO, 2019) noted that “In view of the lack of genotoxicity and the fact that tumours were observed only at doses unlikely to occur in humans, the Meeting concluded that ethiprole is unlikely to pose a carcinogenic risk to humans via exposure from the diet.”

or

[compound carcinogenic with positive in vitro mutagenicity]

As [compound] was not mutagenic in vivo and there is a clear NOAEL for [tumour type] in [sex] [species], it is concluded that [compound] is unlikely to pose a risk for carcinogenicity to humans from the diet. *[There is considerable flexibility in wording here.]*

For example, the evaluation of fenpicoxamid by JMPR in 2018 (WHO, 2019) noted that “As fenpicoxamid is unlikely to be genotoxic in vivo and there is a clear threshold for liver adenomas in male mice and ovarian adenocarcinomas in female rats, the Meeting concluded that fenpicoxamid is unlikely to pose a carcinogenic risk to humans from the diet.”

or

[compound carcinogenic with positive in vitro and in vivo mutagenicity]

As [compound] is mutagenic in a variety of in vivo and in vitro tests and there is no clear NOAEL for [tumour type] in [sex] [species], it is concluded that [compound] should be considered a carcinogen acting by a mutagenic MOA.

or

[compound lacks carcinogenicity data]

If a compound lacks carcinogenicity data or has carcinogenicity data with major limitations, with or without adequate genotoxicity data, it should be noted that a conclusion on carcinogenic potential cannot be reached, and the major limitations of the existing database should be specified. In such a case, establishment of an HBGV may not be appropriate if adequate genotoxicity data are available to support a WOE conclusion that the substance is mutagenic in vivo.

For example, the evaluation of natamycin by JMPR in 2017 (FAO/WHO, 2017b) noted that "In view of the limitations in the available database on carcinogenicity and genotoxicity, the Meeting determined that no conclusions can be drawn on the carcinogenic risk to humans from the diet." JMPR did not establish an ADI or an ARfD for natamycin owing to the inadequate database available to the Meeting. Alternatively, if adequate data on genotoxicity are available, it may be possible to use a WOE approach to reach a conclusion on risk of carcinogenicity from exposure via the diet, even in the absence of data from carcinogenicity bioassays.

The above phrases are intended to cover all standard scenarios that might be encountered in evaluating the carcinogenic potential of a substance. Where no suitable phrase exists, additional phrases will be developed by JMPR and JECFA as necessary.

As with any outcome addressed by JECFA or JMPR, due consideration should be given to the evaluation and communication of major sources of uncertainty in the assessment of mutagenicity. Guidance is available in section 7.2.2 and elsewhere in EHC 240 and in IPCS (2018).

4.5.5 Approaches for evaluating data-poor substances

4.5.5.1 *In silico approaches*

In the regulatory arena, QSAR methods are used to predict bacterial mutagenicity (as well as other end-points). These have been used for drug impurities lacking empirical data, as described in the International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) M7 guidelines (ICH, 2014, 2017) (see Sutter et al., 2013; Amberg et al., 2016; Wichard, 2017). (Q)SAR and read-across approaches[3] have been used (see WHO, 2015a), or have been proposed for use, to assess the genotoxicity of pesticide residues (degradation products and metabolites) for dietary risk assessment (see Worth et al., 2010; EFSA, 2016a). QSAR models are also applied under the aegis of the

[3] For a more detailed explanation of these terms, see Patlewicz & Fitzpatrick (2016).

EU Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation, most commonly, although not exclusively, to support WOE approaches for mutagenicity prediction (e.g. REACH Annex VII).

(a) Available tools (QSARs, SARs/structural alerts) for mutagenicity

In silico approaches pertaining to genotoxicity typically comprise QSARs, SARs (often referred to as structural alerts) and "expert systems", the last comprising QSARs, SARs or both. Expert systems are categorized as statistical (QSAR) or knowledge based (SAR) or hybrids (Patlewicz et al., 2014).

Relative to other hazard end-points, structural alerts for mutagenicity, particularly for DNA-reactive gene mutagenicity, are the most established, and many software tools exist to identify them. The breadth and scope of structural alert schemes may differ between different tools, with the quantity of alerts within a given tool not necessarily being the best or most useful measure of the coverage of the alerts or their performance. The majority of structural alerts available have been derived from Ames test data, although alerts and QSARs are also available for gene mutations in mammalian cells, chromosomal aberrations, MN formation and DNA binding, all of which contribute to mutagenicity assessment – for example, to determine the TTC tier (see section 4.5.5.2). In silico models and tools and the data availability for model development for different mutagenicity end-points have been recently reviewed (Benigni et al., 2019; Hasselgren et al., 2019; Tcheremenskaia et al., 2019). Table 4.5 provides examples of genotoxicity assessment approaches within commercial, open-source or freely available software.

(b) Confidence in approaches

When applying (Q)SAR models, an important consideration is the decision context that will inform the level of confidence needed from one or more models. For example, a different degree of confidence may be required for:

- screening and prioritization of chemicals for further evaluation;
- hazard characterization or risk assessment;
- classification and labelling (under the Globally Harmonized System of Classification and Labelling of Chemicals); and

Table 4.5. Examples of commercial, freely available or open-source in silico tools

Type of model	Effects	Software/availability	Link/reference
Expert system – knowledge based	Alerts for mutagenicity, also subcategorized for chromosomal effects and gene mutations	Derek Nexus – commercial	https://www.lhasalimited.org/products/derek-nexus.htm https://www.lhasalimited.org/products/ICH-M7-assessment-using-derek-nexus.htm
	Alerts to assign concern levels for carcinogenicity	USEPA OncoLogic cancer tool – freely available	https://www.epa.gov/tsca-screening-tools/oncologictm-computer-system-evaluate-carcinogenic-potential-chemicals (USEPA, 2019)
Expert system – hybrid – mix of QSARs and knowledge underpinned by a metabolism simulator	Ames mutagenicity In vitro chromosomal aberration In vivo MN induction In vivo liver genotoxicity In vivo liver transgenic rodent mutagenicity In vivo liver clastogenicity Comet genotoxicity	TIMES – commercial	http://oasis-lmc.org/products/models/human-health-endpoints/?page=2& (Patlewicz et al., 2007)

Type of model	Effects	Software/availability	Link/reference
Expert system – statistical	Various genotoxicity end-points	Leadscope Model Applier – commercial	http://www.leadscope.com/product_info.php?products_id=67
	Various genotoxicity end-points	CASE Ultra – commercial	http://www.multicase.com/case-ultra-models
	Various genotoxicity end-points ("Impurity Profiling Module")	ACD/Percepta – commercial	https://www.acdlabs.com/products/percepta/index.php
	Various genotoxicity end-points	ChemTunes ToxGPS – commercial	https://www.mn-am.com/products/chemtunestoxgps
	Ames mutagenicity	Biovia Discovery Studio – commercial	https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/
	Ames mutagenicity	LAZAR – freely available	https://openrisknet.org/e-infrastructure/services/110/
	Ames mutagenicity	USEPA T.E.S.T. – freely available	https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NRMRL&dirEntryId=232466

Table 4.5 (continued)

Type of model	Effects	Software/availability	Link/reference
Expert system – statistical (continued)	Ames mutagenicity	Sarah Nexus – commercial	https://www.lhasalimited.org/products/sarah-nexus.htm
	Ames mutagenicity	VEGA – freely available	https://www.vegahub.eu/
	Chromosomal aberration	ADMET Predictor – commercial	https://www.simulations-plus.com/software/admetpredictor/toxicity/
Read-across tools – also incorporate WOE QSAR results	Ames mutagenicity	ToxRead – open source	https://www.vegahub.eu/download/toxread-download/
Chemoinformatics system with databases, in silico models and supporting read-across	Prediction tools integrated (e.g. Ames mutagenicity, Toxtree, VEGA models)	AMBIT (Cefic-LRI) – freely available	http://cefic-lri.org/toolbox/ambit/

Type of model	Effects	Software/availability	Link/reference
SARs/structural alerts	Carcinogenicity rules based by ISS (incorporates Ashby–Tennant rules), ISS in vitro Ames test alerts and in vivo mutagenicity (MN); DNA binding alerts (also implemented as DNA binding for OECD in QSAR Toolbox)	Toxtree – open source	https://ec.europa.eu/jrc/en/scientific-tool/toxtree-tool
Profilers – rule based on structural alerts to facilitate grouping of substances for read-across	DNA binding for OECD, DNA binding for OASIS, DNA alerts for Ames, chromosomal aberrations and MN by OASIS, Benigni/Bossa (ISS) alerts for in vitro mutagenicity Ames and in vivo mutagenicity (MN)	OECD QSAR Toolbox – freely available	https://qsartoolbox.org/

ACD: Advanced Chemistry Development, Inc.; Cefic: European Chemical Industry Council; DNA: deoxyribonucleic acid; ISS: Istituto Superiore di Sanità; MN: micronucleus/micronuclei; OECD: Organisation for Economic Co-operation and Development; QSAR: quantitative structure–activity relationship; T.E.S.T.: Toxicity Estimation Software Tool; TIMES: tissue metabolism simulator; USEPA: United States Environmental Protection Agency; WOE: weight of evidence

- addressing specific information requirements depending on regulatory jurisdiction (e.g. EU REACH vs Korea REACH).

(Q)SAR models should follow the OECD (2007) principles for validation to be considered of high quality. When applying a (Q)SAR, it is important that the substance being assessed is within the intended scope of the model – that is, the model is underpinned by substances of like chemistry. Generally, the predictivity of (Q)SAR models is closely related to the data available for model development and their quality. The aim of a recent project was to improve the quality of Ames data as the basis of related (Q)SAR models by extending the data sets with new data and re-evaluating historic Ames test results (Honma et al., 2019).

The performance of different in silico approaches for mutagenicity prediction has been reviewed elsewhere (see Netzeva et al., 2005; Serafimova, Fuart-Gatnik & Worth, 2010; Hanser et al., 2016), including analyses specifically for food ingredients, food contact materials and pesticides (e.g. Worth et al., 2010; Bakhtyari et al., 2013; Cassano et al., 2014; Greene et al., 2015; Vuorinen, Bellion & Beilstein, 2017; Van Bossuyt et al., 2018; Benigni et al., 2019). General aspects of confidence in and applicability of (Q)SAR models have also been reviewed recently, providing a list of guiding assessment criteria (Bossa et al., 2018; Cronin, Richarz & Schultz, 2019).

Quantitative consensus models and expert judgement can be used to deal with multiple QSAR predictions by leveraging the strengths and compensating for the weaknesses of any individual model and quantifying uncertainties in the predictions. For instance, Cassano et al. (2014) evaluated the performance of seven freely available QSAR models for predicting Ames mutagenicity and found that a consensus model outperformed individual models in terms of accuracy. A strategy for integrating different QSAR models for screening and predicting Ames mutagenicity in large data sets of plant extracts has recently been proposed (Raitano et al., 2019). Large-scale, collaborative, consensus model–building efforts have also been undertaken for other end-points, substantiating the benefits of improved performance of consensus models over individual models and the use of a common, harmonized training data set – for example, in vitro estrogenic activity (Mansouri et al., 2016) and acute oral toxicity (Kleinstreuer et al., 2018).

Different perspectives exist on how to combine predictions from one or more models and how to resolve discordant predictions, with some form of expert review and judgement applied to conclude on divergent results (Greene et al., 2015; Powley, 2015; Wichard, 2017). Expert review can also be applied to resolve cases of equivocal and out-of-domain predictions (see Amberg et al., 2019) and is discussed generally in Dobo et al. (2012), Barber et al. (2015), Powley (2015), Amberg et al. (2016) and Myatt et al. (2018). The expert review in a WOE approach can include analogue information (i.e. read-across; see section 4.5.5.3) (Amberg et al., 2019; Petkov et al., 2019).

A decision workflow has been proposed by the international In Silico Toxicology Protocol initiative led by Leadscope Inc. (see Myatt et al., 2018; Hasselgren et al., 2019), which is based on a combination of different experimental and in silico evidence lines to arrive at an overall conclusion about the mutagenic hazard of a substance. This approach includes Klimisch scores extended to more general reliability scores in order to include assessment of in silico results, taking account of consistency of prediction and expert review. In this scheme, in silico results cannot be assigned a score better than 3 (i.e. <3) (Table 4.6).

(c) Mutagenicity assessment

In the context of the present guidance, in silico approaches for mutagenicity assessment can be used (see also Fig. 4.1, boxes 17 and 22):

- When empirical data on a compound are insufficient to reach a conclusion on mutagenicity, additional information should be sought from related analogues (i.e. read-across; see section 4.5.5.3) and in silico approaches (e.g. (Q)SARs) and considered in an overall WOE evaluation of mutagenic potential (see also section 4.5.4.2).

- In silico approaches can be used as the basis for application of the TTC approach, depending on the presence or absence of structural alerts for DNA-reactive mutagenicity (or WOE that the substance might be mutagenic) to determine the TTC tier applied (see section 4.5.5.2).

Table 4.6. Reliability of (geno)toxicity assessments based on in silico models and experimental data

Reliability score	Klimisch score[a]	Description	Summary
1	1	Data reliable without restriction	Well-documented study from published literature Performed according to valid/accepted TG (e.g. OECD) and preferably according to GLP
2	2	Data reliable with restriction	Well-documented study/data partially compliant with TG and may not have been GLP compliant
3	–	Expert review	Read-across Expert review of in silico result(s)[b] or Klimisch 3 or 4
4	–	Multiple concurring prediction results	
5	–	Single acceptable in silico result	
5	3	Data not reliable	Inferences between test system and substance Test system not relevant to exposure Method not acceptable for the end-point Not sufficiently documented for an expert review
5	4	Data not assignable	Lack of experimental details Referenced from short abstract or secondary literature

ECHA: European Chemicals Agency; GLP: Good Laboratory Practice; OECD: Organisation for Economic Co-operation and Development; TG: test guideline

[a] For an explanation of the Klimisch scores, see “Reliability” in section 4.5.4.1(b).
[b] In silico results in this case are broadly intended to capture expert systems, whereas read-across makes reference to expert-driven read-across – e.g. per the ECHA Read-across Assessment Framework.

Source: Modified from Myatt et al. (2018)

When using in silico models for mutagenicity assessment, it is recommended that two complementary models (e.g. a statistics-based model and an expert rule–based system) be applied, as recommended in ICH guideline M7(R1) (ICH, 2017) and EFSA (2016a). As stated by Barber et al. (2017), "the impact of a second system will be dependent upon not only its performance but also on its orthogonality to the first system, particularly in terms of training data, descriptors used and learning methods", in order to allow a WOE evaluation of two independent approaches (see also Greene et al., 2015). Practical application of QSAR models to predict mutagenicity is discussed in Sutter et al. (2013), Barber et al. (2015), Greene et al. (2015), Amberg et al. (2016), Mombelli, Raitano & Benfenati (2016) and Wichard (2017). In particular, the study by Greene et al. (2015) investigated how to best combine existing statistical and rule-based systems to enhance the detection of DNA-reactive mutagenic chemicals.

4.5.5.2 Threshold of toxicological concern (TTC)

Whereas an understanding of the potential for a chemical in the diet to pose a mutagenic hazard is an important element of the overall safety assessment of the chemical in food, it is also recognized that food can contain many contaminants and other constituents at very low levels. These can enter through natural sources (e.g. naturally present in plants or animals or taken up through the environment), through food processing or via migration from storage or packaging materials; they can also be formed during food processing and cooking. Analytical chemists are now able to routinely detect chemicals at sub–parts per billion levels, and, as analytical tools continue to improve, the detection limits will continue to be lowered. At some point, one could consider exposure to a constituent to be so low that it does not pose a safety concern, and testing is not needed. This is the principle behind the TTC concept.

The TTC is a screening tool that can be used to decide whether experimental mutagenicity testing is required for compounds present in the diet at very low levels. However, the TTC approach should not be used to replace data requirements for products, such as pesticides, subject to authorization by regulatory agencies. The TTC is defined as "a pragmatic risk assessment tool that is based on the principle of establishing a human exposure threshold value for all chemicals, below which there is a very low probability of an appreciable risk to human health" (Kroes et al., 2004). The origins of the TTC stem from the USFDA's threshold of regulation (USFDA, 1995), which was developed as a tool to facilitate the safety evaluation of food packaging materials, components of which have the potential to migrate into food at very low levels.

The TTC is used widely to assess low-level exposures to substances with insufficient toxicity data; it was reviewed most recently by EFSA & WHO (2016). It has been expanded from a single value (the USFDA's threshold of regulation) to encompass a range of exposure limits based on potency bins for chemicals. Substances posing a real or potential hazard from DNA-reactive mutagenicity are assigned to the bin with the most stringent exposure limit of 0.0025 µg/kg body weight per day (0.15 µg/day for a 60 kg adult). This exposure limit, first published by Kroes et al. (2004), was based on the distribution of cancer potencies for over 730 carcinogens and has been widely accepted in regulatory opinions on the TTC. Work is ongoing to further substantiate the TTC exposure limit for compounds considered to pose a possible hazard from DNA-reactive mutagenicity (Boobis et al., 2017; Cefic-LRI, 2020). This review is updating the existing database of carcinogens that was evaluated when this exposure limit was first established and will update methods using the state-of-the-science for the safety assessment of (mutagenic) carcinogens. It is also recognized that there are opportunities to refine the 0.0025 µg/kg body weight per day exposure limit for the TTC DNA-reactive mutagenicity tier, which currently assumes daily lifetime exposure, when it is generally recognized that higher exposures can be supported for shorter durations (Felter et al., 2009; Dewhurst & Renwick, 2013). This assumption has been accepted in guidance for mutagenic (DNA-reactive gene mutagens) impurities in pharmaceuticals (ICH, 2017), but is handled on a case-by-case basis in other sectors. It is also recognized that evaluations by the USFDA (Cheeseman, Machuga & Bailey, 1999) have shown that, on average, Ames-positive carcinogens are more potent than Ames-negative carcinogens (see sections 4.5.6.3 and Chapter 9, section 9.1.1, for further details of the TTC approach).

Chemicals are assigned to the "genotox tier" based on existing data (e.g. from mutagenicity assays) and evaluation of chemical structure. The latter is done based on the presence of structural alerts for DNA reactivity, which have been encoded in a number of software programs (e.g. Toxtree, OECD QSAR Toolbox, Derek Nexus; see section 4.5.5.1). Although this approach is generally considered to be robust, it is also recognized that different software programs can result in binning chemicals differently, such that EFSA & WHO (2016) concluded that "a transparent, consistent and reliable source for identifying structural alerts needs to be produced." In the absence of a single globally accepted tool to identify structural alerts, it is generally recognized that the existing tools are adequate to identify the alerts of greatest concern and that discordant results from different software programs do not necessarily raise a concern. As an example, an alert triggered by Toxtree based solely on the presence of a structural alert may be "overridden" by Derek Nexus, which evaluates the entire structure and may recognize that another part of the molecule renders that alert inactive. For example, Solvent Yellow 93 (CAS No. 4702-90-3), an azomethine dye, triggers an alert for genotoxic (DNA-reactive) carcinogenicity based on the presence of an α,β-unsaturated carbonyl. Derek Nexus also triggers this alert, but not if an aryl group is attached to the α,β-bond, as is the case for this chemical. Information available on this substance in a REACH dossier[4] confirms that "The test item did not induce mutagenicity in bacteria and in mammalian cell culture. It did furthermore not induce micronuclei in human lymphocytes." In addition, many scientists have emphasized the role of expert review when using in silico tools (e.g. Barber et al., 2015; Powley, 2015; Amberg et al., 2016). A WOE approach should be taken when binning chemicals into the genotoxicity tier for the TTC. This could be based on a combination of available data, structural similarity to other chemicals with data, evaluation of structural alerts from one or more software programs and expert judgement. Although there remains more work to do on the TTC approach, this is true for all safety assessment approaches. The TTC remains an important tool for evaluating low-level exposures to chemicals in food and can be used as an initial screen to determine whether mutagenicity testing or evaluation is needed. This would be the case when a plausible estimate of exposure to a substance with a clear structural alert for DNA-reactive gene mutagenicity exceeds the respective TTC.

To date, JMPR has applied the TTC approach to single metabolites of pesticides. The issue of how to deal with multiple metabolites that are considered potential DNA-reactive mutagens is

[4] https://echa.europa.eu/registration-dossier/-/registered-dossier/19812/7/7/1.

under active discussion by the OECD Residue Chemistry Expert Group's Drafting Group on Definition of Residues at the time of writing (mid-2020). Once agreed, the recommendations of that group should be adopted in this guidance. The TTC approach is used by JECFA as part of its procedure for assessing the safety of flavouring agents (see section 4.5.6.2).

4.5.5.3 Grouping and read-across approaches

For substances lacking empirical data, grouping approaches can be used to find similar substances for which data exist, which can then be used to infer properties of the data-poor substances ("read-across"). The WOE for evaluating mutagenic potential may come from read-across, structural alerts or QSAR models, using expert judgement on all available information, including empirical data, if limited data exist.

Groups of substances with similar human health or environmental toxicological properties, typically based on an aspect of chemical similarity, are known as chemical categories. When a category comprises two substances (an untested target substance of interest and a source analogue with data from which to read across), the approach is referred to as an analogue approach. Hanway & Evans (2000) were among the first to report read-across as part of the regulatory process for new substances in the United Kingdom. Concerted efforts have since sought to clarify terminology and formalize the linkages between read-across and (Q)SAR approaches, such as in the EU REACH guidance (ECHA, 2008, 2017a), which was developed in collaboration with the OECD to ensure broad consensus of the way in which read-across frameworks were outlined. Read-across, one of the main data gap–filling techniques, can be qualitative or quantitative. Other data gap–filling techniques include trend analysis and (Q)SARs (see also ECHA, 2008; ECETOC, 2012; OECD, 2014b).

The two main approaches to grouping similar chemicals together are "top down" and "bottom up". In a top-down approach, a large inventory of substances is subcategorized into smaller pragmatic groups. In some decision contexts, these "assessment groups" might take on specific context, such as to allow for the consideration of cumulative effects. Examples of a top-down approach are the grouping of food flavouring agents based on chemical structure by JECFA (see section 9.1.2.1) and the grouping of pesticides based either on phenomenological effects by EFSA (2013) or on common

MOAs by the USEPA (Leonard et al., 2019). Top-down groupings might also be used to prioritize large numbers of substances based on specific risk assessment concerns, such as persistence, bioaccumulation and toxicity or carcinogenicity, mutagenicity and reproductive toxicity. In contrast, the bottom-up approach tends to encompass scenarios in which a single target substance is being assessed based on source analogues identified as relevant to infer hazard properties lacking empirical data. In either the top-down or bottom-up approach, the grouping performed is intended to enable the inference of properties between group members (i.e. "reading across" these properties).

In the context of the EU REACH regulation, 63% of the substances submitted for registration used read-across as part of the hazard characterization (ECHA, 2020). In the USA, application of read-across varies widely between and within regulatory agencies and decision contexts (Patlewicz et al., 2019). For example, applications within the USEPA vary from the use of established chemical categories to identify potential concerns and testing expectations as part of the New Chemicals Program to the use of expert-driven read-across to inform screening-level provisional peer review toxicity value derivation in quantitative risk assessments for chemicals of interest to the USEPA Superfund programme (Wang et al., 2012).

Critical aspects in a read-across determination are the identification and evaluation of analogues (i.e. the definition of similarity), which depend on their chemistry and biological activity. In the mutagenicity field, these aspects are facilitated by the understanding of the MOAs and the associated test systems that characterize them. As such, the existence of structural alerts for mutagenicity, clastogenicity and DNA reactivity (see section 4.5.5.1) informs initial chemical categories.

There is a wide range of publicly accessible read-across tools (see Table 4.5 for examples and Patlewicz et al., 2017, for a detailed review), databases with genotoxicity or mutagenicity data (see, for example, Worth et al., 2010; Benigni, Bossa & Battistelli, 2013; Amberg et al., 2016; Corvi & Madia, 2018; Hasselgren et al., 2019; Table 4.2) and other data resources (Pawar et al., 2019) that can help establish sufficient similarity and compile a data matrix for the source and target substances.

Defining adequate similarity or dissimilarity requires a rational hypothesis with empirical evidence and depends on the end-point of concern, decision context and similarity metric chosen. Similarity should be based not only on structural and physicochemical properties, which tend to have been overemphasized (see Mellor et al., 2019, for recommendations on optimal use of molecular fingerprint–derived similarity measures), but also on toxicological (i.e. toxicodynamics and toxicokinetics) similarity (Schultz et al., 2015) supported by biological data (Zhu, Bouhifd & Donley, 2016). It is crucial to reflect on the boundaries of a category and whether specific structural dissimilarities have an impact on category membership.

Existing read-across frameworks rely on expert judgement to assess similarity in structure, reactivity, metabolism and physicochemical properties (Wu et al., 2010; Wang et al., 2012; Patlewicz et al., 2018) and can include a quantitative similarity score between analogues (Lester et al., 2018) or physicochemical similarity thresholds to assess performance (Helman, Shah & Patlewicz, 2018). Reporting templates for read-across assessments also help to identify uncertainties that concern the similarity argumentation and read-across rationale, and also whether the underlying data are of sufficient quality (see, for example, Blackburn & Stuard, 2014; Patlewicz et al., 2015; Schultz et al., 2015; Schultz, Richarz & Cronin, 2019). The ECHA Read-Across Assessment Framework (ECHA, 2017b), which also has been implemented in the OECD QSAR Toolbox (Kuseva et al., 2019), formulates a series of assessment criteria to establish confidence in the prediction and what information might be needed to reduce the uncertainties. New approach methodologies such as high-throughput or high-content screening data and linkages to adverse outcome pathways (AOPs) may help reduce uncertainty in read-across evaluations (see Wetmore, 2015; Zhu et al., 2016; OECD, 2017b,c, 2018a, 2019; Nelms et al., 2018). More recently, efforts to systematize read-across have sought to quantify the performance and uncertainty of the predictions akin to a QSAR-like approach (Shah et al., 2016; Zhu et al., 2016; Helman, Shah & Patlewicz, 2018; Patlewicz et al., 2018).

Read-across and (Q)SAR approaches are underpinned by the same principles and continuum of relating property or activity to a chemical structure, but boundaries between the two approaches are being challenged. (Q)SAR approaches are a more formal means of characterizing the relationship, whereas read-across approaches tend to be more case by case, based on expert review and judgement.

4.5.6 Considerations for specific compounds

4.5.6.1 Mixtures

Extracts from raw natural sources (e.g. plants, animals, algae, fungi, lichens) may be added to food for various purposes – for example, as supplements, flavouring agents or colouring agents. Such extracts are generally complex chemical mixtures, often including many uncharacterized components, rather than simple mixtures that comprise relatively fewer constituents, all with known identities.

Natural extracts from food-grade material generally do not raise safety concerns, based on a history of safe use, unless their use significantly increases exposure to any ingredient above average dietary exposure. In some cases, however, the safety of natural extracts added to food should be evaluated based on experimental or in silico data. Mutagenicity testing, in particular, is complicated by the dilution of individual components, which may hinder their identification using conventional test guidelines.

It is recommended that the selection (i.e. extraction) of test materials for mutagenicity testing follow the suggestions given by the European Medicines Agency's Committee on Herbal Medicinal Products (EMA, 2009). Extracts should be prepared with extremes of extraction solvents in order to maximize the spectrum of materials extracted, assuming that the mutagenicity of any extract produced with intermediate extraction solvents would be represented by the test results of the extremes tested.

Mutagenicity testing of mixtures may apply the tiered approach recommended by EFSA (2019a). The mixture should be chemically characterized as far as possible, providing critical quantitative compositional data, including stability and batch-to-batch variability, to ensure that the test material is representative of the mixture added to food. Useful guidelines exist for the chemical characterization of botanicals (e.g. EFSA, 2009), novel foods (e.g. EFSA, 2016b) and herbal medicinal products (e.g. EMA, 2011; USFDA, 2016) and for assessing the combined exposure to multiple chemicals (e.g. Meek et al., 2011; OECD, 2018b; EFSA, 2019b). Analytical methods to identify and control mutagenic impurities and degradation products of pharmaceuticals (e.g. Görög, 2018; Teasdale & Elder, 2018), although not directly applicable to food, could also be consulted.

For a well-characterized mixture (i.e. a simple mixture in which all components above a certain level[5] are identified and quantified), the mutagenic hazard of the mixture can be evaluated with a component-based approach that assesses all components individually, or at least representative substances for structurally related groups, using existing mutagenicity data and, if limited, supplemental (Q)SAR models. Where appropriate, a quantitative approach can be used for risk characterization, assuming dose addition (Ohta, 2006; EFSA, 2019a).

If the mixture contains a significant fraction of unidentified substances (i.e. complex mixture) or substances lacking empirical data, the chemically identified substances are first assessed individually for potential mutagenicity. If none of the identified substances is mutagenic or likely to be mutagenic, the mutagenic potential of the unidentified fraction should be evaluated. If possible, the unidentified fraction should be isolated for testing (e.g. Guo et al., 2014). Further fractionation of the unidentified material could be considered on a case-by-case basis to remove inert, toxicologically irrelevant components (e.g. high-molecular-weight polymers) in order to minimize the dilution of the components of interest or to remove highly toxic components (e.g. surface-active substances), which may prevent the testing of adequately high doses of the mixture owing to (cyto)toxicity. Testing of the whole mixture can be considered when isolation of the unidentified fraction is not feasible.

The testing strategy for mixtures or their fractions is similar to that for chemically defined constituents. However, as mentioned in OECD TGs 473, 476, 487 and 490, the top concentration may need to be higher than recommended for individual chemicals, in the absence of sufficient cytotoxicity, to increase the concentration of each component. The limit concentration recommended by the OECD for mixtures is 5 mg/mL, compared with 2 mg/mL for single substances (see, for example, OECD TG 473).

[5] Determining an appropriate level for this purpose relies on expert judgement, on a case-by-case basis, as it will depend on several factors, such as the source, process of production and formation of the mixture.

If testing of the whole mixture or fractions thereof in an adequately performed range of in vitro mutagenicity assays provides clearly negative results, the mixture could be considered to lack mutagenicity, and no further testing (e.g. by in vivo assays) would be needed. If testing of the whole mixture or fractions thereof in an adequately performed range of in vitro assays provides one or more positive results, in vivo follow-up testing should be considered on a case-by-case basis, based on the activity profile or MOA observed in vitro, following the same criteria applied to chemically defined substances.

Regulatory guidelines for the assessment of the potential mutagenicity of botanical or herbal medicinal products (EMA, 2006; USFDA, 2016) may also be useful when evaluating complex mixtures used in food.

4.5.6.2 *Flavouring agents*

The Codex Alimentarius Commission guidelines define a flavour as being the sum of those characteristics of any material taken in the mouth, perceived principally by the senses of taste and smell, and also the general pain and tactile receptors in the mouth, as received and interpreted by the brain. The perception of flavour is a property of flavourings (traditionally referred to as flavouring agents by JECFA). Flavourings represent a variety of liquid extracts, essences, natural substances and synthetic substances that are added to natural food products to impart taste and aroma or enhance taste and aroma when they are lost during food processing. Flavourings do not include substances that have an exclusively sweet, sour or salty taste (e.g. sugar, vinegar and table salt) (Codex Alimentarius Commission, 2008).

Depending on the origin and means of production, flavourings identified as a single constituent include those obtained by chemical synthesis or isolated through chemical processes as well as natural substances. Alternatively, flavourings derived from materials of vegetable, animal or microbiological origin by appropriate physical, enzymatic or microbiological processes are usually complex chemical mixtures that contain many different agents, including volatile substances. Constituents that occur naturally in flavourings, owing to their presence in the source materials (e.g. intrinsic fruit water) as well as foods or food ingredients used during the manufacturing process (e.g. ethanol, edible oil, acetic acid), can be considered to be part of the flavouring.

A category of complex flavourings is smoke flavourings and thermal process flavourings. Smoke flavourings include primary smoke condensates and primary tar fractions, flavourings produced by further processing of primary products, the purified water-based part of condensed smoke and the purified fraction of the water-insoluble high-density tar phase of condensed smoke. Thermal process flavourings are obtained by heating a blend of a nitrogen source (e.g. amino acids and their salts, peptides, proteins from foods) and a reducing sugar (e.g. dextrose/glucose, xylose). Owing to the intrinsic chemical complexity of flavourings (e.g. essential oils) that may consist of a number of organic chemical components, such as alcohols, aldehydes, ethers, esters, hydrocarbons, ketones, lactones, phenols and phenol ethers, mutagenicity testing, if needed, should be tailored accordingly. Benzo(*a*)pyrene, a DNA-reactive genotoxic carcinogen, is one of several polycyclic aromatic hydrocarbons (PAHs) that may occur in liquid smoke flavourings and is an indicator of PAH levels in liquid smoke flavourings. Current JECFA specifications limit the total PAH concentration to no more than 2 µg/kg, the lowest practical limit of measurement (FAO, 2001). After reviewing toxicological and carcinogenicity studies on smoke condensates and liquid smoke preparations, JECFA (FAO/WHO, 1987) concluded that such a complex group of products might not be amenable to the allocation of an ADI and that smoke flavourings of suitable specifications could be used provisionally to flavour foods traditionally treated by smoking; however, as the safety data on smoke flavourings were limited, novel uses of smoke flavourings should be approached with caution (FAO/WHO, 1987).

Currently, the JECFA Procedure for the Safety Evaluation of Flavouring Agents considers whether the WOE from empirical mutagenicity data or structural alerts suggests that the flavouring is potentially a DNA-reactive carcinogen (although this should more properly be DNA-reactive in vivo mutagen). If the answer is affirmative, then the Procedure for the Safety Evaluation of Flavouring Agents (described in Chapter 9, section 9.1.2.1, and updated in FAO/WHO, 2016) cannot be applied.

Flavourings that are complex mixtures should be tested according to the procedure recommended for extracts from natural sources (see section 4.5.6.1).

4.5.6.3 *Metabolites in crops/food-producing animals, degradation products and impurities*

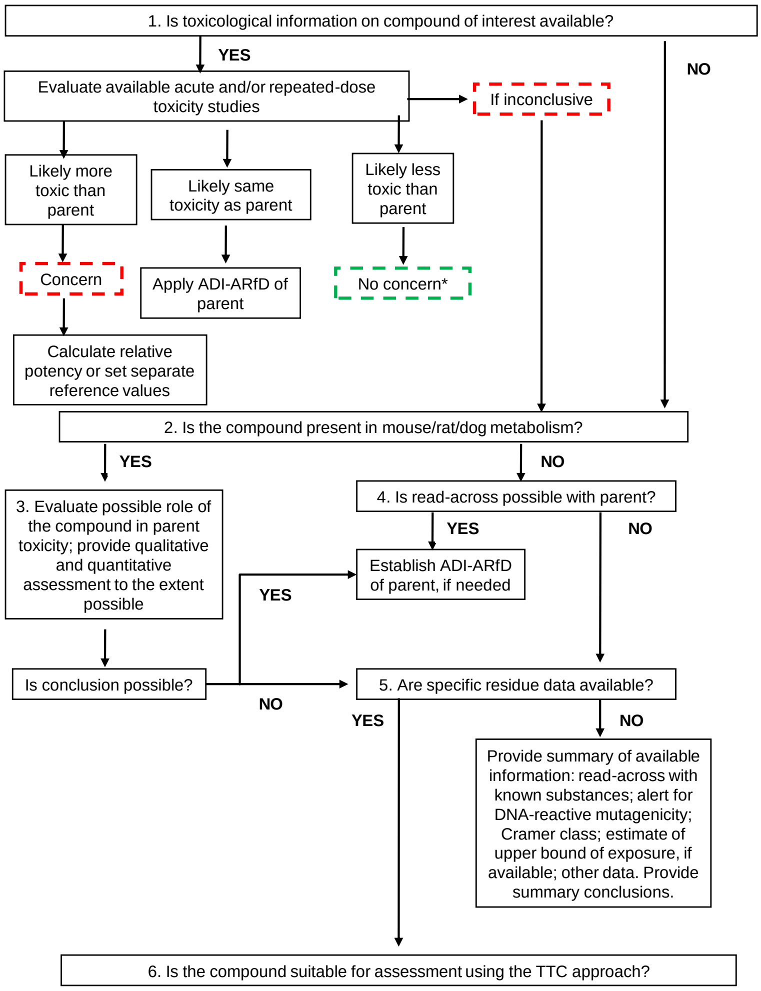
Substances considered here include metabolites of pesticide or veterinary drug active ingredients found as residues in food of plant and animal origin, impurities of the active ingredients, degradation products of pesticides or veterinary drugs due to non-enzymatic processes during food preparation or degradation products found in food commodities following application of pesticides or veterinary drugs.

A stepwise approach to evaluate the mutagenicity of these often minor components is suggested and begins with a non-testing phase. In fact, in many instances, experimental data are limited, but preliminary consideration of available data and information in conjunction with estimated exposure might suffice to reach a conclusion on safety with regard to mutagenicity. Whereas the scheme was first developed by JMPR for metabolites and degradation products of pesticides, the same principles should be applicable to impurities and contaminants in, or derived from, other substances.

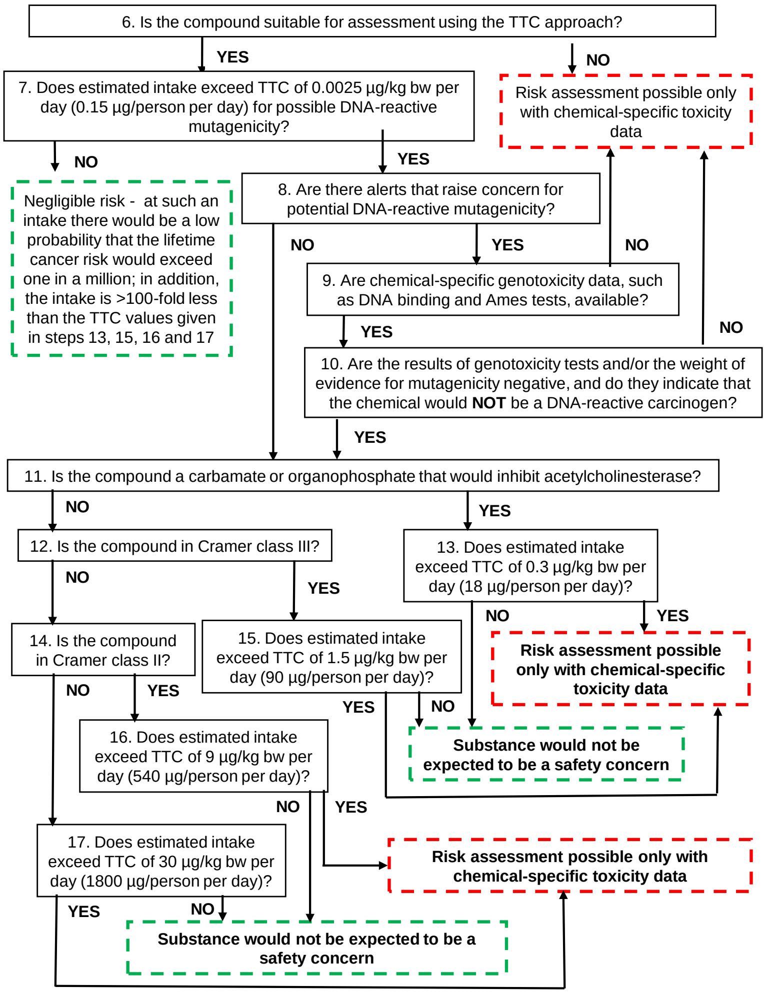
The evaluation of (DNA-reactive) mutagenic potential is part of the general toxicological evaluation of such impurities or degradation products, as illustrated in Fig. 4.2. Sections of the assessment scheme pertaining to mutagenicity are described below, assuming that, for the compound under evaluation, there are no empirical mutagenicity data available:

- *Step 1*: Is toxicological information on the compound of interest available? If so, evaluate the available toxicological information to determine potency relative to that of the parent.
- *Step 2*: If substance-specific data are available on the compound, determine appropriate HBGVs for use in risk assessment. If not, evaluate whether the compound of interest is formed in mice, rats or dogs, and hence whether the compound has been tested for DNA-reactive mutagenicity in tests with the parent compound. As a general rule, the compound is considered to have been tested in studies of the parent compound if urinary levels of the compound of interest represent at least 10% of the absorbed dose. Conjugates and downstream metabolites that derive only from the compound of interest are also included in the total.

Fig. 4.2. Assessment scheme for the safety of plant and animal metabolites/degradation products



* Note: For compounds already included in residue definition.



ADI: acceptable daily intake; ARfD: acute reference dose; bw: body weight; TTC: threshold of toxicological concern

Source: Adapted from WHO (2015a)

- *Step 3*: Evaluate the possible role of the metabolite in the DNA-reactive mutagenicity, if any, of the parent compound. If conclusions cannot be drawn, proceed to step 5.

- *Step 4*: For compounds that are unique plant or livestock metabolites or degradation products, the read-across approach is applied to use the mutagenicity information of compounds, including the parent compound, considered to have sufficient structural similarities to the compound of interest to permit read-across (see section 4.5.5.3 for details). If read-across is not deemed possible, owing to, for example, the lack of sufficiently similar tested analogues, proceed to step 5.

- *Step 5*: This step starts with consideration of whether specific residue data are available, such that dietary exposure can be estimated.[6] If estimation of dietary exposure is possible, proceed to step 6. If not, list all available relevant information, such as:

  - read-across from related substance(s),
  - structural alerts for DNA-reactive mutagenicity,
  - Cramer class,
  - estimate of upper bound of dietary exposure, if available, and
  - other relevant information,

  then determine whether the metabolite is of potential DNA-reactive mutagenicity concern, if possible, and provide advice for further assessment.

- *Step 6*: Determine whether the compound is suitable for assessment using the TTC approach. Substances currently not suitable (see section 4.5.5.2) are non-essential metals or metal-containing compounds, aflatoxin-like, azoxy-, benzidine- or *N*-nitroso- compounds, polyhalogenated dibenzodioxins, dibenzofurans or biphenyls, other chemicals that are known or predicted to bioaccumulate, proteins, steroids, insoluble nanomaterials, radioactive chemicals or mixtures of chemicals containing unknown chemical structures.

- *Step 7*: If the compound does not exceed the TTC for DNA-reactive mutagenic compounds (0.0025 µg/kg body weight per

[6] Dietary exposure assessment is detailed in Chapter 6.

day), the evaluation can be terminated with low concern for carcinogenicity from dietary exposure. Otherwise, proceed to step 8. See section 4.5.5.2 for more details on application of the TTC.

- *Step 8*: A number of models, including structural alert models (see section 4.5.5.1), are available that are suitable for this step. If there are no alerts for DNA-reactive mutagenicity, it can be concluded that there is low concern for this end-point. Similarly, if the only alert is also present in the parent compound, there is no evidence for a differential influence (compared with the parent compound) of the rest of the molecule on its mutagenic potential and the parent compound was negative in an adequate range of mutagenicity tests, it can be concluded that there is low concern for DNA-reactive mutagenicity. Otherwise, proceed to steps 9/10.

- *Steps 9/10*: Adequate in vitro or in vivo mutagenicity data are required to assure that DNA-reactive mutagenicity, carcinogenicity or developmental toxicity is unlikely despite the presence of structural alerts, based on a WOE evaluation (see sections 4.5.4.2 and 4.5.4.5).

Note that, based on structural considerations, if there are several compounds for which read-across would be possible, testing might be limited to one or a few representative compounds.

4.5.6.4 *Secondary metabolites in enzyme preparations*

Many commercial food enzymes are synthesized by microorganisms, which have been improved through classical enhancement techniques, such as mutagenesis and selection, or recombinant DNA technology. The process of manufacturing these food enzymes usually involves large-scale fermentations that necessitate large numbers of microorganisms. The enzymes synthesized de novo by these microorganisms either accumulate inside the cells or are secreted into the culture media of the fermentation tanks. In subsequent steps, the disrupted cells (or the culture media including the enzymes) are subjected to a range of purification processes using chemical, mechanical and thermal techniques (i.e. concentration, precipitation, extraction, centrifugation, filtration, chromatography, etc.).

The issue that is of interest from a safety assessment perspective is the presence of microorganism-derived secondary metabolites in the enzyme-purified extract. This material or extract, which also includes the food enzyme of interest, has traditionally been used in mutagenicity tests. Food enzymes (i.e. proteins) are heteropolymers of amino acids with high molecular weight (>1000 daltons), and they have poor cell membrane penetration potential. Furthermore, most proteins, excluding some allergens, are rapidly hydrolysed to their constituent amino acids in the gastrointestinal tract, so they are unlikely to come into direct contact with the DNA in a cell. Important information about microorganism-synthesized enzymes usually involves a consideration of their susceptibility to degradation in the gastrointestinal tract and the likelihood of them showing immunological cross-reactivity with known allergenic proteins.

The JECFA General Specifications and Considerations for Enzyme Preparations Used in Food Processing (FAO, 2006) are based on Pariza & Foster (1983) and guidelines of Europe's Scientific Committee for Food (SCF, 1991). A decision-tree approach is used for determining the safety of microbial enzyme preparations derived from non-pathogenic and non-toxigenic microorganisms and enzyme preparations derived from recombinant DNA microorganisms (Pariza & Foster, 1983; Pariza & Johnson, 2001) (see also Chapter 9, section 9.1.4.2).

To evaluate the safety of an enzyme preparation, a key initial consideration is an assessment of the production strain, in particular its capacity to synthesize potentially mutagenic secondary metabolites. Microbial secondary metabolites are low-molecular-weight entities that are not essential for the growth of producing cultures. JECFA (FAO, 2006), based on SCF (1991), recommended that the following tests be performed:

- a test for gene mutation in bacteria; and
- a test for chromosomal aberrations (preferably in vitro).

These tests should, where possible, be performed on a batch from the final purified fermentation product (i.e. before the addition of carriers and diluents). It was emphasized that these tests were intended to reveal mutagenic effects of unknown compounds synthesized during the fermentation process. It is recommended that the choice of test to assess these end-points should follow the guidance provided in this section of EHC 240. Hence, the preferred test for chromosomal aberrations would be an in vitro mammalian cell MN assay (OECD TG 487), which will also detect aneugenicity.

However, if the microorganism used in the production has a long history of safety in food use and belongs to a species about which it has been documented that no toxins are produced, and if the actual strain used has a well-documented origin, then it is possible to use the enzyme preparation from such an organism without any mutagenicity testing.

In such situations, a confirmed identification of the microorganism is very important. One example is *S. cerevisiae* (SCF, 1991). An invertase preparation derived from *S. cerevisiae* fermentation did not require toxicity testing (FAO/WHO, 2002) based on a JECFA (FAO/WHO, 1972) conclusion that enzymes from microorganisms traditionally accepted as natural food constituents or normally used in food preparation should themselves be regarded as foods. By 2018, JECFA had evaluated over 80 food enzyme preparations from microorganisms such as *Trichoderma reesei*, *Bacillus subtilis*, *B. amyloliquefaciens*, *B. licheniformis*, *Aspergillus niger* and *A. oryzae*, but had never recorded a positive result in any mutagenicity assay (FAO/WHO, 2019). These data suggest that there are several strains of microorganisms that could constitute safe strain lineages for food enzyme production and would therefore not require mutagenicity testing.

Alternatives to mutagenicity testing for secondary metabolites in fermentation extracts could be chemical characterization of the extracts supported by detailed knowledge of the genomic sequence of any genetically modified microorganisms to exclude the possibility of secondary metabolite toxin genes.

4.5.7 Recent developments and future directions

The need to evaluate the potential mutagenicity posed by thousands of chemicals in commerce remains an urgent priority. There is also a need for the quantitative assessment of the risk associated with realistic environmental exposures. The former necessitates the development and validation of novel, high-throughput tools for mutagenicity/genotoxicity assessment, including in vitro tools that are aligned with the demand to replace and reduce animal use for toxicity assessment (Richmond, 2002; Pfuhler et al., 2014; Burden et al., 2015; Beken, Kasper & Van der Laan, 2016; Riebeling, Luch & Tralau, 2018). The latter will require the establishment of a computational framework for dose–response analysis that includes point of departure determinations for the interpretation of mutagenicity test data in the context of risk assessment (White & Johnson, 2016).

Recently developed high-throughput tools exploit advances in informatics and instrumentation technologies to rapidly assess traditional mutagenicity end-points (e.g. mutations and chromosome damage) and molecular end-points indicative of DNA damage or a DNA damage response. Additionally, (Q)SAR-based models developed by commercial (e.g. Leadscope, MultiCase, Lhasa Ltd) or public sector (e.g. OECD) organizations are increasingly being used for predicting bacterial mutagenicity and chromosomal damage (see Table 4.5 and section 4.5.5.1). High-throughput and in silico methods can rapidly screen and prioritize potential mutagens, but their direct utility for establishing HBGVs (e.g. ADI, ARfD, MOE) is currently limited.

4.5.7.1 Novel in vivo genotoxicity approaches

High-throughput technologies such as flow cytometry and automated microscopy permit the rapid detection and quantification of induced gene mutations and chromosomal aberrations in vivo (see section 4.5.2.3). As many of these assays evaluate mutagenicity biomarkers in peripheral blood, they can be readily integrated into ongoing repeated-dose toxicity studies, thus reducing the need for independent mutagenicity tests (Dertinger et al., 2002; Witt et al., 2007, 2008). Additionally, some methods are amenable to evaluating mutagenicity biomarkers in humans (Witt et al., 2007; Fenech et al., 2013; Collins et al., 2014; Dertinger et al., 2015; Olsen et al., 2017).

In addition to the high-throughput approaches highlighted previously (see section 4.5.2.3), novel in vivo approaches (Table 4.7) can measure MN frequency in liver and, with modification, in small intestine and colon (Uno et al., 2015a,b). Additional novel approaches can measure homologous recombination in virtually any tissue of interest (e.g. FYDR, RaDR mouse; Hendricks et al., 2003; Sukup-Jackson et al., 2014). No international guidelines yet exist for these approaches, but data from these approaches could be used in support of TG data.

4.5.7.2 Novel in vitro genotoxicity approaches

The last few years have seen the development of a range of novel, high-throughput in vitro tools for assessing genotoxicity. Despite

Table 4.7. Novel approaches for genotoxicity assessment

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
In vivo assays				
Liver MN assay	MN frequency in hepatic tissue	Traditional end-point; metabolically competent tissue; can be adapted to other tissues (e.g. colon, intestine)	Technically challenging; not high throughput	Uno et al. (2015a,b)
Recombo-Mouse	Integrated, direct repeat reporter to score homologous recombination events	Flow cytometry or automated imaging to score fluorescent signal; can examine almost any tissue	Rarity of recombinant cells in quiescent tissues; not high throughput	Hendricks et al. (2003); Sukup-Jackson et al. (2014)
Adductomics	Rapid assessment of type and frequency of DNA adducts	Combined with stable isotopes; can differentiate between endogenous and exogenous DNA lesions; can be applied in vivo or in vitro	Indicator test detecting pre-mutagenic lesions; interpretation of results can be complicated, particularly if endogenous and exogenous adducts are not distinguished; no standardized protocols	Rappaport et al. (2012); Balbo, Turesky & Villalta (2014); Hemeryck, Moore & Vanhaecke (2016); Lai et al. (2016); Yao & Feng (2016); Chang et al. (2018); Yu et al. (2018); Takeshita et al. (2019)

Table 4.7 (continued)

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
In vitro assays that assess the frequency of mutations or DNA damage				
Pig-a mutagenicity assay	Flow cytometric detection of *Pig-a* mutant phenotype	Analogous to in vivo assay; automated detection of cells with mutant phenotype; flow cytometry scoring	No consensus on protocol	Krüger, Hofmann & Hartwig (2015); Krüger et al. (2016); Bemis & Heflich (2019)
Transgenic rodent reporter mutagenicity assays	Positive selection assay to detect mutations at a variety of transgenic loci (e.g. *lacI*, *lacZ*, *cII*, *gpt*, Spi^{-})	Scoring protocol identical to in vivo version (i.e. OECD TG 488); scores actual mutations; numerous cell systems available; detects a variety of mutation types; does not require laborious clonal selection; some versions partially validated	Laborious compared with high-throughput reporter-based assays; transgenes, not endogenous loci; no consensus regarding assay protocol; not high throughput	White et al. (2019)
Hupki Mouse	Immortalization of primary embryonic fibroblasts	Measures mutation in human p53; in vitro scoring	Continuous culture maintenance for an extended period (8–12 weeks); not high throughput	Luo et al. (2001); Besaratinia & Pfeifer (2010); Kucab, Phillips & Arlt (2010)

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
Cisbio γH2AX assay	Quantification of H2AX phosphorylation	Positive responses highly predictive of genotoxicity (clastogenicity); homogeneous format with no wash steps required; high-throughput screening compatible; suitable for use with adherent or suspension cells	Requires an HTRF compatible reader and a −60 °C freezer	Hsieh et al. (2019); PerkinElmer-Cisbio (2020)
Microplate comet assay	Automated analyses of DNA “tails”	Increased reproducibility; higher throughput	Same issues of specificity as with conventional comet assay	Ge et al. (2015); Sykora et al. (2018)
In vitro reporter assays (indirect measures of genotoxicity)				
ToxTracker assay	Expression of specific reporter genes upregulated by DNA damage	Simultaneously monitors genes involved in DNA damage response, microtubule disruption, oxidative stress and protein damage response; flow cytometry scoring	Restricted to specifically constructed cell lines	Hendriks et al. (2012, 2016); Ates et al. (2016)

Table 4.7 (continued)

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
MultiFlow DNA Damage assay	In vitro high-content assays for multiple end-points	Determines MOA for MN induction; flow cytometry scoring	Method developed for suspension cell lines only	Bryce et al. (2013); Bemis et al. (2016a); Smith-Roe et al. (2018)
MultiFlow Aneugen Molecular Initiating Event Kit	In vitro follow-up assay for determining MOA of aneugens identified in the MultiFlow assay	Identifies tubulin binders and inhibitors of Aurora B kinase; flow cytometry scoring	Not yet commercially available	Bernacki et al. (2019)
p53-RE assay	Reporter gene assay to assess activation of p53 response element	Assay for cellular signalling pathways activated by DNA damage; automated scoring	Currently limited to a single cell line (HCT-116); can respond to non-genotoxic stressors	Witt et al. (2017)
DT40 differential cytotoxicity assay	Enhanced cytotoxicity in cell lines lacking specific DNA repair enzymes	Highly specific for DNA repair pathways; automated scoring	Limited to isogenic chicken cell lines	Yamamoto et al. (2011); Nishihara et al. (2016)

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
GreenScreen, BlueScreen	GADD45a-based reporter system; green fluorescent protein (GreenScreen) or Gaussia Luciferase (BlueScreen) detection	Highly specific for DNA repair pathways; automated scoring	Currently limited to a single cell line; may respond to non-genotoxic stressors	Hastwell et al. (2006); Simpson et al. (2013)
High-throughput real-time RT-qPCR	Gene expression assessment of 95 genes involved in genomic stability	Can be used for cell lines, primary cells, three-dimensional cultures	Limited to a few cell types, each requiring response characterization	Fischer et al. (2016); Strauch et al. (2017)
TGx-DDI	Gene expression assessment of 64 DNA damage/repair genes	Prediction of DNA-damaging potential	Limited to a few cell types, each requiring response characterization	Li et al. (2015, 2017); Williams et al. (2015); Yauk et al. (2016a); Corton, Williams & Yauk (2018)

DDI: DNA damage–inducing; DNA: deoxyribonucleic acid; HTRF: Homogeneous Time-Resolved Fluorescence; MN: micronucleus; MOA: mode of action; RT-qPCR: reverse transcription quantitative polymerase chain reaction noteworthy advantages related to the cost, throughput and information content of these assays, incorporation of realistic and effective xenobiotic metabolism is a concern. Nevertheless, high-throughput assays are now available to rapidly assess the induction of DNA damage and repair, gene mutations, chromosomal damage or DNA strand breaks (Table 4.7). As mutagenicity screening for regulatory purposes generally requires the assessment of gene mutations and chromosomal damage, assays that streamline detection of these end-points are particularly noteworthy. In vitro versions of the flow cytometric *Pig-a* gene mutation assay and the Transgenic Rodent Somatic and Germ Cell Mutation Assays (OECD TG 488) permit enumeration of mutations at a variety of endogenous and transgenic loci (e.g. *Pig-a*, *lacI*, *lacZ*, *cII*, *gpt*, Spi^{-}). These assays do not require clonal selection and can measure mutagenicity more efficiently than, for example, traditional *Tk* and *Hprt* locus mutation assays.

Some of the high-throughput in vitro assays summarized in Table 4.7 exploit cellular pathways to rapidly measure biomarkers of DNA damage or repair; most are based on genetically engineered cell lines containing a promoter activated by genotoxic insult (e.g. p53 response element) fused to one or more reporter genes (e.g. β-lactamase). Reporter gene activation is visualized via, for example, automated micro-confocal imaging, fluorescent or luminescent readouts, or flow cytometry. Examples include the ToxTracker (Hendriks et al., 2012, 2016; Ates et al., 2016), GreenScreen (Hastwell et al., 2006; Simpson et al., 2013) and several reporter gene and antibody assays (e.g. p53RE, γH2AX, ATAD5) used by the United States Tox21 Program (https://ntp.niehs.nih.gov/whatwestudy/tox21/toolbox/index.html) or the USEPA's ToxCast Program (https://comptox.epa.gov/dashboard/chemical_lists/toxcast). Importantly, in addition to mutagenic hazard, the simultaneous or sequential examination of multiple end-points representing several distinct pathways permits delineation of the mutagenic MOA. Related assays, such as the MultiFlow DNA Damage assay, assess the presence and localization of proteins (e.g. γH2AX, nuclear p53, phospho-histone H3) indicative of DNA damage and alterations in chromosome structure or number (Bryce et al., 2016, 2017, 2018). Proteins are targeted by fluorescently labelled antibodies, and cellular phenotype is scored using flow cytometry. In addition to reporter-based approaches that track and quantify DNA damage response activation, gene expression–based strategies, such as DNA microarray, quantitative polymerase chain reaction (qPCR) and RNA sequencing approaches, have been used as high-throughput approaches for measuring DNA damage signalling. For example, the TGx-DDI assay monitors genes involved in genomic stability (e.g. generalized stress responses, DNA repair, cell cycle control, apoptosis and mitotic signalling) to identify DNA damage–inducing (DDI) substances (Li et al., 2015, 2017; Williams et al., 2015; Yauk et al., 2016a; Corton, Williams & Yauk, 2018; Corton, Witt & Yauk, 2019). Similarly, a high-throughput real-time reverse transcription quantitative polymerase chain reaction (RT-qPCR) assay rapidly scores 95 genes active in maintaining genomic integrity (Fischer et al., 2016; Strauch et al., 2017). These reporter systems rapidly track DNA damage and repair as indirect measures of genotoxicity.

To date, none of the high-throughput tools listed in Table 4.7 have OECD TGs, nor have they been incorporated into widely accepted genotoxicity assessment platforms, such as those recommended by ICH (2011), USFDA (2007) and ECHA (2017a). A future role for these tools in regulatory decision-making would be consistent with global trends to modernize the current mutagenicity assessment frameworks, to reduce and replace the use of experimental animals and to generate mutagenicity MOA information. For example, Dearfield et al. (2017) outlined a paradigm shift whereby a variety of mechanistic end-points indicative of genomic damage are incorporated into a "next-generation testing strategy". Indeed, high-throughput tools are already supporting regulatory evaluations based on traditional in vitro assays. For example, the European Commission's Scientific Committee on Consumer Safety considers additional in vitro tests that include gene expression and recombinant cell reporter assays (SCCS, 2018). Similarly, Corton, Williams & Yauk (2018) outlined how the TGx-DDI assay can be used for regulatory screening of chemicals. Buick et al. (2017) used a TGx-DDI biomarker to evaluate two data-poor substances prioritized by Health Canada for regulatory decision-making due to structural similarity to known mutagens (i.e. Disperse Orange and 1,2,4-benzenetriol), resulting in compound classification consistent with more traditional end-points (e.g. in vitro MN formation). Private sector organizations are now routinely using high-throughput in vitro assays to evaluate the mutagenicity of products in development, such as therapeutic candidates and industrial chemicals (Thougaard et al., 2014; International Antimony Association, 2018; Motoyama et al., 2018; Dertinger et al., 2019; Pinter et al., 2020).

The in vitro tools and approaches summarized in Table 4.7 employ standard cultures of mammalian cells (e.g. two-dimensional attached cultures, suspension cultures). To acquire data that might be deemed more relevant to humans, while also reducing the use of animals in research, three-dimensional cell culture systems have been developed to score end-points such as chromosomal (i.e. MN) and DNA damage (i.e. comet assay). Several novel assays are summarized in Table 4.8.

Another alternative to traditional in vivo testing involves the use of chicken eggs to assess chromosomal damage based on the frequency of MN in extraembryonic peripheral blood (Wolf & Luepke, 1997; Wolf, Niehaus-Rolf & Luepke, 2003; Hothorn et al., 2013).

Advances in high-throughput detection of DNA damage and repair, chromosomal aberrations and gene mutations may soon be eclipsed by error-corrected, next-generation DNA sequencing (NGS) approaches. Whereas previous NGS technologies did not permit detection of rare, exposure-induced mutations (i.e. $<10^{-5}$) in the absence of clonal expansion, recent computational and experimental innovations now allow detection of such rare mutations ($<10^{-8}$) (Salk, Schmitt & Loeb, 2018), with the precision and accuracy required to assess genetic alterations in only a few DNA molecules within a cell population. Although error-corrected NGS technologies are not yet fully validated or widely applied, the technology is rapidly advancing and may soon be routinely available, particularly because it does not require specialized cells, loci or reporters, can score mutations at virtually any locus in any tissue, organism or cells in culture, and can readily be integrated into repeated-dose or translational studies linking observations to humans.

4.5.7.3 *Adverse outcome pathways for mutagenicity*

The OECD AOP framework organizes diverse toxicological data from different levels of biological complexity in order to increase confidence in mechanistic relationships between key events leading to adverse health outcomes. The AOP Knowledge Base,[7] which includes several modules, supports AOP construction to improve application of mechanistic information for both chemical testing and assessment (OECD, 2017d). AOPs also feed into Integrated

[7] https://aopkb.oecd.org/index.html.

Table 4.8. Novel in vitro genotoxicity assessment systems based on multicellular, three-dimensional constructs

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
Three-dimensional MN test	MN frequency in reconstructed skin model	Traditional end-point; simple to score; application in reconstructed skin models	Questions remain concerning metabolism	Aardema et al. (2010); Kirsch-Volders et al. (2011); Chapman et al. (2014); Pfuhler et al. (2014)
Three-dimensional comet assay	DNA damage assay in reconstructed skin model	Traditional end-point; simple to score; application in reconstructed skin models	Questions remain concerning metabolism	Pfuhler et al. (2014); Reisinger et al. (2018)
Hen's egg MN assay	MN frequency in extraembryonic peripheral blood of fertilized hen eggs	Traditional end-point; some metabolic capacity	Non-mammalian test; limited metabolism	Wolf & Luepke (1997); Wolf, Niehaus-Rolf & Luepke (2003); Hothorn et al. (2013)
Avian egg genotoxicity assay	Comet assay and ^{32}P-postlabelling of adducts in hepatocytes isolated from turkey or hen eggs treated ex vivo	Some metabolic activity; traditional end-points; studies of some MOAs	Non-mammalian test; limited metabolism; postlabelling with ^{32}P	Williams, Deschl & Williams (2011); Kobets et al. (2016, 2018, 2019)

DNA: deoxyribonucleic acid; MN: micronucleus; MOA: mode of action

Approaches to Testing and Assessment (IATA), a pragmatic approach to hazard characterization that integrates in silico, in vitro and in vivo assessment tools, including high-throughput in vitro tools based on toxicogenomic or recombinant cell reporter technologies (Sakuratani, Horie & Leinala, 2018). The OECD IATA Case Studies Project reviews case-studies related to different end-points, including mutagenicity and genotoxicity,[1] and publishes the learnings and areas identified where additional guidance is needed (OECD, 2017a,b, 2018a, 2019). The AOP on "alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations" was the first AOP on mutagenicity published in the OECD AOP series (Yauk et al., 2016b). To date, several other AOPs related to mutagenicity are under development in the AOP-Wiki[2] (one module of the AOP Knowledge Base), and several ongoing initiatives should contribute to populating the AOP Knowledge Base with more AOPs on mutagenicity in the near future, increasing the development of AOP networks and supporting further tiered testing and IATA strategies.

4.5.7.4 *Quantitative approaches for safety assessment*

National and international mutagenicity evaluation committees have highlighted a desire to employ quantitative methods for regulatory interpretation of mutagenicity dose–response data (MacGregor et al., 2015a,b; UKCOM, 2018). Lacking carcinogenicity data, quantitative analysis of in vivo mutagenicity dose–response data could be used for deriving MOEs (White & Johnson, 2016). This is particularly relevant for risk assessment and management of unavoidable food contaminants with positive results for gene mutation or DNA-reactive mutagenicity structural alerts and exposures exceeding the TTC of 0.0025 µg/kg body weight per day (see section 4.5.5.2). Moving to a quantitative approach requires a paradigm shift from hazard identification of mutagens and recognizes that compensatory cellular responses (i.e. DNA damage processing) are quantitatively manifested as mechanistically plausible dose–response thresholds (Parry, Fielder & McDonald, 1994; Nohmi, 2008, 2018; Carmichael, Kirsch-Volders & Vrijhof, 2009; Johnson et al., 2014; Nohmi & Tsuzuki, 2016). With respect to threshold determination, this is still under debate, and there is currently no international consensus.

---

[1] http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm.

[2] https://aopwiki.org/.

Several researchers have employed dose–response point of departure values, such as the benchmark dose (BMD), the threshold dose (Td) and the NOGEL, for quantitative interpretation of in vitro and in vivo mutagenicity dose–response data. With respect to in vitro dose–response data, the BMD approach has been used for MOE determinations and to rank potency across test substances, cell types and experimental protocols (Bemis et al., 2016b; Benford, 2016; Tweats et al., 2016; Wills et al., 2016; Verma et al., 2017; Guo et al., 2018). However, it should be noted that not all in vitro guideline mutagenicity tests are suitable for dose–response assessment, as they are optimized to discriminate between "positive" and "negative" compounds. The mutagenicity of ethyl methanesulfonate, an impurity detected in Viracept, an antiretroviral drug, was shown to exhibit a threshold, both in vitro and in vivo. In vivo mutagenicity data were then used to determine a permissible daily exposure to the compound (Gocke & Wall, 2009; Müller & Gocke, 2009). Although the regulatory utility of quantitative interpretation of in vivo dose–response data is increasingly recognized, use of mutagenicity-based BMD values to estimate MOEs for mutagenic food contaminants will require consensus regarding, for example, choice of test/end-point, an appropriate benchmark response for mutagenicity end-points, and appropriate safety factors for exposure limit determination (Ritter et al., 2007; Nielsen, Ostergaard & Larsen, 2008; Dankovic et al., 2015; IPCS, 2018).

4.5.8 References[1]

Aardema MJ, Barnett BC, Khambatta Z, Reisinger K, Ouedraogo-Arras G, Faquet B et al. (2010). International prevalidation studies of the EpiDerm™ 3D human reconstructed skin micronucleus (RSMN) assay: transferability and reproducibility. Mutat Res. 701(2):123–31. doi:10.1016/j.mrgentox.2010.05.017.

Amberg A, Beilke L, Bercu J, Bower D, Brigo A, Cross KP et al. (2016). Principles and procedures for implementation of ICH M7 recommended (Q)SAR analyses. Regul Toxicol Pharmacol. 77:13–24. doi:10.1016/j.yrtph.2016.02.004.

Amberg A, Andaya RV, Anger LT, Barber C, Beilke L, Bercu J et al. (2019). Principles and procedures for handling out-of-domain and indeterminate

[1] Internet links provided in these references were active as of the date of final editing.

results as part of ICH M7 recommended (Q)SAR analyses. Regul Toxicol Pharmacol. 102:53–64. doi:10.1016/j.yrtph.2018.12.007.

Ames BN, Lee FD, Durston WE (1973). An improved bacterial test system for the detection and classification of mutagens and carcinogens. Proc Natl Acad Sci U S A. 70(3):782–6. doi:10.1073/pnas.70.3.782.

Ates G, Favyts D, Hendriks G, Derr R, Mertens B, Verschaeve L et al. (2016). The Vitotox and ToxTracker assays: a two-test combination for quick and reliable assessment of genotoxic hazards. Mutat Res. 810:13–21. doi:10.1016/ j.mrgentox.2016.09.005.

Au WW, Ramanujam VMS, Ward JB, Legator MS (1991). Chromosome aberrations in lymphocytes of mice after sub-acute low-level inhalation exposure to benzene. Mutat Res. 260(2):219–24. doi:10.1016/0165-1218(91)90011-a.

Auerbach C, Robson JM, Carr JG (1947). The chemical production of mutations. Science. 105(2723):243–7. doi:10.1126/science.105.2723.243.

Avlasevich S, Bryce S, de Boeck M, Elhajouji A, Van Goethem F, Lynch A et al. (2011). Flow cytometric analysis of micronuclei in mammalian cell cultures: past, present and future. Mutagenesis. 26(1):147–52. doi:10.1093/mutage/ geq058.

Bakhtyari NG, Raitano G, Benfenati E, Martin T, Young D (2013). Comparison of in silico models for prediction of mutagenicity. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 31(1):45–66. doi:10.1080/10590501.2013.763576.

Balbo S, Turesky RJ, Villalta PW (2014). DNA adductomics. Chem Res Toxicol. 27(3):356–66. doi:10.1021/tx4004352.

Barber C, Amberg A, Custer L, Dobo KL, Glowienke S, Van Gompel J et al. (2015). Establishing best practise in the application of expert review of mutagenicity under ICH M7. Regul Toxicol Pharmacol. 73(1):367–77. doi:10.1016/j.yrtph.2015.07.018.

Barber C, Hanser T, Judson P, Williams R (2017). Distinguishing between expert and statistical systems for application under ICH M7. Regul Toxicol Pharmacol. 84:124–30. doi:10.1016/j.yrtph.2016.12.012.

Beken S, Kasper P, Van der Laan JW (2016). Regulatory acceptance of alternative methods in the development and approval of pharmaceuticals. Adv Exp Med Biol. 856:33–6. doi:10.1007/978-3-319-33826-2_3.

Bemis JC, Heflich RH (2019). In vitro mammalian cell mutation assays based on the *Pig-a* gene: a report of the 7th International Workshop on Genotoxicity Testing (IWGT) workgroup. Mutat Res. 847:403028. doi:10.1016/ j.mrgentox.2019.03.001.

Bemis JC, Bryce SM, Nern M, Raschke M, Sutter A (2016a). Best practices for application of attachment cells to in vitro micronucleus assessment by flow cytometry. Mutat Res. 795:51–9. doi:10.1016/j.mrgentox.2015.10.007.

Bemis JC, Wills JW, Bryce SM, Torous DK, Dertinger SD, Slob W (2016b). Comparison of in vitro and in vivo clastogenic potency based on benchmark dose analysis of flow cytometric micronucleus data. Mutagenesis. 31(3):277–85. doi:10.1093/mutage/gev041.

Benford DJ (2016). The use of dose–response data in a margin of exposure approach to carcinogenic risk assessment for genotoxic chemicals in food. Mutagenesis. 31(3):329–31. doi:10.1093/mutage/gev064.

Benigni R, Bossa C, Battistelli CL (2013). ISSTOX Chemical Toxicity Databases [online databases]. Rome: Istituto Superiore di Sanità (http://old.iss.it/meca/index.php?lang=1&anno=2013&tipo=25).

Benigni R, Battistelli CL, Bossa C, Giuliani A, Fioravanzo E, Bassan A et al. (2019). Evaluation of the applicability of existing (Q)SAR models for predicting the genotoxicity of pesticides and similarity analysis related with genotoxicity of pesticides for facilitating of grouping and read across. Parma: European Food Safety Authority (EFSA supporting publications EN-1598; https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/sp.efsa.2019.EN-1598).

Bernacki DT, Bryce SM, Bemis JC, Dertinger SD (2019). Aneugen molecular mechanism assay: proof-of-concept with 27 reference chemicals. Toxicol Sci. 170(2):382–93. doi:10.1093/toxsci/kfz123.

Beronius A, Ågerstrand M (2017). Making the most of expert judgment in hazard and risk assessment of chemicals. Toxicol Res. 6(5):571–7. doi:10.1039/c7tx00114b.

Besaratinia A, Pfeifer GP (2010). Applications of the human p53 knock-in (*Hupki*) mouse model for human carcinogen testing. FASEB J. 24(8):2612–9. doi:10.1096/fj.10-157263.

Beyersmann D, Hartwig A (2008). Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms. Arch Toxicol. 82(8):493–512. doi:10.1007/s00204-008-0313-y.

Blackburn K, Stuard SB (2014). A framework to facilitate consistent characterization of read across uncertainty. Regul Toxicol Pharmacol. 68(3):353–62. doi:10.1016/j.yrtph.2014.01.004.

Boobis A, Brown P, Cronin MTD, Edwards J, Lodovico Galli C, Goodman J et al. (2017). Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their re-evaluation. Crit Rev Toxicol. 47(8):710–32. doi:10.1080/10408444.2017.1318822.

Bossa C, Benigni R, Tcheremenskaia O, Battistelli CL (2018). (Q)SAR methods for predicting genotoxicity and carcinogenicity: scientific rationale and regulatory frameworks. In: Nicolotti O, editor. Computational toxicology. Methods in molecular biology, vol. 1800. New York (NY): Humana Press; 447–73. doi:10.1007/978-1-4939-7899-1_20.

Brown K (2012). Methods for the detection of DNA adducts. In: Genetic toxicology. New York (NY): Springer; 207–30.

Bryce SM, Avlasevich SL, Bemis JC, Phonethepswath S, Dertinger SD (2010). Miniaturized flow cytometric in vitro micronucleus assay represents an efficient tool for comprehensively characterizing genotoxicity dose–response relationships. Mutat Res. 703:191–9. doi:10.1016/j.mrgentox.2010.08.020.

Bryce SM, Avlasevich SL, Bemis JC, Dertinger SD (2011). Miniaturized flow cytometry–based CHO-K1 micronucleus assay discriminates aneugenic and clastogenic modes of action. Environ Mol Mutagen. 52(4):280–6. doi:10.1002/em.20618.

Bryce SM, Avlasevich SL, Bemis JC, Tate M, Walmsley RM, Saad F et al. (2013). Flow cytometric 96-well microplate–based in vitro micronucleus assay with human TK6 cells: protocol optimization and transferability assessment. Environ Mol Mutagen. 54(3):180–94. doi:10.1002/em.21760.

Bryce SM, Bernacki DT, Bemis JC, Dertinger SD (2016). Genotoxic mode of action predictions from a multiplexed flow cytometric assay and a machine learning approach. Environ Mol Mutagen. 57(3):171–89. doi:10.1002/em.21996.

Bryce SM, Bernacki DT, Bemis JC, Spellman RA, Engel ME, Schuler M et al. (2017). Interlaboratory evaluation of a multiplexed high information content in vitro genotoxicity assay. Environ Mol Mutagen. 58(3):146–61. doi:10.1002/em.22083.

Bryce SM, Bernacki DT, Smith-Roe SL, Witt KL, Bemis JC, Dertinger SD (2018). Investigating the generalizability of the MultiFlow® DNA Damage assay and several companion machine learning models with a set of 103 diverse test chemicals. Toxicol Sci. 162(1):146–66. doi:10.1093/toxsci/kfx235.

Buick JK, Williams A, Kuo B, Wills JW, Swartz CD, Recio L et al. (2017). Integration of the TGx-28.65 genomic biomarker with the flow cytometry micronucleus test to assess the genotoxicity of disperse orange and 1,2,4-benzenetriol in human TK6 cells. Mutat Res. 806:51–62. doi:10.1016/j.mrfmmm.2017.09.002.

Burden N, Mahony C, Müller BP, Terry C, Westmoreland C, Kimber I (2015). Aligning the 3Rs with new paradigms in the safety assessment of chemicals. Toxicology. 330:62–6. doi:10.1016/j.tox.2015.01.014.

Busby WF Jr, Ackermann JM, Crespi CL (1999). Effect of methanol, ethanol, dimethyl sulfoxide, and acetonitrile on in vitro activities of cDNA-expressed human cytochromes P-450. Drug Metab Dispos. 27(2):246–9.

Carmichael P, Kirsch-Volders M, Vrijhof H (2009). Thresholds for genotoxins and their application in risk assessment. Mutat Res. 678(2):71. doi:10.1016/j.mrgentox.2009.04.012.

Cassano A, Raitano G, Mombelli E, Fernández A, Cester J, Roncaglioni A et al. (2014). Evaluation of QSAR models for the prediction of Ames genotoxicity:

a retrospective exercise on the chemical substances registered under the EU REACH regulation. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 32(3):273–98. doi:10.1080/10590501.2014.938955.

Cefic-LRI (2020). B18: Carcinogen dose–response database for threshold of toxicological concern (CDRD-TTC). Brussels: European Chemical Industry Council Long-Range Research Initiative (http://cefic-lri.org/projects/b18-carcinogen-dose-response-database-for-threshold-of-toxicological-concern-cdrd-ttc/).

Chang YJ, Cooke MS, Hu CW, Chao MR (2018). Novel approach to integrated DNA adductomics for the assessment of in vitro and in vivo environmental exposures. Arch Toxicol. 92(8):2665–80. doi:10.1007/s00204-018-2252-6.

Chapman KE, Thomas AD, Wills JW, Pfuhler S, Doak SH, Jenkins GJS (2014). Automation and validation of micronucleus detection in the 3D EpiDerm™ human reconstructed skin assay and correlation with 2D dose responses. Mutagenesis. 29(3):165–75. doi:10.1093/mutage/geu011.

Cheeseman MA, Machuga EJ, Bailey AB (1999). A tiered approach to threshold of regulation. Food Chem Toxicol. 37(4):387–412. doi:10.1016/S0278-6915(99)00024-1.

Chervona Y, Arita A, Costa M (2012). Carcinogenic metals and the epigenome: understanding the effect of nickel, arsenic, and chromium. Metallomics. 4(7):619–27. doi:10.1039/c2mt20033c.

Cimino MC (2006). Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. Environ Mol Mutagen. 47(5):362–90. doi:10.1002/em.20216.

Codex Alimentarius Commission (2008). Guidelines for the use of flavourings. Rome: Food and Agriculture Organization of the United Nations and World Health Organization, Joint FAO/WHO Food Standards Programme, Codex Alimentarius Commission (CAC/GL 66-2008; http://www.fao.org/input/download/standards/11020/cxg_066e.pdf).

Collins A, Koppen G, Valdiglesias V, Dusinska M, Kruszewski M, Møller P et al. (2014). The comet assay as a tool for human biomonitoring studies: the ComNet project. Mutat Res. 759:27–39. doi:10.1016/j.mrrev.2013.10.001.

Corton JC, Williams A, Yauk CL (2018). Using a gene expression biomarker to identify DNA damage–inducing agents in microarray profiles. Environ Mol Mutagen. 59(9):772–84. doi:10.1002/em.22243.

Corton JC, Witt KL, Yauk CL (2019). Identification of p53 activators in a human microarray compendium. Chem Res Toxicol. 32(9):1748–59. doi:10.1021/acs.chemrestox.9b00052.

Corvi R, Madia F (2018). EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database of Ames Positive Chemicals [online database].

Brussels: European Commission, Joint Research Centre (http://data.europa.eu/89h/jrc-eurl-ecvam-genotoxicity-carcinogenicity-ames).

Costa M (2019). Review of arsenic toxicity, speciation and polyadenylation of canonical histones. Toxicol Appl Pharmacol. 375:1–4. doi:10.1016/j.taap.2019.05.006.

Cronin MTD, Richarz A-N, Schultz TW (2019). Identification and description of the uncertainty, variability, bias and influence in quantitative structure–activity relationships (QSARs) for toxicity prediction. Regul Toxicol Pharmacol. 106:90–104. doi:10.1016/j.yrtph.2019.04.007.

Dankovic DA, Naumann BD, Maier A, Dourson ML, Levy LS (2015). The scientific basis of uncertainty factors used in setting occupational exposure limits. J Occup Environ Hyg. 12(Suppl 1):S55–68. doi:10.1080/15459624.2015.1060325.

Dearfield KL, Cimino MC, McCarroll NE, Mauer I, Valcovic LR (2002). Genotoxicity risk assessment: a proposed classification strategy. Mutat Res. 521(1–2):121–35. doi:10.1016/S1383-5718(02)00236-X.

Dearfield KL, Thybaud V, Cimino MC, Custer L, Czich A, Harvey JS et al. (2011). Follow-up actions from positive results of in vitro genetic toxicity testing. Environ Mol Mutagen. 52(3):177–204. doi:10.1002/em.20617.

Dearfield KL, Gollapudi BB, Bemis JC, Benz RD, Douglas GR, Elespuru RK et al. (2017). Next generation testing strategy for assessment of genomic damage: a conceptual framework and considerations. Environ Mol Mutagen. 58(5):264–83. doi:10.1002/em.22045.

Dertinger SD, Torous DK, Hall NE, Murante FG, Gleason SE, Miller RK et al. (2002). Enumeration of micronucleated CD71-positive human reticulocytes with a single-laser flow cytometer. Mutat Res. 515(1–2):3–14. doi:10.1016/s1383-5718(02)00009-8.

Dertinger SD, Bishop ME, McNamee JP, Hayashi M, Suzuki T, Asano N et al. (2006). Flow cytometric analysis of micronuclei in peripheral blood reticulocytes: I. Intra- and interlaboratory comparison with microscopic scoring. Toxicol Sci. 94(1):83–91. doi:10.1093/toxsci/kfl075.

Dertinger SD, Phonethepswath S, Weller P, Nicolette J, Murray J, Sonders P et al. (2011a). International *Pig-a* gene mutation assay trial: evaluation of transferability across 14 laboratories. Environ Mol Mutagen. 52(9):690–8. doi:10.1002/em.20672.

Dertinger SD, Torous DK, Hayashi M, MacGregor JT (2011b). Flow cytometric scoring of micronucleated erythrocytes: an efficient platform for assessing in vivo cytogenetic damage. Mutagenesis. 26(1):139–45. doi:10.1093/mutage/geq055.

Dertinger SD, Avlasevich SL, Bemis JC, Chen Y, MacGregor JT (2015). Human erythrocyte *PIG-A* assay: an easily monitored index of gene mutation requiring low volume blood samples. Environ Mol Mutagen. 56(4):366–77. doi:10.1002/em.21924.

Dertinger SD, Kraynak AR, Wheeldon RP, Bernacki D, Bryce SM, Hall N et al. (2019). Predictions of genotoxic potential, mode of action, molecular targets, and potency via a tiered Multiflow® assay data analysis strategy. Environ Mol Mutagen. 60(6):513–33. doi:10.1002/em.22274.

Dewangan J, Pandey PK, Divakar A, Mishra S, Srivastava S, Rath SK (2018). Detection of gene mutation in cultured mammalian cells. In: Kumar A, Dobrovolsky VN, Dhawan A, Shanker R, editors. Mutagenicity: assays and applications. London: Elsevier; 49–67.

Dewhurst I, Renwick AG (2013). Evaluation of the threshold of toxicological concern (TTC) – challenges and approaches. Regul Toxicol Pharmacol. 65(1):168–77. doi:10.1016/j.yrtph.2012.03.007.

D'Gama AM, Pochareddy S, Li M, Jamuar SS, Reiff RE, Lam AN et al. (2015). Targeted DNA sequencing from autism spectrum disorder brains implicates multiple genetic mechanisms. Neuron. 88:910–7. doi:10.1016/j.neuron.2015.11.009.

Dobo KL, Greene N, Fred C, Glowienke S, Harvey JS, Hasselgren C et al. (2012). In silico methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: an industry survey. Regul Toxicol Pharmacol. 62(3):449–55. doi:10.1016/j.yrtph.2012.01.007.

Douglas GR, Jiao J, Gingerich JD, Gossen JA, Soper LM (1995). Temporal and molecular characteristics of mutations induced by ethylnitrosourea in germ cells isolated from seminiferous tubules and in spermatozoa of lacZ transgenic mice. Proc Natl Acad Sci U S A. 92(16):7485–9. doi:10.1073/pnas.92.16.7485.

Eastmond DA (2017). Recommendations for the evaluation of complex genetic toxicity data sets when assessing carcinogenic risks to humans. Environ Mol Mutagen. 58:380–5. doi:10.1002/em.22078.

Eastmond DA, Hartwig A, Anderson D, Anwar WA, Cimino MC, Dobrev I et al. (2009). Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS harmonized scheme. Mutagenesis. 24(4):341–9. doi:10.1093/mutage/gep014.

ECETOC (2012). Category approaches, read-across, (Q)SAR. Brussels: European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC Technical Report No. 116; http://www.ecetoc.org/wp-content/uploads/2014/08/ECETOC-TR-116-Category-approaches-Read-across-QSAR.pdf).

ECHA (2008). Chapter R.6: QSARs and grouping of chemicals. In: Guidance on information requirements and chemical safety assessment. Helsinki: European Chemicals Agency (https://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf).

ECHA (2011). Chapter R.4: Evaluation of available information. Version 1.1. In: Guidance on information requirements and chemical safety assessment. Helsinki: European Chemicals Agency (https://www.echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf).

ECHA (2017a). Chapter R.7a: Endpoint specific guidance. Version 6.0. In: Guidance on information requirements and chemical safety assessment. Helsinki: European Chemicals Agency (https://www.echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf).

ECHA (2017b). Read-Across Assessment Framework (RAAF). Helsinki: European Chemicals Agency (https://echa.europa.eu/documents/10162/13628/raaf_en.pdf).

ECHA (2020). The use of alternatives to testing on animals for the REACH regulation. Fourth report under Article 117(3) of the REACH regulation. Helsinki: European Chemicals Agency (https://echa.europa.eu/documents/10162/0/alternatives_test_animals_2020_en.pdf/db66b8a3-00af-6856-ef96-5ccc5ae11026).

EFSA (2009). European Food Safety Authority Scientific Committee; Guidance on safety assessment of botanicals and botanical preparations intended for use as ingredients in food supplements, on request of EFSA. EFSA J. 7(9):1249. doi:10.2093/j.efsa.2009.1249.

EFSA (2011). European Food Safety Authority Scientific Committee; Scientific Opinion on genotoxicity testing strategies applicable to food and feed safety assessment. EFSA J. 9(9):2379. doi:10.2903/j.efsa.2011.2379.

EFSA (2013). European Food Safety Authority Panel on Plant Protection Products and their Residues (PPR); Scientific Opinion on the identification of pesticides to be included in cumulative assessment groups on the basis of their toxicological profile. EFSA J. 11(7):3293. doi:10.2903/j.efsa.2013.3293.

EFSA (2015). European Food Safety Authority Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids (CEF); Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs. EFSA J. 13(1):3978. doi:10.2903/j.efsa.2015.3978.

EFSA (2016a). European Food Safety Authority Panel on Plant Protection Products and their Residues (PPR); Guidance on the establishment of the residue definition for dietary risk assessment. EFSA J. 14(12):e04549. doi:10.2903/j.efsa.2016.4549.

EFSA (2016b). European Food Safety Authority Panel on Dietetic Products, Nutrition and Allergies (NDA); Guidance on the preparation and presentation of an application for authorisation of a novel food in the context of Regulation (EU) 2015/2283. EFSA J. 14(11):4594. doi:10.2903/j.efsa.2016.4594.

EFSA (2019a). European Food Safety Authority Scientific Committee; Statement on genotoxicity assessment of chemical mixtures. EFSA J. 17:5519. doi:10.2903/j.efsa.2019.5519.

EFSA (2019b). European Food Safety Authority Scientific Committee; Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA J. 17(3):5634. doi:10.2903/j.efsa.2019.5634.

EFSA, WHO (2016). Review of the threshold of toxicological concern (TTC) approach and development of new TTC decision tree. Parma: European Food Safety Authority and World Health Organization (EFSA supporting publication 2016: EN-1006; https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2016.EN-1006).

EMA (2006). Guideline on non-clinical documentation for herbal medicinal products in applications for marketing authorisation (bibliographical and mixed applications) and in applications for simplified registration. London: European Medicines Agency, Committee on Herbal Medicinal Products (HMPC) (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003576.pdf).

EMA (2009). Guideline on selection of test materials for genotoxicity testing for traditional herbal medicinal products/herbal medicinal products. London: European Medicines Agency, Committee on Herbal Medicinal Products (HMPC) (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500059583.pdf).

EMA (2011). Guideline on specifications: test procedures and acceptance criteria for herbal substances, herbal preparations and herbal medicinal products/traditional herbal medicinal products. London: European Medicines Agency (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/09/WC500113210.pdf).

Erickson RP (2003). Somatic gene mutation and human disease other than cancer. Mutat Res. 543(2):125–36. doi:10.1016/S1383-5742(03)00010-3.

Erickson RP (2010). Somatic gene mutation and human disease other than cancer: an update. Mutat Res. 705(2):96–106. doi:10.1016/j.mrrev.2010.04.002.

FAO (2001). Smoke flavourings. Prepared by the fifty-seventh meeting of JECFA for publication in FNP 52 Addendum 9. Rome: Food and Agriculture Organization of the United Nations (http://www.fao.org/fileadmin/user_upload/jecfa_additives/docs/Monograph1/Additive-386.pdf).

FAO (2006). General specifications and considerations for enzyme preparations used in food processing. Prepared by the sixty-seventh meeting of JECFA for publication in FAO JECFA Monographs 3. Rome: Food and Agriculture Organization of the United Nations (http://www.fao.org/docrep/009/a0691e/A0691E03.htm).

FAO/WHO (1972). Toxicological evaluation of some enzymes, modified starches and certain other substances. Prepared by the fifteenth meeting of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (FAO Nutrition Meetings Report Series, No. 50A; WHO Food Additives Series, No. 1; http://www.inchem.org/documents/jecfa/jecmono/v001je01.htm).

FAO/WHO (1987). Evaluation of certain food additives and contaminants. Thirty-first report of the Joint FAO/WHO Expert Committee on Food Additives.

Geneva: World Health Organization (WHO Technical Report Series, No. 759; https://apps.who.int/iris/bitstream/handle/10665/39108/WHO_TRS_759.pdf?sequence=1).

FAO/WHO (2002). Invertase from *Saccharomyces cerevisiae*. In: Fifty-seventh report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization; 18 (http://apps.who.int/iris/bitstream/handle/10665/42578/WHO_TRS_909.pdf?sequence=1).

FAO/WHO (2011). Pesticide residues in food 2011 – Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations and World Health Organization (FAO Plant Production and Protection Paper 211; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Report11/JMPR_2011_Report.pdf).

FAO/WHO (2016). Evaluation of certain food additives. Eighty-second report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 1000: https://apps.who.int/iris/bitstream/handle/10665/250277/9789241210003-eng.pdf?sequence=1).

FAO/WHO (2017a). Chlormequat. In: Pesticide residues in food – 2017. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper 232; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Report2017/5.6_CHLORMEQUAT__015_.pdf).

FAO/WHO (2017b). Natamycin. In: Pesticide residues in food – 2017. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper 232; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Report2017/5.25_NATAMYCIN__300_.pdf).

FAO/WHO (2018). Evaluation of certain food additives. Eighty-sixth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization and Food and Agriculture Organization of the United Nations (WHO Technical Report Series, No. 1014; https://apps.who.int/iris/bitstream/handle/10665/279832/9789241210232-eng.pdf?sequence=1&isAllowed=y).

FAO/WHO (2019). Joint FAO/WHO Expert Committee on Food Additives: Expert meeting of Working Group established to consider the evaluation of enzyme preparations used in the manufacture of foods, Rome, 11–14 December 2018. Final report. Rome: Food and Agriculture Organization of the United Nations and World Health Organization (http://www.fao.org/3/ca7188en/ca7188en.pdf).

Felter S, Lane RW, Latulippe ME, Llewellyn GC, Olin SS, Scimeca JA et al. (2009). Refining the threshold of toxicological concern (TTC) for risk prioritization of trace chemicals in food. Food Chem Toxicol. 47(9):2236–45. doi:10.1016/j.fct.2009.06.018.

Fenech M, Kirsch-Volders M, Rossnerova A, Sram R, Romm H, Bolognesi C et al. (2013). HUMN project initiative and review of validation, quality control and prospects for further development of automated micronucleus assays using image cytometry systems. Int J Hyg Environ Health. 216(5):541–52. doi:10.1016/j.ijheh.2013.01.008.

Fischer BM, Neumann D, Piberger AL, Risnes SF, Köberle B, Hartwig A (2016). Use of high-throughput RT-qPCR to assess modulations of gene expression profiles related to genomic stability and interactions by cadmium. Arch Toxicol. 90(11):2745–61. doi:10.1007/s00204-015-1621-7.

Ge J, Chow DN, Fessler JL, Weingeist DM, Wood DK, Engelward BP (2015). Micropatterned comet assay enables high throughput and sensitive DNA damage quantification. Mutagenesis. 30(1):11–9. doi:10.1093/mutage/geu063.

Gocke E, Wall M (2009). In vivo genotoxicity of EMS: statistical assessment of the dose response curves. Toxicol Lett. 190(3):298–302. doi:10.1016/j.toxlet.2009.03.008.

Gollapudi BB, Lynch AM, Heflich RH, Dertinger SD, Dobrovolsky VN, Froetschl R et al. (2015). The in vivo *Pig-a* assay: a report of the International Workshop on Genotoxicity Testing (IWGT) workgroup. Mutat Res. 783:23–35. doi:10.1016/j.mrgentox.2014.09.007.

Görög S (2018). Critical review of reports on impurity and degradation product profiling in the last decade. Trends Anal Chem. 101:2–16. doi:10.1016/j.trac.2017.09.012.

Greene N, Dobo KL, Kenyon MO, Cheung J, Munzner J, Sobol Z et al. (2015). A practical application of two in silico systems for identification of potentially mutagenic impurities. Regul Toxicol Pharmacol. 72(2):335–49. doi:10.1016/j.yrtph.2015.05.008.

Guo X, Zhang S, Dial SL, Boudreau MD, Xia Q, Fu PP et al. (2014). In vitro investigation of the mutagenic potential of *Aloe vera* extracts. Toxicol Res. 3(6):487–96. doi:10.1039/c4tx00053f.

Guo X, Heflich RH, Dial SL, De M, Richter PA, Mei N (2018). Quantitative differentiation of whole smoke solution–induced mutagenicity in the mouse lymphoma assay. Environ Mol Mutagen. 59(2):103–13. doi:10.1002/em.22151.

Hanser T, Barber C, Marchaland JFF, Werner S (2016). Applicability domain: towards a more formal definition. SAR QSAR Environ Res. 27:865–81. doi:10.1080/1062936X.2016.1250229.

Hanway RH, Evans PF (2000). Read-across of toxicological data in the notification of new chemicals. Toxicol Lett. 116(Suppl 1):61. doi:10.1016/S0378-4274(00)80001-6.

Hartwig A (2013). Metal interaction with redox regulation: an integrating concept in metal carcinogenesis? Free Radic Biol Med. 55:63–72. doi:10.1016/j.freeradbiomed.2012.11.009.

Hasselgren C, Ahlberg E, Akahori Y, Amberg A, Anger LT, Atienzar F et al. (2019). Genetic toxicology in silico protocol. Regul Toxicol Pharmacol. 107:104403. doi:10.1016/j.yrtph.2019.104403.

Hastwell PW, Chai LL, Roberts KJ, Webster TW, Harvey JS, Rees RW et al. (2006). High-specificity and high-sensitivity genotoxicity assessment in a human cell line: validation of the GreenScreen HC *GADD45a-GFP* genotoxicity assay. Mutat Res. 607(2):160–75. doi:10.1016/j.mrgentox.2006.04.011.

Helman G, Shah I, Patlewicz G (2018). Extending the Generalised Read-Across approach (GenRA): a systematic analysis of the impact of physicochemical property information on read-across performance. Comp Toxicol. 8:34–50. doi:10.1016/j.comtox.2018.07.001.

Hemeryck LY, Moore SA, Vanhaecke L (2016). Mass spectrometric mapping of the DNA adductome as a means to study genotoxin exposure, metabolism, and effect. Anal Chem. 88(15):7436–46. doi:10.1021/acs.analchem.6b00863.

Hendricks CA, Almeida KH, Stitt MS, Jonnalagadda VS, Rugo RE, Kerrison GF et al. (2003). Spontaneous mitotic homologous recombination at an enhanced yellow fluorescent protein (EYFP) cDNA direct repeat in transgenic mice. Proc Natl Acad Sci U S A. 100(11):6325–30. doi:10.1073/pnas.1232231100.

Hendriks G, Atallah M, Morolli B, Calléja F, Ras-Verloop N, Huijskens I et al. (2012). The ToxTracker assay: novel GFP reporter systems that provide mechanistic insight into the genotoxic properties of chemicals. Toxicol Sci. 125(1):285–98. doi:10.1093/toxsci/kfr281.

Hendriks G, Derr RS, Misovic B, Morolli B, Calléja FMGR, Vrieling H (2016). The extended ToxTracker assay discriminates between induction of DNA damage, oxidative stress, and protein misfolding. Toxicol Sci. 150(1):190–203. doi:10.1093/toxsci/kfv323.

HGMD (2017). Human Gene Mutation Database [online database]. Cardiff: Institute of Medical Genetics (http://www.hgmd.cf.ac.uk/ac/index.php).

Honma M, Kitazawa A, Cayley A, Williams RV, Barber C, Hanser T et al. (2019). Improvement of quantitative structure–activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project. Mutagenesis. 34(1):3–16. doi:10.1093/mutage/gey031.

Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW (2014). SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol. 14:43. doi:10.1186/1471-2288-14-43.

Hothorn LA, Reisinger K, Wolf T, Poth A, Fieblinger D, Liebsch M et al. (2013). Statistical analysis of the hen's egg test for micronucleus induction (HET-MN assay). Mutat Res. 757(1):68–78. doi:10.1016/j.mrgentox.2013.04.023.

Hsieh J-H, Smith-Roe SL, Huang R, Sedykh A, Shockley KR, Auerbach SS et al. (2019). Identifying compounds with genotoxicity potential using Tox21 high throughput screening assays. Chem Res Toxicol. 32(7):1384–1401. doi: 10.1021/acs.chemrestox.9b00053.

ICH (2011). Guidance on genotoxicity testing and data interpretation for pharmaceuticals intended for human use. S2(R1). London: European Medicines Agency, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (https://database.ich.org/sites/default/files/S2%28R1%29%20Guideline.pdf).

ICH (2014). ICH guideline M7 on assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. London: European Medicines Agency, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-assessment-control-dna-reactive-mutagenic-impurities-pharmaceuticals-limit-potential_en.pdf).

ICH (2017). ICH guideline M7(R1) on assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. London: European Medicines Agency, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-m7r1-assessment-control-dna-reactive-mutagenic-impurities-pharmaceuticals-limit_en.pdf).

International Antimony Association (2018). Scientific opinion, provisional read-across justification, and further research opportunities – Human health: genotoxicity. Brussels: International Antimony Association (https://ntp.niehs.nih.gov/ntp/roc/pubcomm/01braibant20180809.pdf).

IPCS (2018). Guidance document on evaluating and expressing uncertainty in hazard characterization, second edition. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document 11; https://www.who.int/publications-detail/guidance-document-on-evaluating-and-expressing-uncertainty-in-hazard-characterization-2nd-ed).

Johnson GE, Soeteman-Hernandez LG, Gollapudi BB, Bodger OG, Dearfield KL, Heflich RH et al. (2014). Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. Environ Mol Mutagen. 55(8):609–23. doi:10.1002/em.21870.

Jones NJ (2012). ^{32}P-postlabelling for the sensitive detection of DNA adducts. Methods Mol Biol. 817:183–206. doi:10.1007/978-1-61779-421-6_10.

Kanwal R, Gupta K, Gupta S (2015). Cancer epigenetics: an introduction. Methods Mol Biol. 1238:3–25. doi:10.1007/978-1-4939-1804-1_1.

Khanal S, Singh P, Avlasevich SL, Torous DK, Bemis JC, Dertinger SD (2018). Integration of liver and blood micronucleus and *Pig-a* gene mutation endpoints into rat 28-day repeat-treatment studies: proof-of-principle with diethylnitrosamine. Mutat Res. 828:30–5. doi:10.1016/j.mrgentox.2018.02.005.

Kimoto T, Horibata K, Miura D, Chikura S, Okada Y, Ukai A et al. (2016). The PIGRET assay, a method for measuring *Pig-a* gene mutation in reticulocytes, is reliable as a short-term in vivo genotoxicity test: summary of the MMS/JEMS-collaborative study across 16 laboratories using 24 chemicals. Mutat Res. 811:3–15. doi:10.1016/j.mrgentox.2016.10.003.

Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F et al. (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. doi:10.1016/j.mrgentox.2006.11.008.

Kirkland D, Uno Y, Luijten M, Beevers C, Van Benthem J, Burlinson B et al. (2019). In vivo genotoxicity testing strategies: report from the 7th International Workshop on Genotoxicity Testing (IWGT). Mutat Res. 847:403035. doi:10.1016/j.mrgentox.2019.03.008.

Kirsch-Volders M, Decordier I, Elhajouji A, Plas G, Aardema MJ, Fenech M (2011). In vitro genotoxicity testing using the micronucleus assay in cell lines, human lymphocytes and 3D human skin models. Mutagenesis. 26(1):177–84. doi:10.1093/mutage/geq068.

Kissling GE, Dertinger SD, Hayashi M, MacGregor JT (2007). Sensitivity of the erythrocyte micronucleus assay: dependence on number of cells scored and inter-animal variability. Mutat Res. 634(1–2):235–40. doi:10.1016/j.mrgentox.2007.07.010.

Kleinstreuer NC, Karmaus A, Mansouri K, Allen DG, Fitzpatrick JM, Patlewicz G (2018). Predictive models for acute oral systemic toxicity: a workshop to bridge the gap from research to regulation. Comput Toxicol. 8:21–4. doi:10.1016/j.comtox.2018.08.002.

Kligerman AD, Allen JW, Erexson GL, Morgan DL (1993). Cytogenetic studies of rodents exposed to styrene by inhalation. In: Sorsa M, Peltonen K, Vainio H, Hemminki K, editors. Butadiene and styrene: assessment of health hazards. Lyon: International Agency for Research on Cancer; 217–24 (IARC Scientific Publications, No. 127).

Klimisch HJ, Andreae M, Tillmann U (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol. 25(1):1–5. doi:10.1006/rtph.1996.1076.

Kobets T, Williams GM (2019). Review of the evidence for thresholds for DNA-reactive and epigenetic experimental chemical carcinogens. Chem Biol Interact. 301:88–111. doi:10.1016/j.cbi.2018.11.011.

Kobets T, Duan JD, Brunnemann KD, Etter S, Smith B, Williams GM (2016). Structure–activity relationships for DNA damage by alkenylbenzenes in turkey egg fetal liver. Toxicol Sci. 150(2):301–11. doi:10.1093/toxsci/kfv322.

Kobets T, Duan JD, Brunnemann KD, Iatropoulos MJ, Etter S, Hickey C et al. (2018). In ovo testing of flavor and fragrance materials in Turkey Egg Genotoxicity Assay (TEGA), comparison of results to in vitro and in vivo data. Food Chem Toxicol. 115:228–43. doi:10.1016/j.fct.2018.03.015.

Kobets T, Cartus AT, Fuhlbrueck JA, Brengel A, Stegmüller S, Duan JD et al. (2019). Assessment and characterization of DNA adducts produced by alkenylbenzenes in fetal turkey and chicken livers. Food Chem Toxicol. 129:424–33. doi:10.1016/j.fct.2019.05.010.

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A et al. (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem Toxicol. 42(1):65–83. doi:10.1016/j.fct.2003.08.006.

Krüger CT, Hofmann M, Hartwig A (2015). The in vitro *PIG-A* gene mutation assay: mutagenicity testing via flow cytometry based on the glycosylphosphatidylinositol (GPI) status of TK6 cells. Arch Toxicol. 89(12):2429–43. doi:10.1007/s00204-014-1413-5.

Krüger CT, Fischer BM, Armant O, Morath V, Strähle U, Hartwig A (2016). The in vitro *PIG-A* gene mutation assay: glycosylphosphatidylinositol (GPI)-related genotype-to-phenotype relationship in TK6 cells. Arch Toxicol. 90(7):1729–36. doi:10.1007/s00204-016-1707-x.

Kucab JE, Phillips DH, Arlt VM (2010). Linking environmental carcinogen exposure to *TP53* mutations in human tumours using the human *TP53* knock-in (Hupki) mouse model. FEBS J. 277(12):2567–83. doi:10.1111/j.1742-4658.2010.07676.x.

Kuseva C, Schultz TW, Yordanova D, Tankova K, Kutsarova S, Pavlov T et al. (2019). The implementation of RAAF in the OECD QSAR toolbox. Regul Toxicol Pharmacol. 105:51–61. doi:10.1016/j.yrtph.2019.03.018.

Lai Y, Yu R, Hartwell HJ, Moeller BC, Bodnar WM, Swenberg JA (2016). Measurement of endogenous versus exogenous formaldehyde-induced DNA–protein crosslinks in animal tissues by stable isotope labeling and ultrasensitive mass spectrometry. Cancer Res. 76(9):2652–61. doi:10.1158/0008-5472.CAN-15-2527.

Langie SA, Koppen G, Desaulniers D, Al-Mulla F, Al-Temaimi R, Amedei A et al. (2015). Causes of genome instability: the effect of low dose chemical exposures in modern society. Carcinogenesis. 36(Suppl 1):S61–S88. doi:10.1093/carcin/bgv031.

Leonard JA, Nelms M, Craig E, Perron M, Pope-Varsalona H, Dobreniecki S et al. (2019). A weight of evidence approach to investigate potential common mechanisms in pesticide groups to support cumulative risk assessment: a case study with dinitroaniline pesticides. Regul Toxicol Pharmacol. 107:104419. doi:10.1016/j.yrtph.2019.104419.

Lester C, Reis A, Laufersweiler M, Wu S, Blackburn K (2018). Structure activity relationship (SAR) toxicological assessments: the role of expert judgment. Regul Toxicol Pharmacol. 92:390–406. doi:10.1016/j.yrtph.2017.12.026.

Levy DD, Zeiger E, Escobar PA, Hakura A, Van der Leede BJM, Kato M et al. (2019). Recommended criteria for the evaluation of bacterial mutagenicity data (Ames test). Mutat Res. 848:403074. doi:10.1016/j.mrgentox.2019.07.004.

Li HH, Hyduke DR, Chen R, Heard P, Yauk CL, Aubrecht J et al. (2015). Development of a toxicogenomics signature for genotoxicity using a dose-optimization and informatics strategy in human cells. Environ Mol Mutagen. 56(6):505–19. doi:10.1002/em.21941.

Li HH, Chen R, Hyduke DR, Williams A, Frötschl R, Ellinger-Ziegelbauer H et al. (2017). Development and validation of a high-throughput transcriptomic biomarker to address 21st century genetic toxicology needs. Proc Natl Acad Sci U S A. 114(51):E10881–E10889. doi:10.1073/pnas.1714109114.

Lloyd M, Kidd D (2012). The mouse lymphoma assay. Methods Mol Biol. 817:35–54. doi:10.1007/978-1-61779-421-6_3.

Lorge E, Moore MM, Clements J, O'Donovan M, Fellows MD, Honma M et al. (2016). Standardized cell sources and recommendations for good cell culture practices in genotoxicity testing. Mutat Res. 809:1–16. doi:10.1016/j.mrgentox.2016.08.001.

Luo JL, Yang Q, Tong WM, Hergenhahn M, Wang ZQ, Holstein M (2001). Knock-in mice with a chimeric human/murine *p53* gene develop normally and show wild-type *p53* responses to DNA damaging agents: a new biomedical research tool. Oncogene. 20:320–8. doi:10.1038/sj.onc.1204080.

Lupski JR (2013). Genetics. Genome mosaicism – one human, multiple genomes. Science. 341(6144):358–9. doi:10.1126/science.1239503.

MacGregor JT, Bishop ME, McNamee JP, Hayashi M, Asano N, Wakata A et al. (2006). Flow cytometric analysis of micronuclei in peripheral blood reticulocytes: II. An efficient method of monitoring chromosomal damage in the rat. Toxicol Sci. 94(1):92–107. doi:10.1093/toxsci/kfl076.

MacGregor JT, Frötschl R, White PA, Crump KS, Eastmond DA, Fukushima S et al. (2015a). IWGT report on quantitative approaches to genotoxicity risk assessment. I. Methods and metrics for defining exposure–response relationships and points of departure (PoDs). Mutat Res. 783:55–65. doi:10.1016/j.mrgentox.2014.09.011.

MacGregor JT, Frötschl R, White PA, Crump KS, Eastmond DA, Fukushima S et al. (2015b). IWGT report on quantitative approaches to genotoxicity risk assessment. II. Use of point-of-departure (PoD) metrics in defining acceptable exposure limits and assessing human risk. Mutat Res. 783:66–78. doi:10.1016/j.mrgentox.2014.10.008.

Mansouri K, Abdelaziz A, Rybacka A, Roncaglioni A, Tropsha A, Varnek A et al. (2016). CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. Environ Health Perspect. 124(7):1023–33. doi:10.1289/ehp.1510267.

Marchetti F, Aardema MJ, Beevers C, Van Benthem J, Godschalk R, Williams A et al. (2018). Identifying germ cell mutagens using OECD test guideline 488 (Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays) and integration with somatic cell testing. Mutat Res. 832–833:7–18. doi:10.1016/j.mrgentox.2018.05.021.

Masumura K, Toyoda-Hokaiwado N, Ukai A, Gondo Y, Honma M, Nohmi T (2016a). Estimation of the frequency of inherited germline mutations by whole exome sequencing in ethyl nitrosourea–treated and untreated *gpt* delta mice. Genes Environ. 38:10. doi:10.1186/s41021-016-0035-y.

Masumura K, Toyoda-Hokaiwado N, Ukai A, Gondo Y, Honma M, Nohmi T (2016b). Dose-dependent de novo germline mutations detected by whole-exome sequencing in progeny of ENU-treated male *gpt* delta mice. Mutat Res. 810:30–9. doi:10.1016/j.mrgentox.2016.09.009.

Meek ME, Boobis AR, Crofton KM, Heinemeyer G, Van Raaij M, Vickers C (2011). Risk assessment of combined exposure to multiple chemicals: a WHO/IPCS framework. Regul Toxicol Pharmacol. 60:S1–S14. doi:10.1016/j.yrtph.2011.03.010.

Meek ME, Palermo CM, Bachman AN, North CM, Lewis RJ (2014). Mode of action human relevance (species concordance) framework: evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. J Appl Toxicol. 34(6):595–606. doi:10.1002/jat.2984.

Meier MJ, O'Brien JM, Beal MA, Allan B, Yauk CL, Marchetti F (2017). In utero exposure to benzo[*a*]pyrene increases mutation burden in the soma and sperm of adult mice. Environ Health Perspect. 125:82–8. doi:10.1289/EHP211.

Mellor CL, Marchese Robinson RL, Benigni R, Ebbrell D, Enoch SJ, Firman JW et al. (2019). Molecular fingerprint–derived similarity measures for toxicological read-across: recommendations for optimal use. Regul Toxicol Pharmacol. 101:121–34. doi:10.1016/j.yrtph.2018.11.002.

Miura D, Dobrovolsky VN, Kasahara Y, Katsuura Y, Heflich RH (2008a). Development of an in vivo gene mutation assay using the endogenous *Pig-A* gene: I. Flow cytometric detection of CD59-negative peripheral red blood cells and CD48-negative spleen T-cells from the rat. Environ Mol Mutagen. 49:614–21. doi:10.1002/em.20414.

Miura D, Dobrovolsky VN, Mittelstaedt RA, Kasahara Y, Katsuura Y, Heflich RH (2008b). Development of an in vivo gene mutation assay using the endogenous *Pig-A* gene: II. Selection of *Pig-A* mutant rat spleen T-cells with proaerolysin and sequencing *Pig-A* cDNA from the mutants. Environ Mol Mutagen. 49:622–30. doi:10.1002/em.20413.

Molander L, Ågerstrand M, Beronius A, Hanberg A, Rudén C (2015). Science in Risk Assessment and Policy (SciRAP): an online resource for evaluating and reporting in vivo (eco)toxicity studies. Hum Ecol Risk Assess. 21(3):753–62. doi:10.1080/10807039.2014.928104.

Mombelli E, Raitano G, Benfenati E (2016). In silico prediction of chemically induced mutagenicity: how to use QSAR models and interpret their results. Methods Mol Biol. 1425:87–105. doi:10.1007/978-1-4939-3609-0_5.

Morita T, MacGregor JT, Hayashi M (2011). Micronucleus assays in rodent tissues other than bone marrow. Mutagenesis. 26(1):223–30. doi:10.1093/mutage/geq066.

Motoyama S, Takeiri A, Tanaka K, Harada A, Matsuzaki K, Taketo J et al. (2018). Advantages of evaluating γH2AX induction in non-clinical drug development. Genes Environ. 40:10. doi:10.1186/s41021-018-0098-z.

Muller HJ (1927). Artificial transmutation of the gene. Science. 66(1699):84–7. doi:10.1126/science.66.1699.84.

Müller L, Gocke E (2009). Considerations regarding a permitted daily exposure calculation for ethyl methanesulfonate. Toxicol Lett. 190(3):330–2. doi:10.1016/j.toxlet.2009.03.015.

Myatt GJ, Ahlberg E, Akahori Y, Allen D, Amberg A, Anger LT et al. (2018). In silico toxicology protocols. Regul Toxicol Pharmacol. 96:1–17. doi:10.1016/j.yrtph.2018.04.014.

Nelms MD, Mellow CL, Enoch SJ, Judson RS, Patlewicz G, Richard AM et al. (2018). A mechanistic framework for integrating chemical structure and high-throughput screening results to improve toxicity predictions. Comput Toxicol. 8:1–12. doi:10.1016/j.comtox.2018.08.003.

Nervi C, Fazi F, Grignani F (2008). Oncoproteins, heterochromatin silencing and microRNAs: a new link for leukemogenesis. Epigenetics. 3(1):1–4. doi:10.4161/epi.3.1.5651.

Netzeva TI, Worth A, Aldenberg T, Benigni R, Cronin MTD, Gramatica P et al. (2005). Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships. The report and recommendations of ECVAM Workshop 52. Altern Lab Anim. 33(2):155–73. doi:10.1177/026119290503300209.

Nielsen E, Ostergaard G, Larsen J (2008). Toxicological risk assessment of chemicals: a practical guide. New York (NY): Informa Healthcare USA.

Nishihara K, Huang R, Zhao J, Shahane SA, Witt KL, Smith-Roe SL et al. (2016). Identification of genotoxic compounds using isogenic DNA repair deficient DT40 cell lines on a quantitative high throughput screening platform. Mutagenesis. 31(1):69–81. doi:10.1093/mutage/gev055.

Nohmi H (2008). Possible mechanisms of practical thresholds for genotoxicity. Genes Environ. 30(4):108–13. doi:10.3123/jemsge.30.108.

Nohmi T (2018). Thresholds of genotoxic and non-genotoxic carcinogens. Toxicol Res. 34(4):281–90. doi:10.5487/tr.2018.34.4.281.

Nohmi T, Tsuzuki T (2016). Possible mechanisms underlying genotoxic thresholds: DNA repair and translesion DNA synthesis. In: Nohmi T, Fukushima S, editors. Thresholds of genotoxic carcinogens. From mechanisms to regulation. Tokyo: Academic Press; 49–66.

Nohmi T, Masumura K, Toyoda-Hokaiwado N (2017). Transgenic rat models for mutagenesis and carcinogenesis. Genes Environ. 39:11. doi:10.1186/s41021-016-0072-6.

Nohmi T, Suzuki T, Masumura K (2000). Recent advances in the protocols of transgenic mouse mutation assays. Mutat Res. 455(1–2):191–215. doi:10.1016/S0027-5107(00)00077-4.

OECD (2005). 3.1 Guidance for determining the quality of data for the SIDS dossiers (reliability, relevance and adequacy). In: Manual for investigation of HPV chemicals. Paris: Organisation for Economic Co-operation and Development (http://www.oecd.org/chemicalsafety/risk-assessment/49191960.pdf).

OECD (2007). Guidance document on the validation of (quantitative)structure–activity relationships [(Q)SAR] models. Paris: Organisation for Economic Co-operation and Development (ENV/JM/MONO(2007)2; https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono%282007%292&doclanguage=en).

OECD (2011). 4.3.6 Genetic toxicity. In: Manual for the assessment of chemicals. Paris: Organisation for Economic Co-operation and Development; 15 (https://www.oecd.org/chemicalsafety/risk-assessment/49188998.pdf).

OECD (2014a). Test No. 489: In vivo mammalian alkaline comet assay. Paris: Organisation for Economic Co-operation and Development (https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-tg489-2014.pdf).

OECD (2014b). Guidance on grouping of chemicals, second edition. Paris: Organisation for Economic Co-operation and Development (ENV/JM/MONO(2014)4; OECD Series on Testing & Assessment, No. 194; https://www.oecd-ilibrary.org/docserver/9789264274679-en.pdf?expires=1594050117&id=id&accname=guest&checksum=02B76DA824F325DEFC826C2A1E14F463).

OECD (2016a). Test No. 473: In vitro mammalian chromosomal aberration test. Paris: Organisation for Economic Co-operation and Development (https://www.oecd-ilibrary.org/docserver/9789264264649-en.pdf?expires=1593166230&id=id&accname=guest&checksum=44D55AE3906A5EDAC33CCB80A3EF6A47).

OECD (2016b) Test No. 489: In vivo mammalian alkaline comet assay. Paris: Organisation for Economic Co-operation and Development (https://www.oecd-ilibrary.org/docserver/9789264264885-en.pdf?expires=1593166181&id=id&accname=guest&checksum=D4F62ED21A519E97F6CF00B745D868F8).

OECD (2017a). Overview of the set of OECD genetic toxicology test guidelines and updates performed in 2014–2015. Paris: Organisation for Economic Co-operation and Development (ENV/JM/MONO(2016)33/REV1; OECD Series on Testing & Assessment, No. 238, second edition; http://www.oecd.org/

officialdocuments/publicdisplaydocumentpdf/?cote=ENV-JM-MONO(2016)33/rev1&doclanguage=en).

OECD (2017b). Report on considerations from case studies on integrated approaches for testing and assessment (IATA) – first review cycle (2015). Case studies on grouping methods as a part of IATA. Paris: Organisation for Economic Co-operation and Development (ENV/JM/MONO(2016)48; OECD Series on Testing & Assessment, No. 250; https://www.oecd-ilibrary.org/environment/report-on-considerations-from-case-studies-on-integrated-approaches-for-testing-and-assessment-iata-first-review-cycle-2015_9789264274815-en).

OECD (2017c). Report on considerations from case studies on integrated approaches for testing and assessment (IATA) – second review cycle (2016). Paris: Organisation for Economic Co-operation and Development (ENV/JM/MONO(2017)22; OECD Series on Testing & Assessment, No. 270; http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2017)22&doclanguage=en).

OECD (2017d). Revised guidance document on developing and assessing adverse outcome pathways. Paris: Organisation for Economic Co-operation and Development (ENV/JM/MONO(2013)6; OECD Series on Testing & Assessment, No. 184; http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en).

OECD (2018a). Report on considerations from case studies on integrated approaches for testing and assessment (IATA) – third review cycle (2017). Paris: Organisation for Economic Co-operation and Development (ENV/JM/MONO(2018)25; OECD Series on Testing & Assessment, No. 289; http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2018)25&docLanguage=En).

OECD (2018b). Considerations for assessing the risks of combined exposure to multiple chemicals. Paris: Organisation for Economic Co-operation and Development (OECD Series on Testing and Assessment, No. 296; https://www.oecd.org/chemicalsafety/risk-assessment/considerations-for-assessing-the-risks-of-combined-exposure-to-multiple-chemicals.pdf).

OECD (2019). Report on considerations from case studies on integrated approaches for testing and assessment (IATA) – fourth review cycle (2018). Paris: Organisation for Economic Co-operation and Development (ENV/JM/MONO(2019)26; OECD Series on Testing & Assessment, No. 307; http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2019)26&docLanguage=en).

Ohta T (2006). Mutagenic activity of a mixture of heterocyclic amines at doses below the biological threshold level of each. Genes Environ. 28(4):181–4. doi:10.3123/jemsge.28.181.

Olsen AK, Dertinger SD, Kruger CT, Eide EM, Instanes C, Brunborg G et al. (2017). The *Pig-a* gene mutation assay in mice and human cells: a review. Basic Clin Pharmacol Toxicol. 121(Suppl 3):78–92. doi:10.1111/bcpt.12806.

Pariza MW, Foster EM (1983). Determining the safety of enzymes used in food processing. J Food Prot. 46(5):453–68. doi:10.4315/0362-028X-46.5.453.

Pariza MW, Johnson EA (2001). Evaluating the safety of microbial enzyme preparations used in food processing: update for a new century. Regul Toxicol Pharmacol. 33(2):173–86. doi:10.1006/rtph.2001.1466.

Parry JM, Fielder RJ, McDonald A (1994). Thresholds for aneuploidy-inducing chemicals. Mutagenesis. 9(6):503–4. doi:10.1093/mutage/9.6.503.

Patlewicz G, Fitzpatrick JM (2016). Current and future perspectives on the development, evaluation, and application of in silico approaches for predicting toxicity. Chem Res Toxicol. 29(4):438–51. doi:10.1021/acs.chemrestox.5b00388.

Patlewicz G, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MIH et al. (2007). TIMES-SS – a promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. Regul Toxicol Pharmacol. 48(2):225–39. doi:10.1016/j.yrtph.2007.03.003.

Patlewicz G, Ball N, Becker RA, Booth ED, Cronin MTD, Kroese D et al. (2014). Food for thought: read-across approaches – misconceptions, promises and challenges ahead. ALTEX. 31(4):387–96. doi:10.14573/altex.1410071.

Patlewicz G, Ball N, Boogaard PJ, Becker RA, Hubesch B (2015). Building scientific confidence in the development and evaluation of read-across. Regul Toxicol Pharmacol. 72(1):117–33. doi:10.1016/j.yrtph.2015.03.015.

Patlewicz G, Helman G, Pradeep P, Shah I (2017). Navigating through the minefield of read-across tools: a review of in silico tools for grouping. Comp Toxicol. 3:1–18. doi:10.1016/j.comtox.2017.05.003.

Patlewicz G, Cronin MTD, Helman G, Lambert JC, Lizarraga LE, Shah I (2018). Navigating through the minefield of read-across frameworks: a commentary perspective. Comp Toxicol. 6:39–54. doi:10.1016/j.comtox.2018.04.002.

Patlewicz G, Lizarraga LE, Rua D, Allen DG, Daniel AB, Fitzpatrick SC et al. (2019). Exploring current read-across applications and needs among selected U.S. federal agencies. Regul Toxicol Pharmacol. 106:197–209. doi:10.1016/j.yrtph.2019.05.011.

Pawar G, Madden JC, Ebbrell DJ, Firman JW, Cronin MTD (2019). In silico toxicology data resources to support read-across and (Q)SAR. Front Pharmacol. 10:561. doi:10.3389/fphar.2019.00561.

PerkinElmer-Cisbio (2020). Phospho-H2AX (Ser139) cellular kit. Waltham (MA): PerkinElmer-Cisbio (https://ca.cisbio.net/phospho-h2ax-ser139-cellular-kit-40571#product-insert-assets).

Petkov PI, Kuseva C, Kotov S, Honma M, Kitazawa A, Kulkarni S et al. (2019). Procedure for toxicological predictions based on mechanistic weight of evidences: application to Ames mutagenicity. Comp Toxicol. 12:100009. doi:10.1016/j.comtox.2017.02.004.

Pfuhler S, Fautz R, Ouedraogo G, Latil A, Kenny J, Moore C et al. (2014). The Cosmetics Europe strategy for animal-free genotoxicity testing: project status up-date. Toxicol In Vitro. 28(1):18–23. doi:10.1016/j.tiv.2013.06.004.

Phillips DH (1997). Detection of DNA modifications by the ^{32}P-postlabelling assay. Mutat Res. 378(1–2):1–12. doi:10.1016/s0027-5107(97)00092-4.

Phillips DH, Farmer PB, Beland FA, Nath RG, Poirier MC, Reddy MV et al. (2000). Methods of DNA adduct determination and their application to testing compounds for genotoxicity. Environ Mol Mutagen. 35:222–33. doi:10.1002/(SICI)1098-2280(2000)35:3<222::AID-EM9>3.0.CO;2-E.

Pinter E, Rainer B, Czerny T, Riegel E, Schilter B, Marin-Kuan M et al. (2020). Evaluation of the suitability of mammalian in vitro assays to assess the genotoxic potential of food contact materials. Foods. 9(2):237. doi:10.3390/foods9020237.

Powley MW (2015). (Q)SAR assessments of potentially mutagenic impurities: a regulatory perspective on the utility of expert knowledge and data submission. Regul Toxicol Pharmacol. 71(2):295–300. doi:10.1016/j.yrtph.2014.12.012.

Pozniak A, Müller L, Salgo M, Jones JK, Larson P, Tweats D (2009). Elevated ethyl methanesulfonate (EMS) in nelfinavir mesylate (Viracept®, Roche): overview. AIDS Res Ther. 6:18. doi:10.1186/1742-6405-6-18.

Raitano G, Roncaglioni A, Manganaro A, Honma M, Sousselier L, Do QT et al. (2019). Integrating in silico models for the prediction of mutagenicity (Ames test) of botanical ingredients of cosmetics. Comput Toxicol. 12:100108. doi:10.1016/j.comtox.2019.100108.

Rappaport SM, Li H, Grigoryan H, Funk WE, Williams ER (2012). Adductomics: characterizing exposures to reactive electrophiles. Toxicol Lett. 213(1):83–90. doi:10.1016/j.toxlet.2011.04.002.

Reisinger K, Blatz V, Brinkmann J, Downs TR, Fischer A, Henkler F et al. (2018). Validation of the 3D Skin Comet assay using full thickness skin models: transferability and reproducibility. Mutat Res. 827:27–41. doi:10.1016/j.mrgentox.2018.01.003.

Richmond J (2002). Refinement, reduction, and replacement of animal use for regulatory testing: future improvements and implementation within the regulatory framework. ILAR J. 43(Suppl 1):S63–S68. doi:10.1093/ilar.43.Suppl_1.S63.

Riebeling C, Luch A, Tralau T (2018). Skin toxicology and 3Rs – Current challenges for public health protection. Exp Dermatol. 27(5):526–36. doi:10.1111/exd.13536.

Ritter L, Totman C, Krishnan K, Carrier R, Vézina A, Morisset V (2007). Deriving uncertainty factors for threshold chemical contaminants in drinking water. J Toxicol Environ Health B Crit Rev. 10(7):527–57. doi:10.1080/15287390600975178.

Russell LB (2004). Effects of male germ-cell stage on the frequency, nature, and spectrum of induced specific-locus mutations in the mouse. Genetica. 122(1):25–36. doi:10.1007/s10709-004-1443-7.

Russell LB, Shelby MD (1985). Tests for heritable genetic damage and for evidence of gonadal exposure in mammals. Mutat Res. 154(2):69–84. doi:10.1016/0165-1110(85)90020-X.

Sakuratani Y, Horie M, Leinala E (2018). Integrated approaches to testing and assessment: OECD activities on the development and use of adverse outcome pathways and case studies. Basic Clin Pharmacol Toxicol. 123(Suppl 5):20–8. doi:10.1111/bcpt.12955.

Salk JJ, Schmitt MW, Loeb LA (2018). Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nat Rev Genet. 19(5):269–85. doi:10.1038/nrg.2017.117.

SCCS (2018). The SCCS notes of guidance for the testing of cosmetic ingredients and their safety evaluation, tenth revision. Luxembourg: European Commission, Scientific Committee on Consumer Safety (SCCS/1602/18; https://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_224.pdf).

SCF (1991). Food science and techniques. Reports of the Scientific Committee for Food (twenty-seventh series). Luxembourg: Commission of the European Communities, Scientific Committee for Food (EUR 14181 EN; http://aei.pitt.edu/40836/1/27th_good.pdf).

Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A et al. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicol Lett. 189(2):138–44. doi:10.1016/j.toxlet.2009.05.013.

Schultz TW, Richarz A-N, Cronin MTD (2019). Assessing uncertainty in read-across: questions to evaluate toxicity predictions based on knowledge gained from case studies. Comp Toxicol. 9:1–11. doi:10.1016/j.comtox.2018.10.003.

Schultz TW, Amcoff P, Berggren E, Gautier F, Klaric M, Knight DJ et al. (2015). A strategy for structuring and reporting a read-across prediction of toxicity. Regul Toxicol Pharmacol. 72(3):586–601. doi:10.1016/j.yrtph.2015.05.016.

Serafimova R, Fuart Gatnik M, Worth A (2010). Review of QSAR models and software tools for predicting genotoxicity and carcinogenicity. Luxembourg: European Union, Joint Research Centre (EUR 24427 EN; https://publications.jrc.ec.europa.eu/repository/bitstream/JRC59068/jrc%20report_genotoxicity%20qsar.pdf).

Shah I, Liu J, Judson RS, Thomas RS, Patlewicz G (2016). Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information. Regul Toxicol Pharmacol. 79:12–24. doi:10.1016/j.yrtph.2016.05.008.

Shelby MD (1996). Selecting chemicals and assays for assessing mammalian germ cell mutagenicity. Mutat Res. 352(1–2):159–67. doi:10.1016/0027-5107(95)00222-7.

Simpson K, Bevan N, Hastwell P, Eidam P, Shah P, Gogo E et al. (2013). The BlueScreen-384 assay as an indicator of genotoxic hazard potential in early-stage drug discovery. J Biomol Screen. 18(4):441–52. doi:10.1177/1087057112470858.

Smith-Roe SL, Swartz CD, Shepard KG, Bryce SM, Dertinger SD, Waidyanatha S et al. (2018). Black cohosh extracts and powders induce micronuclei, a biomarker of genetic damage, in human cells. Environ Mol Mutagen. 59(5):416–26. doi:10.1002/em.22182.

Strauch BM, Niemand RK, Winkelbeiner NL, Hartwig A (2017). Comparison between micro- and nanosized copper oxide and water soluble copper chloride: interrelationship between intracellular copper concentrations, oxidative stress and DNA damage response in human lung cells. Part Fibre Toxicol 14:28. doi:10.1186/s12989-017-0209-1.

Sukup-Jackson MR, Kiraly O, Kay JE, Na L, Rowland EA, Winther KE et al. (2014). Rosa26-GFP direct repeat (RaDR-GFP) mice reveal tissue- and age-dependence of homologous recombination in mammals in vivo. PLOS Genetics. 10(6):e1004299. doi:10.1371/journal.pgen.1004299.

Sutter A, Amberg A, Boyer S, Brigo A, Contrera JF, Custer LL et al. (2013). Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Regul Toxicol Pharmacol. 67(1):39–52. doi:10.1016/j.yrtph.2013.05.001.

Suzuki Y, Umemura T, Ishii Y, Hibi D, Inoue T, Jin M et al. (2012). Possible involvement of sulfotransferase 1A1 in estragole-induced DNA modification and carcinogenesis in the livers of female mice. Mutat Res. 749(1–2):23–8. doi:10.1016/j.mrgentox.2012.07.002.

Swenberg JA (1993). Alpha 2u-globulin nephropathy: review of the cellular and molecular mechanisms involved and their implications for human risk assessment. Environ Health Perspect. 101(Suppl 6):39–44. doi:10.1289/ehp.93101s639.

Sykora P, Witt KL, Revanna P, Smith-Roe SL, Dismukes J, Lloyd DF et al. (2018). Next generation high throughput DNA damage detection platform for genotoxic compound screening. Sci Rep. 8(1):2771. doi:10.1038/s41598-018-20995-w.

Takeshita T, Tao F, Kojima N, Kanaly RA (2019). Triple quadrupole mass spectrometry comparative DNA adductomics of Hep G2 cells following exposure to safrole. Toxicol Lett. 300:92–104. doi:10.1016/j.toxlet.2018.10.023.

Tcheremenskaia O, Battistelli CL, Giuliani A, Benigni R, Bossa C (2019). In silico approaches for prediction of genotoxic and carcinogenic potential of cosmetic ingredients. Comput Toxicol. 11:91–100. doi:10.1016/j.comtox.2019.03.005.

Teasdale A, Elder DP (2018). Analytical control strategies for mutagenic impurities: current challenges and future opportunities? Trends Anal Chem. 101:66–84. doi:10.1016/j.trac.2017.10.027.

Thougaard AV, Christiansen J, Mow T, Hornberg JJ (2014). Validation of a high throughput flow cytometric in vitro micronucleus assay including assessment of metabolic activation in TK6 cells. Environ Mol Mutagen. 55(9):704–18. doi:10.1002/em.21891.

Thybaud V, Dean S, Nohmi T, de Boer J, Douglas GR, Glickman BW et al. (2003). In vivo transgenic mutation assays. Mutat Res. 540(2):141–51. doi:10.1016/j.mrgentox.2003.07.004.

Ton ND, Nakagawa H, Ha NH, Duong NT, Nhung VP, Hien LTT et al. (2018). Whole genome sequencing and mutation rate analysis of trios with paternal dioxin exposure. Hum Mutat. 39(10):1384–92. doi:10.1002/humu.23585.

Torous DK, Dertinger SD, Hall NE, Tometsko CR (2000). Enumeration of micronucleated reticulocytes in rat peripheral blood: a flow cytometric study. Mutat Res. 465(1–2):91–9. doi:10.1016/s1383-5718(99)00216-8.

Tweats DJ, Johnson GE, Scandale I, Whitwell J, Evans DB (2016). Genotoxicity of flubendazole and its metabolites in vitro and the impact of a new formulation on in vivo aneugenicity. Mutagenesis. 31(3):309–21. doi:10.1093/mutage/gev070.

UKCOM (2010). Guidance statement: thresholds for in vivo mutagens. Chilton: United Kingdom Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/315698/assessment_of_threshold_for_in_vivo_mutagens.pdf).

UKCOM (2018). Statement on the quantitative approaches to the assessment of genotoxicity data. Chilton: United Kingdom Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment (Statement 2018/S1; https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/704462/COM_statement_on_quantitative_approaches.pdf).

Uno Y, Morita T, Luijten M, Beevers C, Hamada S, Itoh S et al. (2015a). Recommended protocols for the liver micronucleus test: report of the IWGT working group. Mutat Res. 783:13–8. doi:10.1016/j.mrgentox.2014.10.010.

Uno Y, Morita T, Luijten M, Beevers C, Hamada S, Itoh S et al. (2015b). Micronucleus test in rodent tissues other than liver or erythrocytes: report of the IWGT working group. Mutat Res. 783:19–22. doi:10.1016/j.mrgentox.2015.03.001.

USEPA (2019). OncoLogic™ – A computer system to evaluate the carcinogenic potential of chemicals. Washington (DC): United States Environmental Protection Agency (https://www.epa.gov/tsca-screening-tools/oncologictm-computer-system-evaluate-carcinogenic-potential-chemicals).

USFDA (1995). United States Food and Drug Administration; Food additives: threshold of regulation of substances used in food-contact articles: final rule. Fed Regist. 60(136):36582–96 (https://www.govinfo.gov/content/pkg/FR-1995-07-17/pdf/95-17435.pdf).

USFDA (2007). Toxicological principles for the safety assessment of food ingredients. Redbook 2000: IV.C.1. Short-term tests for genetic toxicity. Rockville (MD): United States Food and Drug Administration (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/redbook-2000-ivc1-short-term-tests-genetic-toxicity).

USFDA (2016). Botanical drug development: guidance for industry. Silver Spring (MD): United States Food and Drug Administration (https://www.fda.gov/downloads/Drugs/Guidances/UCM458484.pdf).

USNTP (2011). 12th report on carcinogens. Research Triangle Park (NC): United States Department of Health and Human Services, Public Health Service, National Toxicology Program (https://oehha.ca.gov/media/downloads/proposition-65/crnr/comments/12throc-complete.pdf).

Van Bossuyt M, Van Hoeck E, Raitano G, Vanhaecke T, Benfenati E, Mertens B et al. (2018). Performance of in silico models for mutagenicity prediction of food contact materials. Toxicol Sci. 163(2):632–8. doi:10.1093/toxsci/kfy057.

Verma JR, Rees BJ, Wilde EC, Thornton CA, Jenkins GJS, Doak SH et al. (2017). Evaluation of the automated MicroFlow® and Metafer™ platforms for high-throughput micronucleus scoring and dose response analysis in human lymphoblastoid TK6 cells. Arch Toxicol. 91(7):2689–98. doi:10.1007/s00204-016-1903-8.

Vuorinen A, Bellion P, Beilstein P (2017). Applicability of in silico genotoxicity models on food and feed ingredients. Regul Toxicol Pharmacol. 90:277–88. doi:10.1016/j.yrtph.2017.09.026.

Wang NCY, Zhao QJ, Wesselkamper SC, Lambert JC, Petersen D, Hess-Wilson JK (2012). Application of computational toxicological approaches in human health risk assessment. I. A tiered surrogate approach. Regul Toxicol Pharmacol. 63(1):10–9. doi:10.1016/j.yrtph.2012.02.006.

Wetmore BA (2015). Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. Toxicology. 332:94–101. doi:10.1016/j.tox.2014.05.012.

White PA, Johnson GE (2016). Genetic toxicology at the crossroads – from qualitative hazard evaluation to quantitative risk assessment. Mutagenesis. 31(3):233–7. doi:10.1093/mutage/gew011.

White PA, Luijten M, Mishima M, Cox JA, Hanna JN, Maertens RM et al. (2019). In vitro mammalian cell mutation assays based on transgenic reporters:

a report of the International Workshop on Genotoxicity Testing (IWGT). Mutat Res. 847:403039. doi:10.1016/j.mrgentox.2019.04.002.

WHO (2015a). Pesticide residues in food: Guidance document for WHO monographers and reviewers. Geneva: World Health Organization (https://www.who.int/foodsafety/publications/jmpr_guidance_document_1.pdf?ua=1).

WHO (2015b). Pesticide residues in food – 2014 evaluations. Part II – Toxicological. Geneva: World Health Organization (https://apps.who.int/iris/bitstream/handle/10665/164597/9789241665308_eng.pdf;jsessionid=65B92CA6D1D058F62646E2BCBBAB888D?sequence=1).

WHO (2016). Pesticide residues in food – 2015 evaluations. Part II – Toxicological. Geneva: World Health Organization (http://apps.who.int/iris/bitstream/10665/205165/1/9789241655316_eng.pdf?ua=1).

WHO (2017). Guidance to JECFA experts on systematic literature searches. Geneva: World Health Organization, JECFA Secretariat (https://www.who.int/foodsafety/chem/jecfa/Litertature_Search.pdf).

WHO (2019). Pesticide residues in food – 2018 evaluations. Part II – Toxicological. Geneva: World Health Organization (https://apps.who.int/iris/bitstream/handle/10665/325787/9789241655330-eng.pdf?ua=1).

Wichard J (2017). In silico prediction of genotoxicity. Food Chem Toxicol. 106(Part B):595–9. doi:10.1016/j.fct.2016.12.013.

Williams A, Buick JK, Moffat I, Swartz CD, Recio L, Hyduke DR et al. (2015). A predictive toxicogenomics signature to classify genotoxic versus non-genotoxic chemicals in human TK6 cells. Data Brief. 5:77–83. doi:10.1016/j.dib.2015.08.013.

Williams JG, Deschl U, Williams GM (2011). DNA damage in fetal liver cells of turkey and chicken eggs dosed with aflatoxin B_1. Arch Toxicol. 85:1167–72. doi:10.1007/s00204-011-0653-x.

Wills JW, Johnson GE, Doak SH, Soeteman-Hernández LG, Slob W, White PA (2016). Empirical analysis of BMD metrics in genetic toxicology. Part I: In vitro analyses to provide robust potency rankings and support MOA determinations. Mutagenesis. 31(3):255–63. doi:10.1093/mutage/gev085.

Witt KL, Knapton A, Wehr CM, Hook GJ, Mirsalis J, Shelby MD et al. (2000). Micronucleated erythrocyte frequency in peripheral blood of $B6C3F_1$ mice from short-term, prechronic, and chronic studies of the NTP carcinogenesis bioassay program. Environ Mol Mutagen. 36(3):163–94. doi:10.1002/1098-2280(2000)36:3<163::AID-EM1>3.0.CO;2-P.

Witt KL, Cunningham CK, Patterson KB, Kissling GE, Dertinger SD, Livingston E et al. (2007). Elevated frequencies of micronucleated erythrocytes in infants exposed to zidovudine in utero and postpartum to prevent mother-to-child transmission of HIV. Environ Mol Mutagen. 48(3–4):322–9. doi:10.1002/em.20266.

Witt KL, Livanos E, Kissling GE, Torous DK, Caspary W, Tice RR et al. (2008). Comparison of flow cytometry– and microscopy-based methods for measuring micronucleated reticulocyte frequencies in rodents treated with nongenotoxic and genotoxic chemicals. Mutat Res. 649(1–2):101–13. doi:10.1016/j.mrgentox.2007.08.004.

Witt KL, Hsieh JH, Smith-Roe SL, Xia M, Huang R, Zhao J et al. (2017). Assessment of the DNA damaging potential of environmental chemicals using a quantitative high-throughput screening approach to measure p53 activation. Environ Mol Mutagen. 58:494–507. doi:10.1002/em.22112.

Wolf T, Luepke N-P (1997). Formation of micronuclei in incubated hen's eggs as a measure of genotoxicity. Mutat Res. 394(1–3):163–75. doi:10.1016/S1383-5718(97)00136-8.

Wolf T, Niehaus-Rolf C, Luepke N-P (2003). Investigating genotoxic and hematotoxic effects of *N*-nitrosodimethylamine, *N*-nitrosodiethylamine and *N*-nitrosodiethanolamine in the hen's egg micronucleus test (HET-MN). Food Chem Toxicol. 41(4):561–73. doi:10.1016/S0278-6915(02)00281-8.

Worth A, Fuart-Gatnik M, Lapenna S et al. (2010). Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Scientific report submitted to the European Food Safety Authority. Ispra: European Commission, Joint Research Centre, Institute for Health & Consumer Protection (https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2010.EN-50).

Wu S, Blackburn K, Amburgey J, Jaworska J, Federle T (2010). A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. Regul Toxicol Pharmacol. 56(1):67–81. doi:10.1016/j.yrtph.2009.09.006.

Yamamoto KN, Hirota K, Kono K, Takeda S, Sakamuru S, Xia M et al. (2011). Characterization of environmental chemicals with potential or DNA damage using isogenic DNA repair–deficient chicken DT40 cell lines. Environ Mol Mutagen. 52(7):547–61. doi:10.1002/em.20656.

Yao C, Feng YL (2016). A nontargeted screening method for covalent DNA adducts and DNA modification selectivity using liquid chromatography–tandem mass spectrometry. Talanta. 159:93–102. doi:10.1016/j.talanta.2016.05.074.

Yauk CL (2004). Advances in the application of germline tandem repeat instability for in situ monitoring. Mutat Res. 566(2):169–82. doi:10.1016/j.mrrev.2003.08.001.

Yauk CL, Buick JK, Williams A, Swartz CD, Recio L, Li HH et al. (2016a). Application of the TGx-28.65 transcriptomic biomarker to classify genotoxic and non-genotoxic chemicals in human TK6 cells in the presence of rat liver S9. Environ Mol Mutagen. 57(4):243–60. doi:10.100/Fem.22004.

Yauk C, Lambert I, Marchetti F, Douglas G (2016b). Adverse outcome pathway on alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations. Paris: OECD Publishing (OECD Series on Adverse Outcome Pathways, No. 3; https://doi.org/10.1787/5jlsvvxn1zjc-en).

Youssoufian H, Pyeritz RE (2002). Mechanisms and consequences of somatic mosaicism in humans. Nat Rev Genet. 3:748–58. doi:10.1038/nrg906.

Yu Y, Wang P, Cui Y, Wang Y (2018). Chemical analysis of DNA damage. Anal Chem. 90(1):556–76. doi:10.1021/acs.analchem.7b04247.

Zeiger E (2004). History and rationale of genetic toxicity testing: an impersonal, and sometimes personal, view. Environ Mol Mutagen. 44(5):363–71. doi: 10.1002/em.20062.

Zhu H, Bouhifd M, Donley E, Egnash L, Kleinstreuer N, Kroese DE et al. (2016). Supporting read-across using biological data. ALTEX. 33(2):167–82. doi:10.14573/altex.1601252.

4.6 Carcinogenicity

4.6.1 Introduction

The purpose of testing chemicals for carcinogenicity in experimental animals is to identify potential cancer hazards for humans. Tests are usually conducted for the majority of the lifetime of experimental animals at high multiples of potential human exposures. Under these conditions, the absence of cancer indicates a likely absence of human risk. Positive findings require careful interpretation in relation to mode of action, possible interspecies differences in background incidence and in response and high dose to low dose extrapolation. Virtually all chemicals associated with cancer in humans have been found to increase the incidence of neoplasms in experimental animals (McGregor et al., 1999), although not necessarily the same type of tumour is seen in exposed humans. Accordingly, chronic cancer bioassays are established as relevant for human hazard identification and characterization.

4.6.2 Mechanisms of carcinogenicity and mode of action

In the early days of chemical carcinogenesis, it was initially suspected that carcinogens operated through a common mechanism (Miller & Miller, 1979). With advances in the understanding of the molecular effects of carcinogens, concepts of differing modes of tumour induction were developed (Williams, 1992). It is now widely accepted that two general types of mode of action can be distinguished—genotoxic mechanisms involving chemical interaction of the carcinogen with DNA, and non-genotoxic mechanisms involving other cellular and extracellular effects (Vaino et al., 1992). These different modes of action have major implications for hazard characterization, because a biological threshold is believed to occur for non-genotoxic mechanisms, and a level of human exposure without significant risk can be established. As a precautionary approach, it is considered that a threshold may not exist for direct-acting (alkylating) genotoxic chemicals or that if a threshold does exist, it may be below the level of human exposure; in consequence, any level of human exposure could be associated with some degree of risk. In contrast, a threshold might exist for some forms of genetic damage (genotoxicity) that do not result in potentially irreversible change to DNA leading to a mutation.

The concept of initiation and promotion as distinct steps in carcinogenesis was developed in mouse skin, and a two-step or multistep process is now known to occur in most tissues (McClain, 1993). In general, initiation is produced by DNA-reactive carcinogens, whereas promotion is produced by non-genotoxic carcinogens.

4.6.2.1 *Genotoxic or DNA-reactive mechanisms*

Genetic changes induced by carcinogens are a fundamental part of carcinogenesis (Vaino et al., 1992) and for alkylating compounds arise from the reactivity of the carcinogen with DNA. DNA-reactive carcinogens usually operate as electrophilic reactants to bind to DNA (Williams, 1992). Carcinogens that act through such genotoxic mechanisms are usually multiorgan and trans-species carcinogens, can be active with a single dose and are effective at low exposures.

4.6.2.2 *Non-genotoxic mechanisms*

Non-genotoxic mechanisms of carcinogenesis do not involve a direct chemical attack on DNA, but rather are produced by other effects of the carcinogen on target cells or on the extracellular matrix (Williams, 1992). There are several non-genotoxic effects that can lead to enhancement of tumour development. Adaptive effects may lead to carcinogenicity with chronic, high-level exposure (Dybing et al., 2002; Williams & Iatropoulos, 2002). Thus, carcinogens that act through non-genotoxic mechanisms usually require high, sustained exposure. A common feature of the effects of non-genotoxic carcinogens is enhanced cell proliferation.

4.6.3 Chronic bioassays for the identification and characterization of cancer risk

Methods for the conduct of chronic cancer bioassays are well described (OECD, 1981a; Kitchin, 1999; Williams & Iatropoulos, 2001; VICH, 2002). For regulatory purposes, carcinogenicity bioassays usually consist of a 2-year rat study plus an 18-month mouse study, with 50 animals of each sex per group. Normally, there are at least three dose levels in addition to a concurrent control group; the highest dose should be associated with minimal toxicity as indicated by changes such as a slight decrease in weight gain, without affecting survival, to ensure that the bioassay provides suitable sensitivity for hazard identification purposes. For substances of low toxicity, the substance would normally be added to the diet at up to 5% by weight. Demonstration of a toxic effect in a cancer bioassay that does not compromise survivability or physiological homeostasis ensures that the animals were sufficiently challenged and provides confidence in the reliability of a negative outcome (VICH, 2002).

A positive response in either test species should be considered indicative of carcinogenic potential. With the development of alternative test systems (see section 4.6.4), carcinogenicity studies (e.g. for therapeutic drugs) are sometimes performed in one rodent species, preferably the rat, plus one or more alternative methods. Such an approach may become acceptable for WHO advisory committees in the future.

Extensive results using rats and mice are available (Gold & Zeiger, 1996), and such tests remain the standard. However, issues have arisen over the relevance to humans of an increase in certain types of neoplasms (section 4.6.6) and of mouse bioassays per se (Van Oosterhout et al., 1997).

4.6.3.1 *Statistical methods*

The statistical analysis of multidose cancer bioassays with potential treatment-related differences in survival is a complex and specialist issue. The methods provided by Peto et al. (1980) are widely accepted for statistical analysis, although other methods may be used.

4.6.3.2 Evaluation

Important criteria in the evaluation of positive findings are consistency and reproducibility. Results are more compelling if carcinogenic effects are seen in both rats and mice. In a single experiment, dose-related trends in specific tumour types, the nature and type of tumour, the occurrence of cancer in non-sex-related tissues in both sexes and the presence of related non-neoplastic findings (e.g. hyperplasia or toxicity) are important indicators of treatment-related neoplastic and preneoplastic effects.

4.6.3.3 Interpretation

The interpretation of bioassay results for human risk involves consideration of the relevance of the tumour type to humans and the dose–response in relation to the magnitude of human exposure. Further information is given in sections 4.6.6 and 4.6.7.

4.6.4 Alternative methods for carcinogenicity testing

A variety of alternative tests for carcinogenicity have been introduced in which tumorigenic responses are enhanced and the duration of bioassays is thereby reduced (McGregor et al., 1999; Cohen et al., 2001; Goodman, 2001). None of these have yet been applied to the same extent as the chronic bioassay.

4.6.4.1 Initiation/promotion models

Based upon distinct steps of initiation and promotion in carcinogenesis, models have been developed in which the substance is tested either as an initiator by administration before a promoter for the target organ of interest or as a promoter by administration after an initiator for the target organ (reviewed in Enzmann et al., 1998a,b). As these studies are generally less than 1 year in duration, the background of spontaneous neoplasms is negligible.

One of the major contributions of these models is that they provide information on the mode of action for observed effects. For example, McGregor et al. (1999) concluded that in such models, the

> appearance of tumours after administration of a test chemical as an initiator provides evidence of carcinogenic activity.... Additional evidence of promoting activity makes the evidence compelling. When data are available only on promoting activity, the evidence is suggestive of carcinogenicity..., but the information should be evaluated in conjunction with other data….

On the other hand, caution is needed in data interpretation, as these models assume that the added promoter or initiator is biologically relevant to the corresponding initiator and promoter under test.

4.6.4.2 Neonatal mouse model

In this model, newborn mice, usually of the CD-1 strain, are given the test substance by intragastric instillation on days 8 and 15 postpartum and observed for up to 1 year (Flammang et al., 1997; McClain et al., 2001). At the end of the study, the incidence of spontaneous neoplasms is negligible.

Data suggest that this model responds only to genotoxic carcinogens; as such, its utility for testing unknown substances is limited. In the International Life Sciences Institute (ILSI)–Health and Environmental Sciences Institute (HESI) Collaborative Program on Alternative Models for Carcinogenicity Assessment (ILSI, 2001), only 1 non-genotoxic chemical (17β-estradiol) of the 18 compounds that were evaluated was reported positive (McClain et al., 2001). Thus, a positive response in this model indicates that the test substance probably produced cancer via a genotoxic effect.

4.6.4.3 Transgenic mouse models

Through selective gene activation or deletion, mice of unique genotypes can be produced that may be more susceptible to carcinogenesis (Gulezian et al., 2000). These models have been widely applied in the testing of pharmaceuticals (ICH, 1997) and were evaluated in the ILSI-HESI Collaborative Program on Alternative Models for Carcinogenicity Assessment (ILSI, 2001). Usually the duration of bioassays is 26 weeks (rather than 2 years or 18 months for the rat and mouse, respectively) because of the increase in spontaneous tumours in transgenic animals beyond this time.

(a) p53$^{+/-}$ mice

This model employs mice in which one allele of the *TP53* tumour suppressor gene is disrupted (Donehower et al., 1992); hence, the model is believed to be responsive to genotoxic carcinogens (French et al., 2001). Initially, the inactivated null *Trp53* allele was implanted into C57BL/6 female mice, which produced, after numerous crossings, the C57BL/6-based model (Donehower et al., 1992; French et al., 2001). In a widely used version of this model based on the C57BL/6 mouse, the most common spontaneous neoplasm is subcutaneous sarcoma (Mahler et al., 1998), and increases have been provoked by implantation of devices (Mahler et al., 1998) or injection of irritant materials (Youssef et al., 2001). In addition, malignant lymphoma (both sexes) and osteosarcoma (males) are also known to occur spontaneously (French et al., 2001).

In the ILSI-HESI evaluation (ILSI, 2001), 6 of the 21 compounds tested were human carcinogens. In this model, four of these were positive (cyclophosphamide, melphalan, cyclosporin A and diethylstilbestrol), one was negative (phenacetin) and one was equivocal (17β-estradiol). Moreover, 12 of the 16 genotoxic human or rodent carcinogens were positive, and 2 (chloroform and diethylhexylphthalate) of the 22 non-genotoxic rodent carcinogens were judged equivocal (Storer et al., 2001).

(b) TG.AC model

Homozygous TG.AC mice were developed in the FBV/N strain by the introduction of a construct containing an activated v-Ha-*ras* oncogene (Leder et al., 1990). Either the homozygous TG.AC line or a heterozygous line derived by mating homozygous TG.AC males with FBV/N females can be used for chemical evaluation. Thus far, this model has been used largely for topical application in which the test substance is applied to the shaved dorsal skin (ILSI, 2001). Test substances have been administered in a variety of vehicles.

One issue with this model is the potential for chronic dermal irritation resulting from repeated shaving together with application of irritant vehicles (e.g. acetone) to enhance responses to test substances. This model is not an adequate replacement for a chronic mouse bioassay, as five of seven non-genotoxic mouse carcinogens were negative (Tennant et al., 2001).

(c) K6/ODC

Recently, K6/ODC mice have been evaluated as an alternative for short-term dermal carcinogenicity testing (Miller et al., 2008), as this strain develops epidermal tumours when exposed to genotoxic carcinogens. In a recent study, mice that received 7,12-dimethylbenz[*a*]-anthracene dermally developed papillomas as early as 6 weeks, but progressive adverse health and decreased survival suggested that K6/ODC mice may be an inappropriate alternative model.

(d) Xpa

Xpa$^{-/-}$ homozygous knockout mice have a defect in genes controlling the DNA repair pathway known as nucleotide excision repair. Xpa mice develop skin tumours at high frequency when exposed to ultraviolet light and are susceptible to genotoxic carcinogens given orally (Van Steeg et al., 2001). In an attempt to further increase both the sensitivity and specificity of the Xpa model in carcinogenicity testing, Xpa mice were crossed with p53$^{+/-}$ mice; the resulting Xpa/p53$^{+/-}$ double-knockout mice developed tumours earlier and with higher incidences upon exposure to carcinogens compared with their single-knockout counterparts. There appears to be a good correlation between compounds identified as positive in the Xpa/p53$^{+/-}$ model and human carcinogenicity (Van Steeg et al., 2001).

(e) Tg-rasH2

Unlike the p53$^{+/-}$ mouse, the Tg-rasH2 mouse is sensitive to both genotoxic and non-genotoxic carcinogens, but develops more spontaneous neoplasms compared with wild-type mice (Morton et al., 2002). In carcinogenicity testing, 4 of 6 known/suspected human carcinogens were positive; for 19 non-mutagenic agents testing positive in conventional rodent bioassays, 7 chemicals were positive, 10 chemicals were negative and 2 were equivocal. Results for 15 of 18 mutagenic chemicals agreed with the results of conventional rodent bioassays, and 3 results were equivocal. Thus, the Tg-rasH2 mouse model appears to predict known or suspected human carcinogens as well as the traditional mouse bioassay, but with fewer positive results for non-genotoxic compounds that are not considered human carcinogens (Morton et al., 2002).

(f) Other models

Several other transgenic models are available (Robinson & MacDonald, 2001) but are less widely used and lack adequate validation for regulatory purposes.

4.6.4.4 *Interpretation of the data from alternative methods*

McGregor et al. (1999) considered these alternative models appropriate for identifying carcinogens in rodents. However, the basis for a tumour increase can be obscure. For example, certain agents enhance the development of spontaneous neoplasms only; these could simply arise from a shortening of the latent period for these tumours, which appear in high incidence later.

In medium-term assays with preneoplasia as the end-point, McGregor et al. (1999) concluded that "the occurrence of preneoplasia ... within a period of 20–40 weeks provides evidence of potential carcinogenic activity".

More recently, IARC suggested that under certain circumstances, data from alternative assays could be used in safety evaluation in place of a second bioassay and that some of these models might be useful in hazard identification if used in conjunction with information from other sources in a weight of evidence, integrated analysis approach to risk assessment (Cohen et al., 2001).

4.6.5 *End-points in carcinogenicity studies*

4.6.5.1 *Spontaneous neoplasms*

The rodent strains used in chronic cancer bioassays have high incidences of certain tumour types (Williams & Iatropoulos, 2001) that may be irrelevant for human health, especially if increases are found only in such common neoplasms. Any increase may have arisen by enhancement of an endogenous spontaneous rodent mechanism, providing evidence of a cancer-promoting potential rather than a cancer-initiating potential. As such, the dose–response would be expected to exhibit a threshold.

4.6.5.2 *Pathological classification of neoplasms*

Standard criteria for the diagnosis of rodent neoplasms have been developed (Faccini et al., 1992). These are generally used in studies conducted for regulatory purposes, but not always in investigator-originated studies. The precision with which diagnostic criteria are applied is, of course, a function of the skill of the study pathologist. Guidance for the performance of the pathological evaluation is available (Williams & Iatropoulos, 2001).

For veterinary drugs, it has been recommended that in-life observations and pathological examination, consistent with OECD Test Guideline No. 451 (OECD, 1981a), are undertaken in carcinogenicity studies and that clinical pathology (haematology, urinalysis and clinical chemistry) is not considered necessary and does not contribute to the assessment of neoplastic end-points.

A valuable component of the pathological evaluation is peer review, in which a second pathologist examines a representative sampling of the material. Such peer review is particularly valuable when the pathologist is not informed as to which slides are from treated animals and which are from control animals (blind analysis).

4.6.5.3 *Benign and malignant neoplasms*

The distinction between benign and malignant neoplasms in experimental animals is usually made on the basis of histopathology; neoplasms classified as benign are usually not invasive or metastatic. There is controversy over whether an agent that induces only benign neoplasms should be classified as carcinogenic, and these data should therefore be used in an overall weight of evidence approach. Often a combination of histogenetically related benign and malignant neoplasms is used to arrive at a conclusion that the test substance is carcinogenic (Faccini et al., 1992; Williams & Iatropoulos, 2001).

4.6.5.4 *Preneoplastic lesions*

Preneoplastic lesions are part of the continuum of neoplastic development (Williams, 1999). Accordingly, their presence in a tissue at the end of a bioassay, together with related neoplasms, supports the conclusion of a chemical-induced carcinogenic effect. By themselves, however, they do not justify the conclusion that the substance is carcinogenic.

4.6.6 Characterization of carcinogenic effects

IARC has developed guidelines on the use of information on mechanisms in evaluating carcinogenicity findings of this type (Capen et al., 1999), which have been applied to assessment of human hazard of specific chemicals (McGregor et al., 1999).

IPCS developed a conceptual framework on the evaluation of an animal mode of action for chemical carcinogenesis. This framework provides a generic approach to the principles commonly used for evaluating mode of action. It outlines a list of elements to be considered in analysing whether available data support a particular mode of action (Sonich-Mullin et al., 2001).

Subsequently, this framework was extended to address the issue of human relevance of animal cancer data. The IPCS framework for analysing the relevance of a cancer mode of action for humans, along with three case-studies, was published in 2006 (Boobis et al., 2006). The application of this framework is intended to increase transparency in analysing and interpreting cancer data and will result in improved communication of the bases for scientific conclusions and decision-making.

4.6.6.1 Mechanisms relevant to humans

(a) DNA reactivity or genotoxicity

Carcinogens that are DNA reactive are usually trans-species carcinogens and therefore are presumed to be potential human carcinogens (McGregor et al., 1999); indeed, most human carcinogens are clearly DNA reactive (Thorgeirsson et al., 1994; Williams & Iatropoulos, 2001). Thus, assessment of genotoxicity is an important component of chemical evaluation and critical in the hazard characterization approach adopted (see chapter 7). Barlow et al. (2002) concluded that “specific markers of DNA damage or adducts will not only assist mechanistic understanding, but can assist in risk assessment". It should be noted that some forms of genotoxicity may exhibit a threshold—for example, aneugenicity as a consequence of spindle inhibition (Parry et al., 1994). In rare circumstances, toxicokinetic factors may be such that there is a de facto threshold for genotoxicity in vivo—for example, for phenol when exposure is via the oral route (EC, 2006).

Substances that produce cancer via modes of action that do not involve direct DNA reactivity and alkylation tend to show species differences in susceptibility and are often associated with cancer incidence at a single site. In addition, these non-genotoxic carcinogens usually show a biological threshold in their dose–response relationship. Normally, other effects that may be precursors are seen at doses below those that increase the incidence of cancer, and these effects are usually the focus of hazard characterization and derivation of a health-based guidance value.

4.6.6.2 *Mechanisms not relevant to humans*

(a) Surface and luminal tissue chronic irritation

It has long been known that wounding of surface and luminal tissues can elicit tumour development at the wound site. As blocking of cellular communication channels, an increase in the intensity of tissue metabolic reactions and even induction of sustained tissue ischaemia differ between laboratory animals and humans, their relevance to humans is limited.

(b) Mouse liver neoplasms

The relevance of the production of increases only in mouse liver neoplasms has long been questioned (Stevenson, 1990). No agent that produces increases only in mouse liver tumours is associated with comparable effects in humans (Williams, 1997).

(c) Hormonal disruption

Several hormone systems in rodents are more susceptible to disruption with consequent increase in neoplasia than the corresponding systems in humans. For example, thyroid tumours in rats can arise from thyroid–pituitary disruption, whereby reduced thyroid hormone levels lead to a negative feedback increase in thyroid-stimulating hormone levels and subsequent hyperplasia and neoplasia (Thomas & Williams, 1991; Hill et al., 1998; Rice et al., 1999) that are of negligible relevance to humans.

(d) Inhibition of tissue trophic activity

Interference with neuroendocrine immune feedback pathways can result in neoplasia that is species or sex specific and not relevant to humans (Iatropoulos & Williams, 1996; Williams & Iatropoulos, 2001).

(e) α2u-Microglobulin-induced rat nephropathy

Kidney tumours in male rats arising indirectly through binding to and increases in renal excretion of α2u-microglobulin are considered not relevant to humans, because humans do not synthesize α2u-microglobulin (USEPA, 1991d).

(f) Rat stomach neuroendocrine neoplasm

Neoplasia of gastric neuroendocrine cells is stimulated by gastrin in rats and to a lesser degree in mice, because rodents have a high density of neuroendocrine cells, giving high levels of gastrin (>1000 pg/ml). Because these high gastrin levels are not achieved in humans and other primates, this type of neoplasm is not relevant to humans (Tuch et al., 1992; Thake et al., 1995).

(g) Peroxisome proliferation

Rodent hepatic peroxisome proliferators cause tumours in rodent liver but do not produce these effects in primate or human liver (Williams & Perrone, 1996) as a result of species differences in levels of the peroxisome proliferator activated receptor of the class α (PPARα) (Tugwood & Elcombe, 1999) and other mechanistic differences between rodents and humans (Klaunig et al., 2003). Because of this, IARC (1995) has recommended that a tumour response in mice or rats secondary to peroxisome proliferation should modify the evaluation of carcinogenicity.

(h) Cytotoxicity and regenerative hyperplasia

Sustained, chemically induced cytotoxicity of various types can lead to regenerative hyperplasia and subsequent preneoplastic foci and tumours. However, the relevance of this to human exposure is questionable, as this mechanism is often a "high-dose" phenomenon that may be species specific.

4.6.7 Assessment of carcinogenic response

Carcinogenicity is a major concern in the risk assessment of chemicals in food, particularly if a genotoxic mechanism in known or suspected. In part, this is because risk management options for such substances can vary with jurisdiction. Hence, it is important that any possible carcinogenic effect be fully and consistently assessed. There are a number of issues that should be considered.

4.6.7.1 Nature of the test substance

The chemical purity of the substance and the possibility that impurities or co-formulants such as the vehicle (e.g. corn oil) might have influenced the response should be considered. The physicochemical form of the substance tested should be appropriate to the substance to which the population may be exposed. For example, the carcinogenicity of some metals (e.g. chromium) depends markedly on speciation. In the case of airborne particulates, the geometry and solubility of the particle will profoundly influence the response.

4.6.7.2 Relevance of study design

The route of exposure needs to be considered. Where irritant substances are administered at high local concentrations—for example, by oral gavage—they may produce tumours at the site of contact that are of limited or no relevance to humans under the exposure scenarios of concern. Some routes of exposure—for example, intraperitoneal—are not relevant to human exposure. These need to be considered on a case-by-case basis. In some instances, the avoidance of presystemic metabolism may lead to quantitatively, or even qualitatively, erroneous conclusions.

Duration of exposure should also be considered. Where study duration is less than that recommended by the relevant test guidelines, the likelihood that carcinogenic effects would have been missed needs to be assessed. This also applies to situations where survival at the end of a study is less than the minimum recommended. In some instances, it may still be possible to obtain meaningful conclusions from the study—for example, where survival is still high until a couple of months before the normal end of the study.

4.6.7.3 *Are the tumours substance related?*

As discussed above, the possibility that tumours are a consequence of the vehicle used or the method of administration—for example, physical irritation by the gavage needle—should be considered, particularly where the response is specific to a particular set of experimental conditions and is negative in other studies with different experimental conditions (e.g. when using another vehicle). The statistical significance of the tumour response should be considered, together with historical control data. For example, was the tumour incidence in the control group lower or higher than the extremes in the historical control data?

The nature of the dose–response relationship can be of value in interpreting the data. For example, where a statistically significant response is observed only at the lowest dose and no response is seen in any of the higher dose groups, the plausibility of a substance-related response needs to be considered carefully. The lesion in question should be a malignant tumour, although, on occasion, benign tumours may be informative in assessing carcinogenicity, as discussed above. However, the relationship between preneoplastic and neoplastic effects needs to be considered; where there is no substance-related malignancy, the relevance of preneoplastic findings alone needs to be addressed.

Food intake can influence longevity and tumour incidence as a consequence of nutritional status or altered lifespan. Hence, substance-related effects and other factors influencing food consumption may indirectly affect tumour incidence, and due consideration should be given to this possibility when there are appreciable changes in either food consumption or lifespan (increased) in a study.

4.6.7.4 *Can a mode of action for the tumour response be established?*

A mode of action has been defined as a series of key events leading to an observed effect supported by robust experimental observations and mechanistic data (Boobis et al., 2006). Examples of key events include specific metabolic transformation, receptor–ligand changes, increased cell growth and organ weight, and hormonal or other physiological perturbations. Identification of the mode of action for a carcinogenic response in experimental animals can be of considerable value in addressing issues such as human relevance, dose–response and CSAFs. Identification of a mode of action is based on a weight of evidence approach that has been described in detail in publications from IPCS (Sonich-Mullin et al., 2001; Boobis et al., 2006). Whereas formal mode of action analysis may not be necessary for every carcinogenic response, some consideration of mode of action will be necessary in all cases, if only to determine whether the response is likely to exhibit a threshold or not (see section 4.6.2).

4.6.7.5 *Is the mode of action relevant to humans?*

IPCS has published an analytical framework for assessing whether the mode of action for a tumour response observed in an experimental study is relevant to humans (Boobis et al., 2006). A number of modes of action are not relevant on the basis of qualitative or quantitative considerations (see section 4.6.6.2). Application of the framework will not be necessary in all cases—for example, where a compound is clearly a direct-acting DNA-reactive genotoxic carcinogen. However, in other cases, the framework can be invaluable in determining the strength of evidence of a conclusion regarding human relevance, in a transparent and consistent manner. Hence, in cases where there is possible ambiguity as to the conclusion regarding human relevance, it is recommended that the framework be applied and the results presented in the report of the assessment. Even where human relevance cannot be excluded, application of the framework can provide insight into species differences, dose–response relationships and potential susceptible subpopulations—for example, on the basis of life stage.

4.6.7.6 *Historical control data*

The incidence of spontaneous tumours can vary, sometimes appreciably, among control groups of the same species and strain in different studies, even when conducted within the same laboratory under carefully controlled conditions. Hence, for a response to be considered substance related, not only should it differ significantly from that in the control group, but in general it should also differ from the background incidence in that species and strain of experimental animal. Hence, suitable data on historical controls should be available to help in interpretation of the findings. Although historical control data can be of considerable value in data interpretation, they should not be viewed as a substitute for concurrent control data. An overall weight of evidence approach is necessary.

Ideally, historical control data will have been obtained in the same species and strain, from the same supplier, and maintained under the same conditions in the same laboratory as that generating the study data being evaluated. The data should be from control animals over a 5-year period, centred as closely as possible on the date of the study being evaluated. The historical control data should be presented for each discrete group, indicating sex and age of the animals. In addition, information on the following should be provided:

- species, strain, name of the supplier and specific colony identification if the supplier is based in more than one location;
- name of the laboratory and date on which the study was performed;
- description of general conditions under which the animals were maintained, including details of diet and, where possible, the amount consumed;
- the approximate age, in days, of the animals at the beginning of the study and at the time of death;
- details of the mortality pattern observed during or at the end of the study and of any other relevant observations (e.g. infections);
- identity of the pathology laboratory and the pathologist responsible for analysing the pathology data from the study; and
- which tumours were combined, if any, in generating the incidence data.

In evaluating historical control data, the following points should be considered:

- If the tumour incidence in the concurrent control group is lower than that in the historical control groups but is within the historical control range in the treated groups, it would be concluded that there is no biologically relevant substance-related response.

- If the tumour incidence in the treated groups is above the historical control range but not statistically significantly different from that of the concurrent controls, it would be concluded that there is no substance-related response (although it is always possible that this was a false negative).

- Where the tumour incidence in the treated groups is significantly greater than that in the concurrent controls and is above the historical control range, it would be concluded that the carcinogenic effect is likely to be substance related, with a low probability of a false positive.

4.7 Reproductive and developmental toxicity

4.7.1 Introduction

Adverse effects on reproduction may be expressed through reduced fertility or fecundity in either the parents or offspring as a result of morphological, biochemical, genetic or physiological disturbances. Adverse effects on development may be expressed through altered viability, growth or structural or functional abnormalities due to either mutations or biochemical/physiological disturbances. Adverse effects on development induced by chemicals may be expressed immediately or they may be delayed, sometimes for many years, as exemplified by transplacental carcinogens.

Typical developmental toxicity studies investigate the effects of exposure to test substances starting at implantation and continuing through the period of organogenesis. More recent study protocols extend the period of exposure to include the fetal period. Effects due to chemical exposure during the fetal period, the developmental period after the major organ systems have formed, generally involve growth retardation and functional disorders, although the external genitalia and the central nervous system are also susceptible to injury during this period. These studies were previously called “teratogenicity studies” but are now called “prenatal toxicity” or “developmental toxicity” studies in recognition that they cover more than just structural malformations. Subtle structural or functional abnormalities often do not become obvious until some time after birth and in some cases not until adulthood.

Because of the differential rates of development between species and the relative states of maturity of neonates at birth, it is important to understand equivalencies of developmental stages when comparing exposure scenarios across species (i.e. what is the equivalent human stage for a particular window of exposure in a rodent?). Comparative rates of development, as well as spontaneous rates of malformations for a number of species and strains, are provided by Schardein (2000). The developmental processes at risk and their critical stages of vulnerabilities during prenatal and postnatal life have been reviewed by IPCS (2006b).

Neonatal development may be influenced by chemicals (or their metabolites) that are present in the maternal diet and subsequently transferred into maternal milk. Chemical exposure of the mother may also affect neonatal development by influencing maternal behaviour, hormonal balance or nutrition. Direct neonatal exposure to xenobiotic compounds can also occur via consumption of infant formula. Examples include the limited number of additives that are used in infant formula, phytoestrogens in soy-based formula and migrants from infant feeding bottles.

Guidelines for reproductive and developmental toxicity tests have been developed by various legislative and international organizations, including the OECD (see http://www.oecd.org/department/0,2688,en_2649_34377_1_1_1_1_1,00.html), the ICH (1994c), the USEPA (1991b, 1996; see also http://www.epa.gov/opptsfrs/home/guidelin.htm) and IPCS (2001b). A guideline for developmental neurotoxicity has also been developed by OECD, in which postnatal function and behaviour can be investigated in offspring exposed to chemicals during the prenatal and in the early postnatal period (OECD, 2007). Such studies are discussed in section 4.8.3.3 and will not be further addressed here.

4.7.2 End-points of concern

The range of reproductive functions that are observed in reproductive toxicity studies includes gametogenesis, mating, fertility, maintenance and duration of pregnancy, parturition, litter numbers, lactation, puberty, viability and growth of offspring and reproductive senescence. These aspects can be investigated in the parental and filial generations through end-points such as the following:

- Parents and offspring:
  - Sperm measures (number, motility, morphology, sperm production rate)

- Vaginal cytology (estrous cycles)
- Hormone measurements
- Evidence of mating
- Pregnancy rate
- Organ weights (gonads, uterus, epididymis and accessory sex glands)
- Histopathology of the reproductive tissues
- Reproductive behaviour

- Offspring:
  - Litter size and viability
  - Body weight
  - Sex ratio
  - Anogenital distance
  - Nipple/areola retention in males
  - Vaginal opening
  - Testes descent
  - Preputial separation

For all the outcomes and end-points, it is necessary to determine the normal range and the extent of deviation that should be considered adverse.

The range of adverse effects on offspring arising from maternal exposure to chemicals during pregnancy includes death and resorption of the embryo or fetus, teratogenic defects (structural malformations), growth retardation or specific developmental delays, and decreased postnatal functional capabilities.

For a developmental toxicant, the effects that will be expressed depend on the level and gestational timing of the dose of the chemical and the duration of the treatment period. Thus, a substance given at one dose level may result in growth retardation, whereas at a higher level it may result in death and resorption of the embryo. Sometimes the slope of the dose–response curve for these effects is very steep. The concept of critical period is important to recognize, as an exposure at one developmental stage could be without effect, whereas the effect could be severe at another developmental stage because the target tissue is at an exceptionally vulnerable point as a result of the progression of developmental events that are occurring. Similarly, an exposure at one point in development may induce growth retardation, whereas malformations could be observed during a different exposure window. In addition, because of differences in the rates of development and toxicokinetics, it is not expected that a particular experimental outcome will translate with fidelity across species. Thus, an agent that induces, for example, limb malformations in a mouse would not necessarily yield that same result in humans (but for human risk assessment purposes, it would generally be assumed to have the potential to produce some manifestation of developmental toxicity). Because all of these outcomes are adverse, the most important consideration when evaluating these studies should not be what effect is observed, but rather at what dose level the adverse effect became evident (USEPA, 1991b) and whether there was also any evidence of maternal toxicity.

4.7.3 Study design

4.7.3.1 Overview

A number of reviews of procedures and methodologies for assessing the effects of chemicals on reproductive function are available (USEPA, 1996, 1998b,c, 2002; IPCS, 2001b). The procedures described in these publications are designed to assess the potential for reproductive and developmental toxicity of test substances using lower mammals as model systems. It is important to take into account the existing toxicological database on the chemical to make sure that appropriate end-points are being adequately covered. The knowledge can be used for more individualized study designs that go beyond the minimum core guideline requirements in order to better understand the full potential of the chemical to affect reproductive function and development.

Regardless of the actual experimental design, the goal of reproductive and developmental toxicity protocols is to assess the sensitivity of various processes and life stages to alterations brought about by exposure to the substance under study and to characterize the most vulnerable target tissue. Therefore, the highest dose of a food chemical that is administered is generally the amount that would be expected to cause slight systemic toxicity, with lower doses being geometrically spaced to a level not expected to induce significant adverse effects. If there is a significant reduction in maternal body weight or other indication of excessive maternal toxicity, caution should be applied in interpreting any adverse outcomes in the offspring, as the effects could be secondary to maternal toxicity. It is important that appropriate sensitive end-points be evaluated, that exposures cover all of the known critical periods and that sufficient sample sizes be used in order to ensure adequate statistical power to detect effects when present. Thus, in the case of developmental toxicity studies, where either half or all (depending on the particular protocol) of the fetuses are examined for soft tissue and skeletal morphology, it has been estimated (USEPA, 1991b) that the minimum change detectable is an increased incidence of malformations of 5- to 12-fold over control levels and a 3- to 6-fold increase in embryonic or fetal death. This contrasts with the ability to detect a 0.15- to 0.25-fold reduction in fetal weight, which is a continuous variable. As a number of chemicals have now been identified as endocrine disruptors that can cause malformations of the reproductive tract that would not be readily observable in the fetal examinations conducted in developmental toxicity tests (e.g. hypospadias), it is likely that in reproductive toxicity tests, the numbers of offspring evaluated in filial (F_1, F_2, etc.) generations (where subsequent postnatal development allows the malformations to be expressed and readily observed) will need to be increased.

4.7.3.2 *Reproductive toxicity*

Generally, effects on reproduction are evaluated in multigeneration studies such as OECD Test Guideline No. 416: Two-Generation Reproduction Toxicity Study (OECD, 2001b), the USEPA's Reproduction and Fertility Effects test guideline (USEPA, 1998b) and the Reproductive Assessment by Continuous Breeding protocol of the United States NTP (Chapin & Sloane, 1997). Rats are the usual species of choice for multigeneration-type studies, and generally only one species is tested because of the length, cost and complexity of such studies.

For hazard identification, several other protocols exist that evaluate various aspects of reproduction and development, such as OECD Test Guideline No. 415: One-Generation Reproduction Toxicity Study (OECD, 1983), OECD Test Guideline No. 421: Reproduction/

Developmental Toxicity Screening Test (OECD, 1995d), OECD Test Guideline No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test (OECD, 1996) or the NTP 35-day screening protocol (Harris et al., 1992). One-generation studies usually evaluate the effects of subchronic exposure of adult animals in the parental generation and the F_1 generation through to weaning, whereas in multigeneration studies, exposure of the F_1 generation continues through weaning to adulthood, at which point they are mated to produce the F_2 generation. Because the parental and subsequent filial generations have different exposure histories, different outcomes may be observed. In particular, effects may be observed in the F_1 and F_2 generations that are not apparent in the parental generation because of their exposure during the full period of development. More recently, with the concerns raised for chemicals that could interact with the endocrine system and thus disrupt a number of processes critical for successful development and reproduction, a series of screening assays have been proposed that evaluate specific aspects of physiology related to estrogen, androgen and thyroid hormone action (see section 4.7.3.5).

It should be borne in mind that some end-points in reproductive toxicity studies are also inherently insensitive to chemical exposure (USEPA, 1996). For example, because of a large reserve capacity in sperm numbers, daily sperm production can be drastically reduced in the adult male rat without any apparent effect on fertility. This is in contrast to the situation in humans, where relatively small decrements in sperm production would be expected to elevate the probability of infertility or subfertility. To address this discrepancy and to add more sensitive end-points, recent revisions to test guidelines (e.g. USEPA, 1998b; OECD, 2001b) include guidance for the assessment of testicular function (e.g. daily sperm production and epididymal sperm counts, sperm motility and sperm morphology). Similarly, to be more sensitive to endocrine-active agents, some designs include determination of the age at vaginal opening in the female and preputial separation in the male as indices of puberty and options for measurement of anogenital distance, an androgen-dependent, sexually dimorphic trait, in the neonate and nipple retention in male offspring.

Single-generation and multigeneration reproduction studies are particularly useful for assessing potentially deleterious effects on reproduction and development through birth to weaning. Although the basic protocols have been in existence for at least 30 years, new end-points have been added to them over time in order to increase the breadth of the end-points covered, as well as the sensitivity of the end-points to perturbations (Kimmel & Makris, 2001). There is also discussion about the sample sizes used to evaluate the offspring in multigeneration studies for malformations. Existing guidelines generally require one male and one female from each of the litters to be evaluated for malformations. Such small sample sizes require that a very high incidence of an effect be present before it would be confirmed statistically (see discussion of statistical power in section 4.7.3.1).

Conversely, other components of earlier multigeneration test protocols have been dropped over time, most notably the need to rear two litters per generation (nowadays, only one is recommended) and the need to use three generations (nowadays, only one or two is recommended). The general consensus now is that these additional components did not provide qualitatively new information.

4.7.3.3 Developmental toxicity

Effects on prenatal development are examined using protocols such as OECD Test Guideline No. 414: Prenatal Developmental Toxicity Study (OECD, 2001a) and the USEPA's Prenatal Toxicity Study (USEPA, 1998c), which expose pregnant animals during the period of major organ formation and examine fetuses for growth and structural development. Generally, developmental toxicity tests are conducted in two species, usually a rodent and a non-rodent, as greater confidence is gained when results are available from more than one species. This is especially true in instances where the lack of developmental toxicity is noted in the first species tested. However, in situations where the first study shows evidence of developmental toxicity, it may be possible to complete the assessment with adequate confidence (see section 4.7.3.4). The species of choice for routine studies are usually rat and rabbit, but in cases where the rabbit is unsuitable (see section 4.7.4), the mouse is often used.

The basic protocol for the evaluation of developmental toxicity has been largely unchanged for more than 25 years, although later modifications have increased their scope and sensitivity (Kimmel &

Makris, 2001). One change has involved the extension of the dosing period from just covering the period from implantation through to closure of the palate (known as "organogenesis" and corresponding to days 6–15 of pregnancy in the rat) to include the late gestation period to the day before sacrifice. This allows better coverage of late-developing organ systems, such as the reproductive tract and the central nervous system. There are still recognized limitations in detecting alterations in some systems using the standard fetal examination process that focuses on morphology and examines tissues that are not fully mature (and hence may not yet express the developmental effect), such as the central nervous system (Rodier et al., 1994; Harry, 1998), the immune system (Holladay & Luster, 1994) and the heart, lungs and kidneys (Lau & Kavlock, 1994). These limitations can be addressed, at least partially, in the newer multigeneration and developmental neurotoxicity study protocols (e.g. OECD, 2007), which include assessments of animals after birth. Another significant change to developmental toxicity protocols has been to increase the numbers of non-rodents per dose group from 12 to 20 animals. This change was made in recognition of the fact that studies in non-rodents were statistically underpowered relative to those in rodents, which themselves still have limitations in terms of detecting rare events. A final modification relates to the examination of cartilage in addition to bone, as this can provide information for judging whether a skeletal alteration represents a variation or a true structural malformation.

As in reproductive toxicity studies, rats are commonly used in developmental toxicity studies, but experience has indicated that the use of a second species (generally a non-rodent like the rabbit) affords greater confidence in identifying agents that are likely to be hazardous to humans because of the recognized variability among species in response to developmental toxicants. Additional information on the use of rabbits in reproductive and developmental toxicity studies has been summarized by Foote & Carney (2000).

Regardless of the approach taken, evaluation of developmental toxicity data is facilitated by the use of common terminology. Glossaries of common developmental abnormalities (Wise et al., 1997) and skeletal anomalies (Solecki et al., 2001), as well as accompanying images, are available on the Internet at http://www.devtox.org/.

4.7.3.4 *Tiered and combined approaches to reproductive and developmental toxicity testing*

A proposal has been developed recently, in the context of pesticide safety assessment, for a tiered approach to toxicity testing at different life stages (Cooper et al., 2006). The aim of the approach is to assess the potential of a chemical to cause adverse effects on reproduction and assess the nature and severity of any effects on development and adolescence. It proposes, for Tier 1, an F_1-extended one-generation reproduction study in the rat and a prenatal developmental toxicity study in the rabbit. Pharmacokinetic studies are rarely performed routinely in pregnant or young animals, but such information is helpful in better understanding dose–response relationships and in placing the results in context with potential human exposure situations. This proposed approach emphasizes the value of using kinetic data in the design and interpretation of life stage studies. A draft protocol for an extended one-generation reproduction study is currently under development by OECD.

The International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products (VICH) also recommends a tiered approach to testing for the safety assessment of veterinary drug residues in human foods. In the first instance, a two-generation reproduction study in the rat and a developmental toxicity study in the rat should be conducted. If clear evidence of teratogenicity is observed, regardless of maternal toxicity, testing for developmental toxicity in a second species would not be required, unless teratogenicity in the rat was the critical effect for the setting of the ADI. If a negative or an equivocal result for teratogenicity is observed in the rat, a developmental test in a second species, preferably the rabbit, should be conducted. In the absence of teratogenicity in the rat, a developmental toxicity test in a second species would be required even if there were other signs of developmental toxicity in the rat (i.e. fetotoxicity or embryolethality). The VICH guidelines are available at http://www.vichsec.org/en/guidelines2.htm.

4.7.3.5 *Endocrine toxicity*

The state of the science in the area of endocrine toxicity was extensively reviewed by IPCS (Damstra et al., 2002). It is now recognized that the well-established tests for reproductive and developmental toxicity described above do not necessarily cover the full range of effects that might be induced by chemicals that interfere with the endocrine system. Moreover, these tests are resource intensive and not suited to the initial screening of large numbers of chemicals for endocrine toxicity. Spurred on by the concerns raised during the last decade about chemicals acting as endocrine disruptors and by legislative mandates such as the Food Quality Protection Act of 1996 in the USA, considerable effort has been directed at developing a battery of assays that can evaluate chemicals that interact with the estrogen, androgen and thyroid signalling pathways.

A tiered screening battery was proposed by the United States Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC, 1998) and is in the process of being validated through international cooperation between the USEPA and OECD. Tier 1 of the battery includes in vitro tests of receptor binding and gene activation for estrogens and androgens, a uterotrophic assay to identify estrogens, a Hershberger assay to identify androgens/anti-androgens, a female pubertal assay to evaluate neuroendocrine (estrogenic and thyroid) control of puberty, a frog metamorphosis test to evaluate thyroid effects and a short-term fish reproduction test to evaluate alterations in steroid hormone homeostasis in a lower vertebrate (Gray et al., 2002). As the Tier 1 screening tests are directed at detecting modes of action and not necessarily adverse effects, they serve primarily to trigger other tests (e.g. multigeneration tests) that could confirm a hazard and establish dose–response relationships. Because they can provide insight into potential modes of action, these screening assays should be highly informative at directing attention to specific outcomes in any follow-up dose–response studies, which could be customized to detect the more sensitive end-points. However, it should be noted that for many of the food chemicals that are evaluated by JECFA and JMPR, a reproductive toxicity test is conducted routinely, irrespective of whether the chemical is suspected to be an endocrine disrupter.

It is clear that the methodology for investigating endocrine toxicity is still evolving, and there are currently no generally accepted core requirements beyond the standard developmental and reproductive testing guidelines. The current status of the validation and use of the EDSTAC screening battery (EDSTAC, 1998) by the USEPA can be found at http://www.epa.gov/scipoly/oscpendo/index.htm. The current status of method validation by the OECD through its programme on Endocrine Disrupter Testing and Assessment can be found at http://www.oecd.org/document/62/0,3343,en_2649_34377_2348606_1_1_1_1,00.html.

4.7.4 Issues specific to category of chemical

There are relatively few examples in reproductive or developmental toxicity where a species is inappropriate for evaluation of a particular class of chemicals. One such example is chemicals that interfere with prolactin, which is essential for the maintenance of early pregnancy in the rat but not in humans. Another example, relevant to the work of JECFA on veterinary drug residues, is oral administration of certain Gram-negative antibiotics in rabbits. The intestinal flora of rabbits is particularly sensitive to this type of antibiotic, and treated dams can develop diarrhoea with reductions in food consumption and body weight, resulting in abortions, resorptions, malformations and fetal growth retardation (reviewed in Chernoff et al., 1989).

Schardein (2000) discussed the appropriateness of various animal models for assessing human risk. As with any toxicity test, it would be most appropriate to utilize a species that metabolizes a chemical in a manner similar to that of humans. However, in practice, such information is usually not available. Another consideration is whether the type of placentation in a particular species influences the degree or nature of the outcome in the fetus. For example, trypan blue is a developmental toxicant in rodents because of its effects on the yolk sac placenta, which is critical for the nutrition of the embryo in rodents. Such effects do not occur in other species in which, like humans, the embryo does not rely on the yolk sac for nutrition.

4.7.5 Interpretation of data

There are a number of publications, mostly developed by regulatory agencies or other bodies, that provide excellent information on the evaluation of reproductive and developmental toxicity data (e.g. USEPA, 1991b, 1996; IPCS, 2001b; Hood, 2006). In addition, the Center for the Evaluation of Risks to Human Reproduction (CERHR), established by the United States National Institute of Environmental Health Sciences, convenes expert panel meetings dealing with chemicals, chemical classes or generic issues related to the evaluation of data. The basis for the CERHR evaluative process can be found at http://cerhr.niehs.nih.gov/aboutCERHR/index.html#evalprocess.

In interpreting data from both reproductive and developmental toxicity studies, it is important to look for biologically related patterns of response and the relationship of outcomes across end-points and to relate any findings to the larger body of toxicological data available from other bioassays. Outcomes from other toxicity studies can be useful in targeting those end-points in developmental or reproductive toxicity tests that might be expected to be responsive to the agent, as well as assisting in determining potential modes of action. The incidence and severity of the findings should be noted, with comments on the extent to which the effects might be expected to be reversible upon cessation of exposure. Attention should be paid to which life stage is the most sensitive to exposure, although initial studies may not pinpoint the origin of the adverse effect because of the possibility of delay in its appearance.

In developmental toxicity studies, a *malformation* is usually defined as a permanent anatomical structural change that may adversely affect survival, development or function. The term *variation* is used to indicate an alteration in anatomical structure that generally does not adversely affect survival or health. When interpreting the significance of some structural variants, it is important to consider the stage of the fetus at the time of observation. Under most regulatory guidelines, fetuses are removed from the mother 12–14 h prior to the anticipated time of birth, a period of very rapid growth. Even slight perturbations in the growth trajectory can lead to changes in the rate of ossification and increases in the number of variants recorded. Double-staining the skeleton for bone with alizarin R and for cartilage with alcian blue can help distinguish whether bone development is merely delayed or whether there is an underlying morphological alteration. However, distinguishing between variations and malformations is difficult, as there is a continuum of responses from the normal to the extremely abnormal. There is no generally accepted classification of malformations and variations. Other terms that are often used, but no better defined, include anomalies, abnormalities, birth defects, deformations and aberrations.

Appropriate historical control data can sometimes be very useful in the interpretation of data on the incidence of malformations and variations. Comparison of data from treated animals with data from concurrent study controls should always take precedence over comparison with historical control data. The most appropriate historical control data are those from the same laboratory in which studies were conducted. Even data from the same laboratory, however, should be used cautiously and examined for subtle changes over time that may result from genetic alterations in the strain or stock of the species used, changes in environmental conditions, both in the breeding colony of the supplier and in the laboratory, and changes in personnel conducting studies and collecting data. Study data should be compared with recent as well as cumulative historical data. Although a dose-related increase in malformations is readily interpreted as an adverse developmental effect of exposure to a chemical, the biological significance of an altered incidence of anatomical variations is more difficult to assess and must take into account what is known about developmental stage (e.g. with skeletal ossification), background incidence of certain variations (e.g. 12 or 13 pairs of ribs in rabbits) or other strain-specific or species-specific factors. However, if variations are significantly increased in a dose-related manner, these should also be evaluated as a possible indication of developmental toxicity (USEPA, 1991b).

Because standard study designs require that the top dose exert some minimal indication of maternal toxicity (e.g. a 10% reduction in maternal body weight gain during pregnancy), there is sometimes difficulty in distinguishing whether a developmental effect seen at such a dose is a direct result of the action of the chemical on the embryo or fetus or an indirect result of altered maternal homeostasis. Although there have been several examples of the latter, it is important not to infer causation from an association of developmental toxicity with maternal toxicity without additional analysis and experimentation. Some aspects that should be considered include the following: Is the nature of the developmental manifestation a rare or common event in control offspring? What is the statistical power to detect a maternal versus a developmental event? Does the incidence or intensity of the effect tend to correlate with the intensity of the corresponding maternal response? Does the response occur in common across a number of members of a chemical class? Chernoff et al. (1989), Daston (1994) and Schardein (2000) have discussed various aspects of this issue. For example, significant impairment of maternal renal function by mercury(II) chloride in the rat has relatively minimal effect on rat embryonic development

(Kavlock et al., 1993), whereas the induction of maternal nutritional deficiencies (e.g. zinc deficiency following metallothionein induction) has been causally related to altered pregnancy outcomes (Keen et al., 2003). In any event, maternal and developmental toxicity should not be causally linked merely because of their concurrent appearance on the dose–response curve. However, the larger the spacing between the dose causing a maternal effect and a lower dose causing a developmental effect, the more likely a chemical will pose a developmental hazard to humans, as there would be no warning from maternal toxicity of the impending developmental effect. It is also important to note that some human developmental toxicants, such as lead, methylmercury and alcohol, exert effects on the embryo and fetus at doses that induce maternal toxicity, but the adverse effects are not secondary to the maternal toxicity, and thus the expected exposure conditions for humans are also an important consideration in interpreting such data.

4.7.6 Other considerations

4.7.6.1 In vitro tests

A number of assays have been proposed for use in screening chemicals for developmental toxicity. These include the use of lower organisms (e.g. *Drosophila* or *Xenopus* embryos), cell lines (e.g. human epithelial mesenchymal cells, mouse ovarian tumour cells, chick embryo neural retinal cells and various embryonic stem cell lines), primary cell cultures (e.g. neuronal and limb bud cells), avian embryos in ovo and mammalian embryos in culture. None of these tests has yet achieved international acceptance for use in hazard assessment, but they have proven valuable in some situations for understanding structure–activity relationships within chemical classes, as well as potential modes of action for toxicity.

4.7.6.2 Paternally mediated effects

Paternally mediated effects are those that are expressed in the offspring via exposure of the male prior to mating. A workshop (Robaire & Hales, 2003) reviewed evidence showing that such effects can occur with certain types of chemical. Most of the emphasis on paternally mediated effects has traditionally been in relation to infertility (e.g. dominant lethal effects), as opposed to evaluations of abnormal pregnancy outcomes (e.g. structural malformations or transplacental carcinogenesis).

In general, chemicals that have been associated with the induction of paternally mediated effects are DNA reactive and exert effects through DNA damage to the sperm. As a consequence, a number of new tests have been developed to serve as biomarkers of genetic and chromosomal integrity of sperm (e.g. chromosome-specific fluorescence in situ hybridization probes, the sperm chromatin structure assay and the comet assay). Because these biomarker tests tend to be technically difficult to perform, they have not received widespread use. For risk assessment purposes, it is important to understand the exposure paradigm in relation to the spermatogenic cycle, the nature of the end-points evaluated and the characterization of any dose–response relationships.

4.7.7 Information gaps

There are also several gaps in current approaches for the assessment of reproductive toxicity, including 1) the lack of longitudinal studies that assess exposed individuals through to senescence, 2) little evaluation of reproductive senescence in particular, 3) very limited evaluations of endocrine function, 4) little or no information regarding pharmacokinetics (this includes age-related studies, sex studies and target organ dosimetry) and 5) no use of acute or chronic exposures for the evaluation of reproductive effects or consideration of latent effects.

Likewise, there are gaps in the testing protocols for assessment of developmental toxicity. These include 1) the limited exposure of the neonatal animal, 2) the general limitation that the studies focus primarily on morphological changes and do not evaluate functional alterations in important systems such as the immune, cardiovascular, respiratory and renal systems, 3) the lack of pharmacokinetic information and 4) the paucity of information related to identification of latent manifestations of toxicity.

4.8 Neurotoxicity

4.8.1 Introduction

Neurotoxicity has been defined as an adverse change in the structure or function of the central nervous system and/or peripheral nervous system following exposure to a chemical (natural or synthetic) or physical agent (Tilson, 1990b; ECETOC, 1992; Ladefoged et al., 1995). The Nordic Council of Ministers defined neurotoxicity as the capability of a chemical to induce adverse effects in the central nervous system, peripheral nervous system or sense organs and cause a consistent pattern of neural dysfunction or lesion (Johnsen et al., 1992). The crucial term within these definitions is “adverse”. Exactly what defines an effect as adverse remains a major point of debate. In a toxicological sense, “adverse” can indicate a detrimental change in structure or function of the nervous system. A commonly accepted definition of adversity is an exposure-related alteration from baseline functioning that diminishes an organism’s ability to survive, reproduce or adapt to its environment (ECETOC, 1992; Ladefoged et al., 1995; USEPA, 1998a; IPCS, 2001a). IPCS has also defined an adverse effect as a change in morphology, physiology, growth, development or lifespan of an organism that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other environmental influences (IPCS, 2004).

Neurotoxic effects include a spectrum of biochemical, morphological, behavioural and physiological abnormalities whose onset can vary from immediate to delayed following exposure to a toxic substance and whose duration may be transient or persistent. These effects may be due to a direct action of the substance or metabolites on the nervous system or an indirect action on other biological systems that in turn adversely affect the nervous system (ECETOC, 1992, 1998; O’Donoghue, 1994; Ladefoged et al., 1995; USEPA, 1998a; USFDA, 2000).

4.8.2 Nervous system features

The basic structure and function of the nervous system, as they relate to neurotoxicity, have been comprehensively presented in EHC 60 (IPCS, 1986b) and EHC 223 (IPCS, 2001a). Additional descriptions are available in USEPA testing and risk assessment guidelines (USEPA, 1991a,c), in the IPCS-sponsored workshop efforts on in vitro techniques for neurotoxicity (Harry, 1998) and in other reports (United States Congress, Office of Technology Assessment, 1990; USNRC, 1992; SGOMSEC, 1996).

4.8.3 Evaluation of neurotoxicity

Conventional toxicity studies do allow some evaluation of neurotoxicity; however, these studies provide little information concerning less severe, but important, types of neurotoxic effects, including behavioural and physiological dysfunction and developmental neurotoxicity. Historically, neurotoxicity was equated with structural changes involving frank neuropathological lesions or overt neurological dysfunctions, such as seizure, paralysis or tremor. However, a significant body of scientific literature has demonstrated a variety of functional and structural abnormalities associated with chemically induced changes at the cellular and molecular level that may occur in the absence of evident structural changes identified using routine neuropathological techniques. Thus, reliance on routine neuropathology does not adequately reflect contemporary concerns about the broader spectrum of potential neurotoxic effects on the organism.

Methods to assess morphological, physiological, biochemical, behavioural and interactive components of nervous system functioning have been included in specific testing guidelines. Current guidelines for neurotoxicity studies have been developed by various national and international bodies, including assessments of general toxicity, gross histopathology and evaluations of behavioural functions (USEPA, 1991a,c, 1998a; ICME, 1994; OECD, 1995b,c, 1997; USFDA, 2000). Guidelines for developmental neurotoxicity studies recommend dosing during defined periods of gestation and lactation and the assessment of postnatal physical and behavioural development, including learning and memory, and neuroanatomical alterations, as appropriate (USEPA, 1991b; USFDA, 2000; OECD, 2007).

4.8.3.1 *Morphological evaluations*

The complexity and integrative nature of the nervous system make reliance on a single end-point problematic. The presence of a gross histopathological lesion in the brain would clearly identify a compound as being neurotoxic; however, discrete lesions are not always detected, even with known neurotoxicants. Any requirement that histopathological or morphological changes must be present as evidence of neurotoxicity is inappropriate and limits the discovery of neurotoxic potential (Ladefoged et al., 1995). Dissociation of neuropathology from functional changes may involve a number of factors, including the intrinsic toxicity of a chemical, the dose and regimen of exposure, the age of the animals exposed and the sensitivity of the tests. In addition, the nervous system maintains a level of compensatory capacity as a mechanism of repair and has been shown to possibly retain a level of regenerative capacity in certain brain regions. However, although such repair processes exist, they are not fully understood and do not appear to result in the nervous system returning to a completely normal state. Rather, the nervous system returns to a relatively normal state in which it remains somewhat altered and possibly compromised in its response to future insults. Greater understanding of the structural complexity, connectivity and various cell–cell interactions has clearly demonstrated that the level of examination required to identify such discrete changes is significantly greater than that conducted in a general morphological or histopathological examination. However, the level of sensitivity in detection of neuropathological changes can be enhanced by a more careful histopathological examination of the nervous system.

Various types of neuropathological lesions may be classified according to the site where they occur (Spencer et al., 1980; Spencer & Schaumburg, 1985; IPCS, 1986b; Krinke, 1989; Griffin, 1990). Within each general class of nervous system structural alteration, there are various histological changes that can occur. The degenerative process of the nerve cell can be either relatively rapid or prolonged, depending on the underlying mechanism responsible. For example, neurons can degenerate following a direct action on the cell body, following loss of synaptic target site influences, loss of trophic factors or loss of stimulus innervation from other neurons. Each process may require examination along the neuronal projection field to detect the level of injury induced. Guidelines exist for tissue preparation and examination of the nervous system (IPCS, 1986b). However, guidance remains sparse regarding the neuroanatomy of the brain, such as specific brain regions for examination, associated neural pathways, types of cellular alterations and other unique features of “screening” nervous system tissue for damage as compared with other organ systems.

Histological evaluation often relies solely on routine stains such as haematoxylin and eosin; however, the addition of immunohistochemical staining for specific cell types and cell processes can serve to complement traditional histological evaluations. One special stain recommended in various guidance documents is an immunological stain for the major structural protein of astrocytes, glial fibrillary acidic protein. In response to injury and excessive neural activity, the astrocytes will increase in size, resulting in an increase in this structural protein. This can occur at both the primary site of injury as well as the projection sites of injured neurons. The detection of astrocyte hypertrophy in distinct brain regions can serve as an indicator for additional detailed examination. More recently, microglia, associated with inflammatory processes, have been examined in brain tissue following chemically induced injury, with the initial data suggesting that this response may serve as an early indicator of injury. Unlike the neuron, the astrocyte/microglia response does not appear to be influenced by ischaemia/hypoxia and cell shrinkage that can occur with immersion fixation. At low exposure levels, gross neuronal necrosis and astrocyte hypertrophy may not be evident and indeed may not even play a significant role in the neurotoxicity.

Issues with regard to histological examination of the developing brain have been extensively discussed by Garman et al. (2001). Structural evaluation of adverse effects on the developing nervous system poses a set of questions additional to those associated with histopathology. While acute degenerative lesions can occur in the developing brain, quite often the neuropathology assessment is primarily one of identifying chemically induced alterations in determination of cell fate (numbers and locations) and the normal developmental process. With low levels of exposure, one may assume that a gross necrotic lesion would not be the likely manifestation of damage, but rather a disarrangement of the normal cytoarchitecture of the brain. Some of the proposed methods to evaluate such effects have included both qualitative and quantitative morphological assessment. In addition to histological assessment, quantitative evaluations can be conducted, including end-points such as brain weight and, although not yet validated, morphometric dimensions. Differential sensitivity in the degrees of retardation of brain development may be expected from one area of the brain to another. For example, areas that mature after birth (e.g. cerebral cortex, cerebellum and hippocampus) might be more affected by chemical exposure than are subcortical structures that develop in utero. When examining a delay in development of the brain or an effect on a specific cellular structure, biochemical and molecular methods can be used to more closely examine such effects. For example, ontological profiles of developmentally regulated structural proteins and associated messenger RNAs (mRNAs) can provide evidence of delayed or altered synapse formation, astrocyte maturation or myelin formation (Toews & Morell, 1999) that can be used to complement morphological findings.

Unlike other organs, the actual size and weight of the brain are relatively unaffected by mild to moderate changes in total body weight. Such "brain sparing" is typically seen in undernourished adult animals but may also occur in the developing animal and does not necessarily preclude delayed or otherwise abnormal brain development. Delayed brain development and smaller brains can be seen in undernourished juvenile animals, yet the ratios of brain weight to body weight for undernourished pups are generally equal to or slightly greater than the ratios for adequately nourished rat pups. Undernutrition can be the result of increased litter size, decreased lactation, decreased maternal nutrition or maternal neglect. Thus, it is critical to control these factors in order to adequately interpret study findings as evidence of chemical-specific neurotoxicity.

Quantitative neuropathological approaches include morphometric evaluation of specific regional structures using linear (linear measurements of a brain or brain region, such as width or length between two specific sites), areal (measurements of the two-dimensional area of a brain region) or stereological measurements (measurements that are assumed to provide a more three-dimensional compilation of two-dimensional measurements of a brain region). Although such quantitative evaluations may offer discrete measurements, there is considerable debate as to the validity of such methods to uniformly represent the brain region of interest, both within a subject as well as between subjects. This debate involves, for example, the variability of these measurements, the many factors that can contribute to these measurements, such as plane of cut through the brain that must be standardized in each study, ill-defined topographical markers, insufficient database, lack of validation of methods for toxicological assessment and varied assumptions underlying each method. More recent imaging methods allow for three-dimensional reconstruction of a brain and the determination of total volume of any specific brain region. Magnetic resonance imaging may allow for an accurate evaluation of altered brain development and identification of specific target sites. However, this is based on the assumption that structural components of the region would be disrupted in a manner that would cause a change in volume. Alterations in the connectivity of a region would not necessarily be detected using any of these types of structural evaluations.

4.8.3.2 *Neurobehavioural evaluation*

Evaluation of neurotoxicity is not performed routinely for all chemicals, but only when indicated (e.g. from structure–activity considerations or the results of other toxicity tests). Among the various approaches for assessing neurotoxicity, behavioural testing in conjunction with neuropathological evaluation has been considered a practical approach to assess functional integrity of the nervous system. Behaviour is an adaptive response of an organism, orchestrated by the nervous system, to internal and external stimuli. A behavioural response represents the integrated end-product of multiple neuronal subsystems, including sensory, motor, cognitive, attention and integrative components, as well as an array of physiological functions. Thus, behaviour can serve as a measurable index of the status of multiple functional components of the nervous system.

Behavioural testing has been established as a reliable toxicological index, and considerable progress has been made in the standardization and validation of neurobehavioural testing procedures (IPCS, 1986b, 2001a; Tilson, 1990a; Eisenbrandt et al., 1994; OECD, 1995a,b, 1997; EC, 1996, 1997; Catalano et al., 1997; Moser, 1997; Moser et al., 1997a,b,c,d; Tilson et al., 1997). Neurobehavioural assessment methods are used routinely to evaluate the effects of developmental neurotoxicants on sensory, motor and cognitive functions (Tilson, 1998; Cory-Slechta et al., 2001). It is important to recognize that as neural function interacts dynamically with the status of other organ systems (e.g. cardiovascular, endocrine and immunological systems), certain patterns of behavioural change may indirectly reflect significant primary toxicity in those other organ systems.

4.8.3.3 *Developmental neurotoxicity*

Developmental neurotoxicity has been defined as any effect on the developing nervous system before or after birth that interferes with normal nervous system structure or function. IPCS (1986b, 2001a) addressed some of these concerns and highlighted specific differences between the adult and immature nervous systems. The developing nervous system as a unique target system for adverse effects has been addressed in an ILSI-sponsored workshop with a review of testing methods and assessments of nervous system injury. This review considered available testing guidelines and identified approaches that can be used to assess adverse effects following exposure during development (Cory-Slechta et al., 2001; Dorman et al., 2001; Garman et al., 2001; Mileson & Ferenc, 2001). Since then, the OECD has adopted a guideline for developmental neurotoxicity (OECD, 2007). Additional concern for adverse effects on the developing nervous system has been presented in many reviews regarding endocrine disrupting agents (USNRC, 1993, 1999; USEPA, 1998a,b; EC, 1999; Damstra et al., 2002).

It has long been known that critical windows of vulnerability exist during the formation and maturation of the nervous system (e.g. the period of the brain growth spurt) (Rodier, 1990; Isaacson & Jensen, 1992a,b). The mammalian central and peripheral nervous systems are complex structures resulting from critically timed developmental processes, including cell proliferation, differentiation, apoptosis, migration, synaptogenesis and myelination. Each brain region develops according to specific and unique temporal profiles, with a critical interdependence between each structure for stimulus input and projection target sites. The final neural network pattern is dependent upon the integration of selective neural connections between all cell types of the brain. This process begins during prenatal life and continues through adolescence, with plasticity throughout adult life.

In evaluating the potential of a chemical to disrupt the formation and maturation of the neural network, a number of factors must be considered. These include 1) the developmental stage of the target tissue or the specific nervous system component, 2) the mode or mechanism of action of the toxic agent, 3) the dose of the agent delivered to the target tissue, 4) the toxic end-point of interest, 5) the age of the offspring during testing and 6) the method used to evaluate the outcome. Toxicological effects on the nervous system depend on the delivered dose, exposure duration and the developmental stage at which exposure occurred. Pharmacokinetic processes governing chemical disposition within the adult and in the offspring will also have an influence (see review by Dorman et al., 2001). In addition, unique physical features such as the placental barrier and the maturation of the blood–brain and blood–nerve barriers significantly influence chemical disposition. Neonatal exposure may depend on maternal pharmacokinetic processes and transfer of the substance through the milk, although direct exposure can occur from other routes.

4.8.4 Tiered testing strategy

A number of expert groups have recommended tiered testing strategies for the evaluation of chemically induced neurotoxicity (e.g. IPCS, 1986b; United States Congress, Office of Technology Assessment, 1990; USNRC, 1992; EC, 1996; USFDA, 2000). The initial phase of a tiered testing strategy is the identification of neurotoxicity at some dose level (hazard identification). Tests designed to measure the presence or absence of an effect are usually different from those used to assess the degree of toxicity or type of toxicity or to determine the lowest exposure level required to produce an effect (Tilson, 1990a).

Screening procedures are first-tier tests typified by their capability to assess a large number of animals. Such procedures do not require extensive resources, are usually simple to perform and can yield semi-quantitative data (Moser, 1989, 1995; O'Donoghue, 1989; Schulze & Boysen, 1991; Moser et al., 1997a,b). Systematic clinical observation, such as the USEPA's functional observational battery, is considered an essential part of first-tier testing. Clinical signs have been criticized as being highly variable and poorly documented. Thus, numerous efforts have been made to place observation of clinical signs under a systematic protocol. For any first-tier test, a screening technique should include the following: 1) clearly defined methods and end-points, 2) quantified end-point using an explicitly stated rating scheme, 3) trained and experienced observers and 4) an adequate number of end-points assessed to evaluate multiple modalities of nervous system function. Observations should detect signs of significant neurological disorders, behavioural abnormalities, physiological dysfunctions and any other signs of nervous system toxicity. In addition to the animal's physical appearance, body posture and weight, the clinical screen should provide sufficient information to assess the incidence and severity of such end-points as seizure, tremor, paralysis or other signs of neurological disorder, the level of motor activity and alertness, the animal's reactivity to handling or other stimuli, motor coordination and strength, gait, sensorimotor response to primary sensory stimuli, excessive lacrimation or salivation, piloerection, diarrhoea, polyuria, ptosis, abnormal consummatory behaviour and any other signs of abnormal behaviour or nervous system toxicity. Assessment of cognitive functioning is not usually a component in first-tier screens. The specific composition of the screen and the end-points to be recorded should be consistent with the particular focus of the study and be appropriate for the age and species of the animals to be tested.

Although observational methods are conceptually the most straightforward, they are also the easiest to confound and can sometimes be difficult to interpret without some internal or external corroboration of results. A quantitative measure of locomotor activity, limb grip strength and hindlimb foot splay can be considered as first-tier tests. Often, such functional tests are used in conjunction with other methods, including neuropathology. Given the various biological modalities encompassed in nervous system function and the numerous end-points examined, questions can arise concerning the significance of a change in any one specific screening end-point. As a result of the IPCS-sponsored international collaborative study on neurobehavioural methods for the functional observational battery, motor activity and grip strength, a clustering approach was proposed as one method to deal with such data (Moser et al., 1997a,b,c,d). This approach clusters the various observations into functional domains that represent common neurobiological processes (i.e. autonomic, motor and sensory function), generating a composite response to reflect the functional integrity of a given subset of neurological processes. This approach would allow data to be evaluated within a small number of neurobiologically meaningful clusters rather than numerous isolated end-points. In all cases, it is important that the neurotoxicity screening information be supplemented with any other relevant toxicological findings.

There are a number of publications to guide the design and conduct of testing appropriate for neurotoxicity screening of the adult (Deuel, 1977; Tupper & Wallace, 1980; Gad, 1982, 1989; Vorhees, 1987; O'Donoghue, 1989; Broxup, 1991; Schulze & Boysen, 1991; USEPA, 1991c; Tilson & Moser, 1992; Chang & Slikker, 1995; Moser et al., 1997a,b) and the developing organism (Buelke-Sam et al., 1985; Wier et al., 1989; Rees et al., 1990; Rodier, 1990; Nelson, 1991; USEPA, 1991b; Slikker, 1997).

The second tier of neurotoxicity testing utilizes more specific tests than the first tier and is designed to characterize the nature and dose–response for the neurotoxic effect. A decision to test at the next tier is based on data suggesting that an agent produces neurotoxicity, including neurotoxicological data already in the literature, structure–activity relationships, data from first-tier testing or reports of specific neurotoxic effects in humans. The choice of the most appropriate approach is dependent on the scientific questions generated by the results of the first-tier testing or other available data. These specialized tests are often more sensitive, may contribute information concerning mode of action and are aimed at objectively quantifying effects and determining NOAELs or BMDs. Second-tier tests often yield graded or continuous data amenable to routine parametric statistical analysis.

Third-tier testing may involve mechanistic studies that attempt to establish a detailed profile of a chemical's effect at several levels of nervous system organization (i.e. behavioural, physiological, cellular, molecular). Such tests could provide detailed information on enzyme function, ionic balance, signal transduction, transmitter systems, receptor modulation and underlying molecular mechanisms as they relate to the pathogenesis of effects. It is from such studies that understanding of the processes underlying neurotoxicity and specificity of effect is gained. Mechanism or mode of action studies, when linked to the pathogenesis, provide a basis for the development of biologically based models of neurotoxicity.

4.8.5 Cholinesterase-inhibiting compounds

Inhibition of a specific enzyme, acetylcholinesterase (AChE), has been shown to occur with some neurotoxicants, such as the organophosphate and carbamate pesticides. This enzyme hydrolyses the neurotransmitter acetylcholine, and inhibition results in prolonged action of acetylcholine at receptor sites. Objective clinical measures of cholinergic overstimulation (e.g. salivation, sweating, muscle weakness, tremor, blurred vision) can be used to identify such an effect and the dose–response relationship (Moser, 1995). Generally, the acute cholinergic effects of anticholinesterase compounds are viewed as reversible (ECETOC, 1998), although longer-lasting effects have been reported in animals (Tandon et al., 1994; ECETOC, 1998). Tolerance may be observed following repeated exposure to cholinesterase-inhibiting chemicals; however, the cellular mechanisms associated with this process may lead to other effects not present at the time of initial exposure (Bushnell et al., 1991). There is currently no experimental evidence for lasting or persistent effects of repeated exposure to organophosphates at levels that do not produce significant inhibition of brain AChE (Ray, 1999). Depending on magnitude and time course, a given depression in red blood cell or brain AChE activity may or may not be accompanied by clinical manifestations. Reductions in brain AChE are usually considered as adverse, whereas reductions in plasma and red blood cell cholinesterase are considered as indicative of possible adverse effects. Reductions in plasma butyrylcholinesterase serve as biomarkers of exposure. Low levels of inhibition of AChE are tolerated, whereas inhibitions of 20% or more are considered to be significant for risk assessment purposes. All available data on brain, blood and other tissue cholinesterase activity, as well as the presence or absence of clinical signs and neuropathology, should be evaluated for cholinesterase-inhibiting chemicals on a case-by-case basis using a weight of evidence approach (ECETOC, 1992; Padilla et al., 1994; USEPA, 1998a).

A subset of organophosphate agents, such as tri-*o*-cresylphosphate and leptophos, can produce a delayed neuropathy (organophosphate-induced delayed neuropathy [OPIDN]) after acute or repeated exposure. This degenerative process involves primarily demyelination of long axons of both the peripheral nerves and the spinal cord. It is not clear whether this process occurs in all species; however, humans are known to be highly susceptible, and the adult hen is the experimental animal model of choice. Chemicals that can cause OPIDN in the hen are generally regarded as unacceptable for use as pesticides. The observed ataxia is clinically "irreversible", although the picture can change from a flaccid paralysis (peripheral nerve plus central nervous system lesions) to a spastic paralysis (central nervous system lesions only). Initiation of OPIDN has been associated with the inhibition and "ageing" of neuropathy target esterase (NTE) (Johnson, 1990; Richardson, 1995). Comparison of the semi-log relationship between dose and NTE inhibition and clinical manifestation suggests that more than 70% of NTE inhibition/ageing is required for OPIDN to develop.

4.8.6 Alternative test methods

Attention has been directed to the development of in vitro systems for assessing the neurotoxicological impact of chemical agents (United

States Congress, Office of Technology Assessment, 1990; Harry, 1998; USEPA, 1998a; USFDA, 2000; IPCS, 2001a). The nervous system is composed of highly specialized, heterogeneous, integrated populations of cells. Thus, it is unlikely that a single in vitro test or even a battery of in vitro tests would be able to mimic the responses of the nervous system to a broad range of chemically induced toxicity. Given the complicated nature of the interdependent interactions of the various cell types and network processes in the nervous system, it would be unwise to conclude that a chemical does or does not have neurotoxic potential based upon data from in vitro systems alone. However, batteries of in vitro tests do offer the possibility of developing additional or more appropriate first-tier screening methods for inclusion in a test battery.

This does not diminish the value of information gained from in vitro test systems; it just emphasizes the requirement that any such data be placed within the framework of a limited representation of nervous system function and the toxicokinetics of a given substance. In general, the consensus is that in vitro/alternative test systems offer the greatest strength in hypothesis-based mechanistic studies (Harry, 1998) that may allow one to refine subsequent second-tier study designs, resulting in an overall reduction in animal use.

4.8.7 Interpretation of data

Neurotoxicity is one of several non-cancer end-points that share common default assumptions and principles. The evaluation of the validity of the database is a primary step in the interpretation of data as indicative of a potential neurotoxic effect. This requires four principal questions to be addressed to provide a useful framework for evaluating either laboratory animal or human studies or the weight of evidence for any given chemical (McMillan, 1987; Sette & MacPhail, 1992; Health Canada, 1994; Hertel, 1996; IPCS, 2001a):

1) Do the effects result from exposure?

2) Are the effects neurotoxicologically significant?

3) Is there internal consistency among behavioural, physiological, neurochemical and morphological end-points?

4) Are the effects predictive of what will happen under various conditions?

Although there are known differences between experimental animals and humans in sensitivity to some neurotoxicants, available data support the general assumption that an agent that produces an effect in the laboratory animals will pose a potential hazard to humans (Kimmel et al., 1990; Kulig, 1996; Spencer et al., 2000). Criteria for the quality of data necessary for use in risk assessment to represent the pattern of effects seen in vivo or to define neurotoxicity have been addressed in detail by IPCS (2001a). In general, the value of test methods for quantitative neurotoxicity risk assessment is related to a number of criteria, including 1) sensitivity of the test method to detect differences between exposed and non-exposed groups, 2) specificity for neurotoxicity end-point in a chemical exposure, 3) reliability (consistency of measurement over time) of both the measurement and the effect and 4) validity (concordance with other behavioural, physiological, biochemical or anatomical measurements of neurotoxicity). A relationship between exposure level and severity of response or inclusion of additional functional effects adds support for the observed neurotoxicity. Impairment across a number of functional domains lends support to characterization of an effect within a specific component of the nervous system (e.g. motor, sensory). Comparability of test methods across experimental animals and humans as well as information on underlying mechanisms associated with the neurotoxic response are of particular value. These issues are discussed in detail in USEPA (1998a) and IPCS (1986b, 2001a).

4.9 Immunotoxicity

4.9.1 Introduction

Immunotoxicology focuses on unintended modulation of the immune system. Effects that may occur include immunosuppression, immunostimulation, hypersensitivity and autoimmunity. These may result in outcomes such as increased incidences of infectious or neoplastic diseases, allergy/asthma or autoimmune diseases. To date, immunotoxicity risk assessment efforts have focused primarily on the potential for chemicals to suppress the immune system, as there is a general acceptance of the relevance of immunosuppression end-points in humans and experimental animals for the determination of human risk (see reviews by Vos & Van Loveren, 1998; Descotes, 2003; Luebke et al., 2006), and on identifying allergic contact sensitizers (see section

4.10 and reviews by Basketter et al., 2002; Gerberick et al., 2007; Van Loveren et al., 2008).

Numerous studies have been published suggesting that while immunosuppression is not a common occurrence in the human population, it is not rare. A number of epidemiological studies suggest that alterations in immune responses have arisen as a result of exposure to chemical contaminants in foods (reviewed in Luster et al., 2005).

4.9.2 Assessment of immunotoxicity

4.9.2.1 Laboratory animal studies

Although the toxicokinetics of some chemicals may differ between experimental animals and humans, rodents have proven to be useful models for examining the immunotoxicity of compounds that do not have species-specific effects because of the similarities in rodent and human immune systems. However, some degree of caution must be exercised, as there are instances where concordance between the effects in humans and other species, or even between different rodent species, does not occur. Toxicokinetic data may provide useful information with regard to interspecies differences. Immune system changes observed at overtly toxic dose levels should be interpreted cautiously, as stress and malnutrition are known to impair immune responsiveness. Inclusion of a positive control group, exposed to a well-characterized immunosuppressant, is important in data interpretation and to validate the robustness of the assays conducted.

(a) Standard toxicology studies

Data from standard toxicology studies, such as those conducted in accordance with OECD Test Guideline No. 407 (OECD, 2008) and the ICH S8 guideline (ICH, 2005), provide insensitive, but sometimes useful, information on immunological end-points. Changes in immune system parameters may accompany generalized toxicity affecting all organ systems, reduced body weight secondary to reduced food consumption and significantly reduced protein or micronutrient intake, or stress responses that induce increased corticosteroid production. Under these conditions, altered immune system end-points should be interpreted with caution, as they are unlikely to occur at doses that do not cause generalized toxicity. In the absence of overt toxicity, lymphoid organ weights (absolute and relative) are useful, as they are suggestive of dystrophic or dysplastic changes. However, alterations in mean organ weights are by themselves poor predictors of immunotoxicity, and changes in immune system organ weights should not be the sole criteria used to determine immunotoxicity. Instead, these data should be considered along with other changes (e.g. functional immune response, histopathological parameters) as part of a weight of evidence approach to evaluate whether immunosuppression has occurred.

Haematological data, including erythrocyte counts, haemoglobin, haematocrit, mean corpuscular volume, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, platelet count, total number of leukocytes and leukocyte differentials, as well as clinical chemistry data, such as the ratio of albumin to globulin, total immunoglobulin levels (if available) and a liver enzyme panel, are often included in standard toxicology studies. These end-points provide baseline information on other organ systems that may affect the immune system, as well as basic information on the supply of immune cells. For example, changes in erythrocyte parameters or leukocyte counts may indicate altered bone marrow function and the potential for decreased production of immune cell precursors, and shifts in the ratio of albumin to globulin may signal decreased antibody synthesis. Changes in these end-points may suggest that specific immune function assays are necessary to determine the existence of immunosuppression; however, these data alone are not considered to be reliable predictors of immunotoxicity, as these end-points may be within normal limits, even in children with primary immunodeficiencies.

(b) Immunology studies

Immunotoxicologists have applied tiered panels of assays to identify suppressive immunomodulatory agents in laboratory animals. The configurations of testing panels vary, but they typically include assessment of more than one of the following: 1) lymphoid organ weights and histopathology, 2) quantitative assessment of lymphoid tissue cellularity and peripheral blood haematology, 3) immune cell function at the effector or regulatory level and 4) host resistance studies involving infectious or neoplastic challenge. The first tier is usually a screen for immunotoxicity, whereas subsequent tiers consist of more specific or confirmatory studies, host resistance studies or in-depth mechanistic studies.

Histopathology. From a histological standpoint, assessment of the mammalian immune system is fairly complex. It is composed of multiple organs and tissues, some of which are responsible for haematopoiesis (bone marrow), others for lymphocyte maturation (thymus) and still others that generate responses to antigen (lymph nodes and spleen). In addition, there are specialized tissues located throughout the body that are responsible for responding to antigens or pathogens locally (e.g. lymphoid tissues associated with the skin, lung and gut). Alterations in function in these tissue-associated lymphoid tissues can result in unique adverse effects. The biological processes responsible for the immune response suggest that immunotoxic chemicals that operate by altering antigen recognition or antigen-dependent responses would most likely manifest histopathology in secondary lymphoid organs (spleen, lymph node), coinciding with an active immune response. In contrast, agents that operate through nonspecific cytotoxic or antiproliferative processes would be expected to present histopathology in both primary (thymus) and secondary lymphoid organs, being more apparent in lymphoid organs that undergo extensive proliferation and self-renewal.

Gross and microscopic examinations of lymphoid tissues are important steps in the assessment of the potential for compounds to induce immunotoxicity. A number of studies indicate that histopathological evaluations of lymphoid tissues can be good predictors of potential immunotoxicity, provided that an appropriate level of stringency (histological score) is applied when assessing lesions and that standardized scoring, quality assurance and controls are used to ensure that subtle histopathological lesions can be consistently identified (ICICIS Group Investigators, 1998; Harleman, 2000; Germolec et al., 2004a,b). Histological lesions, particularly in the thymus, have been shown to be sensitive indicators of immunotoxicity, and lesions in the thymic cortex correlate well with altered antibody production. The use of histopathology as a screening tool for immune system toxicity would be advantageous, as these evaluations could be conducted during routine toxicology studies, such as the 28-day rodent study, without the need for additional animals (Kuper et al., 2000).

A working group within the Society of Toxicologic Pathology has developed and published a Best Practice Guideline for the routine pathology evaluation of the immune system, which identifies specific methodology and standardized terminology most appropriate for the detection and reporting of histopathological alterations to immune tissues (Haley et al., 2005). This working group agreed that three primary points should be emphasized when following the recommended "semiquantitative" evaluation of changes in lymphoid tissues: 1) lymphoid tissue sections should contain separate compartments that support specific immune functions, 2) these separate compartments should be evaluated individually for changes and 3) descriptive, rather than interpretive, terminology should be used to document changes within each compartment.

Histopathological evidence may be available from a range of tissues, and the utility of the data for risk assessment would depend on the degree of pathology, the extent of involvement of multiple organs and the biological rationale and likelihood of the histopathology to represent an adverse response to chemical exposure. For example, a lesion within the thymus or bone marrow may suggest suppression. However, a bone marrow lesion that is characterized by reduced progenitor cells in the bone marrow with a resulting reduction in specific cell types in the thymus or peripheral blood is stronger evidence that functional defects are likely to occur. Histopathology, haematology and clinical chemistry changes can also provide information in a weight of evidence approach to support immunotoxicity.

Lymphocyte phenotyping. Lymphocyte phenotyping is one of the most commonly utilized clinical measures of the immune system. Lymphocyte counts do not usually correlate with changes in immune function or host resistance unless marked changes occur. However, reductions in specific lymphocyte populations can be good indicators of overall changes in immune function (Luster et al., 1992). In addition, because lymphocyte phenotyping can be conducted in human studies, use of this measure in laboratory studies allows for comparison of effects across species. A number of different flow cytometry protocols are available for lymphocyte phenotyping, and standard protocols have been established following interlaboratory comparisons (e.g. Burchiel et al., 1997). To perform the assay, single-cell suspensions are prepared from blood or spleen (although thymus, lymph nodes or bone marrow preparations are also used), stained with cell surface marker–specific antibodies and analysed by flow cytometry. A wide variety of commercial cell–type specific antibodies are available that bind to cell surface antigens, such as OX19+, the pan T cell marker in rats, or OX8+, which, when combined with OX19+ antibodies, identifies CD8+ T cells. Changes in lymphocyte subpopulations can be expressed as either a change in the absolute number of a specific cell type or a change in relative cell populations (i.e. ratio of CD4 to CD8).

Functional measures of immune responses. A detailed description of tests and methods used to screen compounds, evaluate resistance to infection or neoplastic challenge or determine mode or mechanism of action is beyond the scope of this chapter. Reference works (e.g. Burleson et al., 1995; Vohr, 2005) are an excellent source of detailed protocols and discussions of assay merits and shortcomings. The information that follows is a brief description of the tests that are commonly used to evaluate immune function in laboratory animals.

Humoral immunity—The utility of the T cell–dependent antibody response (TDAR) as a marker for immunosuppression hazard identification is 2-fold: 1) antibody synthesis is crucial for successfully controlling a wide range of infectious agents and associated toxins, whether immunity is the result of a previous infection or the result of deliberate immunization; and 2) antibody synthesis requires that a complex series of events take place, involving multiple cell types and multiple cellular products. The TDAR requires functional macrophages (antigen processing), T_H cells (source of stimulatory cytokines) and B cells (differentiation into antibody-producing plasma cells) and is generally considered to be an excellent indicator of overall immune function, especially when combined with certain routine toxicology tests, such as thymus weights (Luster et al., 1992). A variety of methods have been used to evaluate TDARs, particularly measuring antibody responses following immunization with sheep red blood cells or keyhole limpet haemocyanin. The number of antigen-specific antibody-producing cells can be measured in the spleen (plaque-forming cell assay or enzyme-linked immunosorbent spot [ELISPOT]) or from serum samples (enzyme-linked immunosorbent assay [ELISA] or haemagglutination assays). By varying the detecting antibodies in the latter assay systems, specific antibody subclasses can be quantified.

Cell-mediated immunity—Cellular immunity is traditionally thought of as reactions mediated by T cells, exclusive of the T_H component of antibody responses. Cytokines released by antigen-specific T cells amplify inflammatory responses against intracellular pathogens, down-regulate normal immune responses to prevent tissue damage, affect contact-dependent killing of altered host cells and suppress the activity of self-reactive cells associated with autoimmunity. In cell-mediated responses to pathogens, sensitized CD4+ T cells (from an earlier encounter or from immunization with specific proteins) respond to a challenge by producing cytokines that provide the activation signals required by macrophages to become bactericidal or cytolytic and participate in eliminating the infection. The delayed-type hypersensitivity (DTH) response provides a comprehensive assessment of the ability of T cells to respond to intracellular infections. The DTH response is used not only clinically to determine whether individuals have been previously exposed to a certain organism (e.g. *Mycobacterium tuberculosis*), but also as a measure of T cell reactivity, by testing with antigens that the majority of the population will respond to. Following intradermal injection of an extract of the organism, significant swelling and redness will be apparent 24–48 h later in individuals who have been sensitized by prior exposure to the organism. The response is referred to as "delayed" because of the time lag between antigen challenge and the host response. Immunotoxicologists evaluate the DTH response by immunizing animals to antigens such as egg or bovine serum albumin or keyhole limpet haemocyanin, typically by subcutaneous injection in combination with an adjuvant. The animal is subsequently challenged by intradermal injection of the same antigen, and swelling at the injection site is carefully measured after an additional 24 h.

Cytotoxic T lymphocytes play a central role in destroying chemically or virally modified host cells and neoplastic cells bearing tumour antigens. Their function is typically assessed by culturing antigen-primed T cells, generated either in vivo or in vitro, with labelled tumour cells or foreign lymphocytes and measuring label release. Because clonal expansion of antigen-specific cells is critical to immune function, the proliferative capacity of T cells has been used as an ex vivo correlate of clonal expansion, although the predictive value of the assay is limited (Vos & Van Loveren, 1998). Thus, an in vitro proliferative response to foreign cells such as allogeneic lymphocytes (e.g. the mixed lymphocyte response) or direct stimulation of the T cell receptor using an antibody to the receptor (anti-CD3) can be used as a functional correlate of T cell replication. The potential ability of lymphocytes to proliferate in response to nonspecific agents, known as mitogens, which stimulate lymphocytes to enter the S-phase of the cell cycle, has also been utilized as an indicator of overall immune system health, both clinically and in experimental animals. Mitogens are commercially available that stimulate proliferation of T cells, B cells or both subsets of lymphocytes. Because antigen receptors are not engaged and the normal process of responding to an antigen is bypassed, these relatively nonspecific measures of cell-mediated and humoral-mediated immunity have proven to be of limited predictive value (Luster et al., 1992).

Innate immunity—Innate immunity refers to responses that do not require antigen recognition or cell division/maturation. Some measure of innate immune function is generally included in tiered testing panels, although the specific end-points may vary depending on potential targets or regulatory requirements. The methods employed to evaluate the functional status of macrophages and neutrophils following exposure to suspected immunotoxicants vary considerably, ranging from measures of phagocytic activity to release of a growing list of soluble mediators to complex bactericidal or tumoricidal activities, including the release of reactive oxygen or nitrogen. Tumor cell lysis by natural killer (NK) cells is one of the primary tests of innate immune function and immunotoxicity associated with chemical exposure. Lytic function is measured by quantifying the proportion of tumour cells (target cells) that have been lysed following co-incubation with NK cells (effector cells) collected from the spleen or peripheral blood.

Disease resistance measures or host resistance assays. The major function of the immune system is to protect the individual from infectious or neoplastic disease. As practised in immunotoxicology, experimental animals are challenged with sufficient numbers of transplantable tumour cells or pathogenic organisms to produce disease at a low level or in a small number of control animals. These “host resistance assays” are often considered particularly relevant for validating the usefulness of other methods to evaluate immune function and for extrapolating the potential of environmental agents to affect clinical disease in the human population. Host resistance models that utilize human pathogens have been developed for use in experimental animals; these and others that closely mimic human disease processes are most commonly employed. In general, host resistance assays represent the final level of the screening process and are conducted only when there are indications of alterations in immune function in the primary screen. Although host resistance assays are often considered to be the ultimate predictor of adverse effects, functional immune tests are predictive of host resistance (Luster et al., 1993). Although it is relatively rare for compounds that produce no alterations in functional immune tests to affect disease resistance in the commonly used models with the increasing sensitivity of the end-points used in host resistance tests, these types of studies may detect suppression of immunity at dose levels where no effects are observed in specific functional tests (Van Loveren, 1995).

Because the immune mechanisms that mediate resistance differ for different pathogens, a single host resistance model is usually not suitable to study all possible consequences of immunosuppression. Selection of particular challenge models (see Table 4.2) is based upon experimental considerations, such as the route of chemical exposure and results obtained from initial immune evaluations, which provide an indication of which immune cells or processes are targeted by the toxicant. Although some models have been adapted for use in both rats and mice, to date, the majority of host resistance studies conducted have been in the mouse. Reference materials are available that contain background information and specific protocols for the conduct of these studies (e.g. Burleson et al., 1995; Coligan et al., 2005).

(c) Evaluation of allergic contact dermatitis

Guinea-pigs were traditionally used to test the sensitizing potential of chemicals, but animal costs, sensitivity issues and subjectivity of the assay end-point led to the development of other assays (Burleson et al., 1995). The mouse ear swelling test (MEST) is similar to the guinea-pig assay in that both immune sensitization and elicitation of an immune response phase are required. In the MEST, a compound is applied to the ear pinna and evaluated by measuring changes in ear thickness following challenge. An alternative test is the local lymph node assay (LLNA), in which the test material or appropriate control is applied topically in three successive daily applications to both ears of the test species, usually the mouse. Cell proliferation is subsequently measured in the lymph nodes draining the ears. At least one

Table 4.2. Commonly employed disease resistance models

Challenge agent	End-point measured
Listeria monocytogenes	Colony-forming units in spleen and liver, morbidity
Streptococcus pneumoniae	Morbidity
Plasmodium yoelii	Parasitaemia
Influenza virus	Morbidity, tissue burden
Cytomegalovirus	Morbidity, tissue burden
Trichinella spiralis	Numbers of parasites in muscle or intestine
PYB6 sarcoma	Tumour incidence (subcutaneous)
B16F10 melanoma	Tumour burden (lung nodules)

concentration of the test chemical must produce a 3-fold increase or greater in lymphocyte proliferation in the draining lymph nodes of test animals compared with vehicle-treated control mice to be considered a positive. The LLNA is currently the method of choice for determining skin sensitizing potential, as it provides a marked refinement and reduction in animal use compared with guinea-pig assays without a loss of accuracy (Dean et al., 2001; Basketter et al., 2002; Gerberick et al., 2007).

4.9.2.2 *Human studies*

Retrospective epidemiological studies have typically been employed to detect potential immunotoxicity in humans following inadvertent exposure to chemicals. The method has been used to evaluate individuals with transient high-level occupational exposure, small cohorts following accidental exposures or large cohorts with chronic low-level exposures. The assessment of immunotoxicity in humans is complicated by the need to account for confounding factors, such as genetic diversity, age and lifestyle factors (e.g. tobacco, alcohol or drug use). Testing strategies for assessing immunological effects in individuals potentially exposed to immunotoxic chemicals have been detailed in EHC No. 180 (IPCS, 1996), EHC No. 212 (IPCS, 1999) and EHC No. 236 (IPCS, 2006a), and the reader should refer to these documents for a more comprehensive discussion of the clinical measures that may be employed. In general, immunological testing has been limited to one or two assays that are relatively insensitive measures (e.g. lymphocyte counts or immunoglobulin levels) and are best at identifying severe immunological effects, rather than mild to moderate changes in immune responses. Some of the more comprehensive immunotoxicology studies in humans have demonstrated immunosuppression in different populations of children following prenatal or postnatal exposure to persistent organochlorine compounds (e.g. polychlorinated biphenyls [PCBs]) via maternal diet and breast milk (reviewed in Luster et al., 2008).

Although human immune function data are generally not incorporated in human retrospective epidemiological studies, these types of data represent the strongest evidence of immunosuppression. However, a few studies have measured antibody titres to common vaccine antigens following immunization in adults (Sleijffers et al., 2003). Similar studies, conducted in conjunction with established vaccination programmes for newborns and young children (e.g. measles, diphtheria, tetanus and poliomyelitis), present a significant opportunity to assess chemical-induced alterations in immune status in populations with identified chemical exposure. Reduced antibody responses following immunization with several childhood vaccines have been observed in infants and children with perinatal exposures to PCBs (Weisglas-Kuperus et al., 2000; Heilmann et al., 2006).

Surface marker analysis (immunophenotyping) and serum immunoglobulin levels are the most commonly employed tests to evaluate immunological changes in human studies. These tests are routinely conducted in large hospitals and have provided considerable information on the ontogeny and activation state of the human immune system. In many human studies, statistically significant differences have been found between the control and case populations with respect to serum immunoglobulin levels and cell surface marker analysis of lymphocytes. However, because of the large variability in historical control values, case values may be significantly different from control values, while being within historical normal ranges. This was observed in a study of children with halogenated aromatic hydrocarbon exposure (Weisglas-Kuperus et al., 1995). However, exposure was also associated with a significant increase in inner ear and respiratory infections (Weisglas-Kuperus et al., 2000). These data indicate that exposure may result in minimal to mild shifts in observational end-points, essentially clustering at one end of the normal range. As such, when evaluating observational immune system data collected during epidemiological studies, data obtained from routine toxicity testing (e.g. immunoglobulin levels, white blood cell counts, immunophenotyping) or functional data (e.g. vaccine titres) to identify potential immune system hazards, emphasis should be placed on statistically significant differences in values for exposed and appropriately matched controls, rather than on whether values for the exposed population fall within a broad range of normal.

4.9.3 Interpretation of data on immunotoxicity

As of 2009, formal guidance for chemical immunotoxicity risk assessment has not been published, although efforts are under way in the USA and Europe to develop guidelines.

In order to accurately predict the immunotoxic risk of exposures in human populations, a scientifically sound framework should be used to support an accurate and quantitative interpretation of experimental and epidemiological studies. Thus, when reviewing immunotoxicology data, it is important to examine multiple end-points and to determine that the results are biologically plausible. Regardless of the end-point being measured, data generated to assess immunotoxicity must be considered in their entirety, including dose responsiveness, general indications of toxicity, the appropriateness of the test methods and the historical predictive value of the data. It is important that information on immunosuppression be considered together with other health effects in the overall characterization of risk.

4.9.4 Conclusions

Immunosuppression represents a series of complex cascading cellular and organ-related events that can lead to an increased incidence or severity of infectious and neoplastic diseases. Unintended immune stimulation is not well understood, but can lead to increased allergic and autoimmune responses. Therefore, it is not surprising that the data from experimental immunotoxicology or epidemiological studies that are used to address quantitative risk assessment issues require careful interpretation. To improve risk assessment for immune system toxicity, it will be necessary to increase our understanding of the underlying immunomodulatory mechanisms that cause adverse effects and the quantitative relationships between the immunological tests conducted in the laboratory and actual disease in human populations. This is particularly true when the magnitude of immunological effects is slight to moderate, as may be expected from inadvertent exposures to immunosuppressive agents in the food supply that have been linked to adverse health effects. It is therefore critical to address the potential risks of immune effects following dietary exposures to chemicals, as they have the potential to increase both the burden of disease and the costs of caring for affected individuals.

4.10 Food allergy and other food hypersensitivities

4.10.1 Introduction

Food allergy and other food hypersensitivities are adverse reactions to specific foods and food ingredients occurring in sensitive individuals within the general population (Ebo & Stevens, 2001). These food hypersensitivities are considered individualistic responses, in that most individuals are able to consume these foods without adverse consequences (Taylor & Hefle, 2001). Hence, these types of sensitivities do not include general toxic reactions to foods and food ingredients that could affect any consumer without discrimination provided the ingested dose of the toxic agent is sufficient.

Previously, food allergy was identified as a "form of food intolerance", where there existed "evidence of abnormal immunological reaction to a food" that is "mediated by immunoglobulin E" (IgE). Food intolerance has been defined as "a reproducible, unpleasant reaction to a food or food ingredient, including reactions due to immunological effects, biochemical factors such as enzyme deficiencies, and anaphylactoid reactions, which often include histamine release" (IPCS, 1987).

Since then, there have been several attempts to classify adverse reactions to food (Figures 4.2 and 4.3) (Sampson, 1999; Johansson et al., 2001).

The World Allergy Organization concluded in 2004 (Johansson et al., 2004) that the appropriate term is *food allergy* when immunological mechanisms have been demonstrated. If IgE is involved in the reaction, the term *IgE-mediated food allergy* is appropriate. *Non-IgE-mediated immunological reactions* are called either

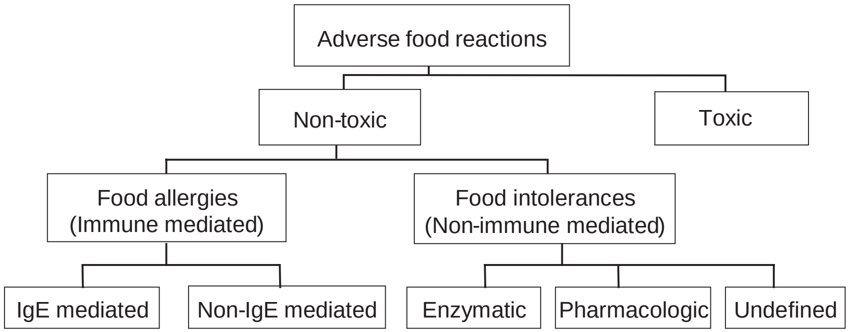


Fig. 4.2. Classification according to the European Academy of Allergology and Clinical Immunology nomenclature task force (adapted from Johansson et al., 2001)

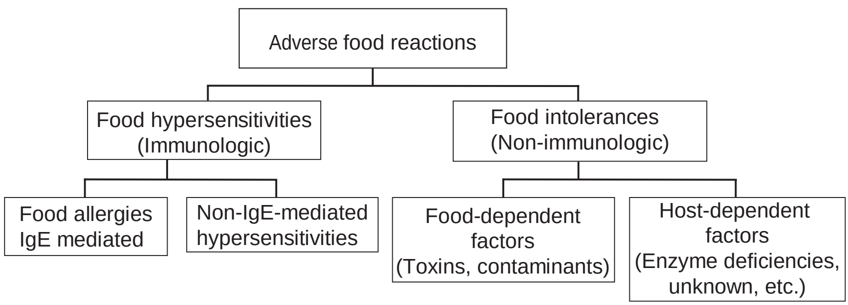


Fig. 4.3. Classification adapted from Sampson (1999)

non-IgE-mediated allergy or non-IgE-mediated hypersensitivity. All other reactions should be referred to as *non-allergic food hypersensitivity.*

A varied range of pathological mechanisms underlie food hypersensitivities. Some conditions involve immunological mechanisms, and others do not. The mechanism can be IgE mediated (Taylor & Hefle, 2001) or partially IgE mediated, as seen with conditions such as eosinophilic oesophagitis or asthma (Sampson, 1999). Immunological reactions can also be non-IgE mediated, being IgG mediated or cell mediated, as seen with disorders such as coeliac disease (Troncone et al., 2008). Finally, some adverse reactions do not involve the immune system (IPCS, 1987; Taylor & Hefle, 2001). These sensitivities may be attributed to the existence of metabolic disorders or the occurrence of reactions with unknown mechanism.

4.10.2 Prevalence

A meta-analysis of food hypersensitivity prevalence studies showed that it is not possible to make an overall worldwide estimate of the prevalence of food allergy or of the prevalence of specific foods, even based on well-conducted studies of prevalence, either self-reported or based on challenge studies (Rona et al., 2007; Zuidmeer et al., 2008).

The heterogeneity in the prevalence reported in different studies could be a result of differences in study design and methodology. Another possibility is that the findings reflect real differences between populations.

In studies of self-reported food allergies, 3–38% answer that they have food allergies, although only a few studies had figures above 20%. If those people who believe that they have a food allergy are challenged with the food that they think causes their allergy, only 1–11% have their food allergy confirmed. Most of the studies in which food allergy is clinically proven report percentages between 1% and 5% of the total population as having any food allergy. So there is a large gap between the percentage of people who think they have a food allergy and the percentage of people who are diagnosed as allergic. In general, the same effect is apparent when specific foods (with the exception of soy and wheat) are investigated: self-reported food allergy is overestimated compared with clinically proven food allergy (Rona et al., 2007; Zuidmeer et al., 2008).

4.10.3 IgE-mediated food allergy

4.10.3.1 Sensitization

The normal reaction to dietary proteins is development of tolerance, where the immune response is downregulated by an active immunological process (Brandtzaeg, 2002; Sampson, 2004).

Food allergies are a consequence of the undesired or uncontrolled immune response to a food antigen in susceptible individuals. They are based on the body's aberrant interpretation of certain dietary proteins as "foreign", which leads to a heightened response of the immune system.

Allergy develops through the process of sensitization. During the sensitization phase, exposure to the food allergen stimulates production of antigen-specific IgE (Taylor & Hefle, 2001).

Sensitization may occur via the intestinal tract. This is called traditional food allergy or class 1 allergy and is often caused by stable allergens. Class 2 food allergy develops after sensitization to airborne allergens via the lung and is typically caused by pollen cross-reacting with food allergens (Asero et al., 2007). Sensitization via the skin may also be possible (Lack et al., 2003). Class 1 food allergy is most prevalent in children, whereas class 2 food allergy is most prevalent in young adults and adults.

In general, milk and egg are the most common food allergens in children, and this is worldwide (Hill et al., 1997; Dalal et al., 2002; Osterballe et al., 2005). Eating habits may influence the development of food allergies. For instance, sesame allergy is frequent in Israel, probably because of early introduction of tahini (Dalal et al., 2002).

Most infants develop cows' milk allergy in the 1st year of life, but about 85% become clinically tolerant by the 3rd year of life (Host et al., 2002). Allergy to hen eggs often develops in the 2nd year of life. Approximately half of these patients become tolerant in 3 years, and up to 66% of children become tolerant in 5 years (Boyano Martínez et al., 2001). Peanut allergy tends to persist throughout adulthood, although up to 20% of peanut-allergic children lose their allergy (Skolnick et al., 2001; Hourihane, 2002).

The foods that most often cause allergy in adults are fruits and vegetables (Kanny et al., 2001; Zuberbier et al., 2004; Osterballe et al., 2005). Here, the primary sensitization comes mainly from pollen, and thus sensitization does not reflect eating habits, but rather exposure to flora.

Factors such as age, genetic predisposition and amount and frequency of food consumption may play a role in sensitization, but there is no current consensus regarding a threshold dose for sensitization for food allergens (see section 4.10.3.4).

It is important to remember that sensitization (e.g. the induction of specific IgE upon exposure to an allergen) is not the same as clinical disease. This means that detection of specific IgE in serum or a positive skin prick test is not always accompanied by clinical disease (Asero et al., 2007).

4.10.3.2 *Symptoms and diagnosis*

The symptoms of food allergies range from mild discomfort to severe, life-threatening reactions (anaphylaxis), which require immediate medical treatment. Symptoms may be triggered in the skin (e.g. itching, redness, swelling), gastrointestinal tract (e.g. pain, nausea, vomiting, diarrhoea, itching and swelling of oral cavity), respiratory tract (e.g. itching and swelling of the nose and throat, asthma), eyes (e.g. itching and swelling) or cardiovascular system (e.g. chest pain, abnormal heart rhythm, very low blood pressure causing fainting, and even loss of consciousness). Fortunately, anaphylaxis is much less frequent than skin rashes or symptoms in the gastrointestinal tract.

Allergic reactions to foods may occur within a few minutes after eating the offending food, but symptoms may also (rarely) develop after hours, making the relationship with ingestion of food less clear. Symptoms can last for days. The specific symptoms and severity of an allergic reaction are affected by the type and amount of the allergen consumed, by the form in which the food containing the allergen was eaten, by the intake of alcohol, aspirin and other drugs such as beta-blockers and angiotensin-converting enzyme inhibitors, by exercise or stress, and by the sensitivity of the allergic person.

The most frequent symptoms of food allergies are itching and swelling of the mouth. Oral itching (known as oral allergy syndrome) can be an initial symptom in any kind of food allergy. Oral itching is, however, a well-known symptom in food allergy induced by cross-reaction with pollen, such as by apple, kiwi, hazelnut, walnut, celery, carrot, tomato, cherry and melon. Most of the allergens in cross-reacting foods will be destroyed in the gastrointestinal tract. This explains why the symptoms are frequently mild and limited to the mouth. Most of the allergens in the cross-reactive foods will be destroyed if the food is cooked. Many people allergic to birch pollen cannot eat raw apples without experiencing symptoms, but stewed apples and apple juice might not be a problem (Asero et al., 2007).

Anaphylaxis is an uncommon, acute, potentially life-threatening allergic reaction involving the whole body. A person who has this type of reaction will typically experience the following symptoms: itching of the skin or tingling in the mouth and throat followed quickly by feeling unwell and dizzy with an accelerated heart rate and nausea. At the same time, there may be a nettle rash or skin flushness, hay fever and asthma. Blood pressure may drop dangerously, and the person may collapse. Untreated anaphylaxis can rapidly result in death.

An unusual form of this condition can be triggered by eating problem foods within 2–3 h of vigorous exercising and is called "food-dependent, exercise-induced" anaphylaxis.

In Europe and the USA, peanuts and nuts are the foods most commonly reported to cause anaphylaxis (Pumphrey & Gowland, 2007; Shah & Pongracic, 2008). In Japan, milk, egg and wheat seem to be the most common foods associated with anaphylaxis (Immamura et al., 2008). Prompt administration of the medicine adrenaline after eating suspected problem foods has helped minimize life-threatening episodes. Applicators to administer adrenaline can be carried by people who are aware that they are at risk of anaphylaxis (Shah & Pongracic, 2008).

4.10.3.3 Common characteristics of food allergens

Virtually all known food allergens are proteins. The traditional food allergens (class 1) are water-soluble glycoproteins 10–70 kilodaltons in size and fairly stable to heat, acid and proteases (Sampson, 2004).

The food allergen component of a food represents only a few of a vast number of different proteins found in the complex mixture that comprises a food (Taylor & Lehrer, 1996; Becker & Reese, 2001). They can be less prominent proteins in the allergenic foods (Taylor & Lehrer, 1996). Most allergenic foods contain multiple allergenic proteins. When assessed with regard to the nature of their reactivity in sensitive individuals, the allergenic food proteins can be considered as a "major" food allergen or a "minor" food allergen, depending on whether, respectively, a majority or a minority of atopic or allergic individuals react to it (Taylor & Lehrer, 1996; Bredehorst & David, 2001).

A relatively small number of specific foods or food groups are responsible for the vast majority of food-related allergic reactions (Hefle et al., 1996; Sampson, 1999). The foods or food groups identified as key in this regard by an international expert panel (FAO, 1995) are cows' milk, eggs, peanuts, soybeans, wheat, tree nuts (e.g. almond, walnut, pecan), fish (e.g. finfish: cod, salmon) and crustaceans (e.g. shrimp, crab, lobster). Some food additives may also give IgE-mediated allergic reactions (Kägi et al., 1994; Wüthrich et al., 1997; Chung et al., 2001).

The relevant component of the primary protein structure of food allergen is an epitope. Epitopes are the part of the whole allergenic proteins or glycoproteins that are detected immunologically by antibodies (Lehrer et al., 1996; Becker & Reese, 2001). They serve as the interface between the chemical structure of the food allergen protein and the immune system. Different types of epitopes exist (Huby et al., 2000). Continuous epitopes are peptides of a length of 6–16 amino acid residues in a linear sequence (Lehrer et al., 1996; Becker & Reese, 2001). Discontinuous epitopes comprise different components or several different adjacent non-continuous amino acid sequences of the primary protein structure and depend on conformational or tertiary three-dimensional structure of the protein (Lehrer et al., 1996; Becker & Reese, 2001). The latter type of epitopes have the most potential to be altered or destroyed by denaturation and thus factor in the stability of food allergens, especially with respect to aspects of food processing (Becker & Reese, 2001). Epitopes can also be composed of glycoconjugate carbohydrate determinants, possibly causing glycosylated food allergens to be resistant to denaturation (Huby et al., 2000; Becker & Reese, 2001).

Systematic analysis of plant food allergens has shown that the majority belong to only a few protein structural families, the prolamin, Bet v 1 and cupin superfamilies (Breiteneder & Mills, 2005; Jenkins et al., 2005). Animal food allergens can be classified into three main families—tropomyosins, EF-hand proteins and caseins—along with 14 minor families, each composed of 1–3 allergens. The evolutionary relationships of each of the animal allergen superfamilies showed that, in general, proteins with more than approximately 62% sequence identity with a human homologous protein were rarely allergenic (Jenkins et al., 2007). These observations indicate that the structural features and properties of food proteins may play a role in determining their allergenicity.

For class 1 allergy, where sensitization occurs via the gastrointestinal tract, resistance to digestion may be important (Astwood et al., 1996). Thus, the ability of a protein to sensitize and to elicit allergic reactions via the gut may depend on the extent to which it survives digestion. This has been shown for a number of prolamin superfamily members, with IgE epitopes having been found to resist digestion for the 2S albumin allergens from Brazil nut (Moreno et al., 2005) and peanut (Sen et al., 2002) and for the lipid transfer protein allergens from grape and various Rosaceae fruits (Asero et al., 2000; Scheurer et al., 2004; Vassilopoulou et al., 2006). However, this hypothesis does not hold for the cupin allergens, such as the peanut allergen Ara h 1, which, despite being susceptible to proteolysis, retains its allergenic properties (Eiwegger et al., 2006). There is evidence that low molecular weight peptides form aggregates of a size sufficient both to sensitize and to elicit an allergic reaction (Bøgh et al., 2008).

In addition to digestive processes, allergenic food proteins are potentially altered by food preparation processes, including heat (e.g. roasting, cooking), proteolysis and hydrolysis (Bredehorst & David, 2001). The allergenicity of certain food proteins has been demonstrated to be less potent, more potent or, more commonly, unaltered to any significant degree after food processing or cooking procedures. These differences in reactivity that result from changes in food allergen proteins may vary across allergic individuals. Recently, a workshop concluded that it is not currently possible to identify specific variables that could be used to reliably determine how processing will influence protein allergenicity (Thomas et al., 2007).

Class 2 food allergy develops as a consequence of an allergic sensitization to inhalant allergens cross-reacting with allergens in fruits and vegetables. These class 2 allergens are in general more labile than allergens causing class 1 allergy and most often cause oral allergy syndrome (e.g. typical for the birch–apple syndrome), but they can also cause anaphylaxis, which is not rare in the mugwort–celery syndrome (Breiteneder & Ebner, 2000).

Not all allergies to fruits and vegetables are caused by labile pollen cross-reacting with allergens (Fernandez-Rivas et al., 2006). For example, lipid transfer proteins in peach and apple are very resistant to processing (Asero et al., 2000).

4.10.3.4 Thresholds

(a) Sensitization

There is no current consensus regarding a threshold dose for sensitization for food allergens. Nor is there information delineating the differences in sensitization threshold across age groups, routes of sensitization or the combination of both. In addition, the parameters that define the process of sensitization—for example, the amount of allergen ingested per exposure, the number of exposures, the duration of exposure, the pattern of exposures and even the total dose of exposure—are not well defined.

(b) Clinical food allergy (elicitation)

Exposure to low or minimal amounts of an allergenic food is potentially hazardous to individuals with an allergy to that food. Hence, determination of a "safe" or tolerable level of exposure is critical to those individuals with an allergy to a specific food. Risk assessment methodologies allow for the estimation of this level.

For food allergy, knowledge about hazard and adverse effect levels comes from case-reports and case-series or from challenge studies performed on sensitive individuals. Food challenge tests are typically conducted to diagnose the presence of a food allergy in individuals suspected of sensitivity to a particular food. The data from challenge tests available in the literature are from open challenge tests, single-blind placebo-controlled food challenge (SBPCFC) tests, meaning only the patient is unaware of the food or placebo being tested, and double-blind placebo-controlled food challenge (DBPCFC) tests, meaning neither the patient nor the test administrator is aware of the food or placebo being tested. Of these food challenge tests, the findings from DBPCFC test protocols are considered the more reliable and valid source of dose–effect information (e.g. Bock et al., 1988; Hourihane et al., 1997; Taylor et al., 2002; Bindslev-Jensen et al., 2004). It is sometimes referred to as the "gold standard" protocol.

Oral food challenge trials have shown large individual differences in human reactivity to allergenic food, from 0.01 mg to several grams of protein (Taylor et al., 2002; Wensing et al., 2002; Ballmer-Weber et al., 2007).

Over recent years, more focus has been directed towards the performance of low-dose DBPCFC tests to determine the NOAEL as well as the LOAEL for allergenic foods (e.g. Hourihane et al., 1997; Taylor et al., 2002, 2004; Wensing et al., 2002; Flinterman et al., 2006; Ballmer-Weber et al., 2007). A part of this process has been to publish consensus standardized clinical protocols for low-dose DBPCFC tests. The goal of these protocols is to be able to more confidently compare food challenge results across studies and to reduce the variability in these results (Taylor et al., 2004; Crevel et al., 2008).

Different allergenic foods may have different NOAELs or LOAELs. This may reflect real differences in potency or differences in the allergic population investigated in challenge trials. Reviews of challenge data can be found in Taylor et al. (2002), in EFSA (2004) and at http://www.foodallergens.info.

Because of potentially severe reactions (anaphylaxis), some patients are excluded from food challenge procedures. In addition, patients are included in challenge trials when their symptoms are stable and they have no infections. For these reasons, it is often debated whether results from challenge trials reflect the reactivity in the whole population allergic to the food investigated. On the other hand, low-dose DBPCFC trials are conducted at university allergy clinics where the patient group may be more sensitive than the ordinary food-allergic patient (Crevel et al., 2008).

4.10.3.5 *Risk assessment in food allergy*

It is assumed that food-allergic persons are able to avoid the food to which they are allergic if the allergenic food is an ingredient in the food they eat. This means that risk assessment is typically conducted in situations where the allergenic food occurs not as an ingredient, but as a "contaminant" (e.g. milk in dark chocolate). Another important area is the exemption from labelling requirements (e.g. to determine if the level of residual protein in highly refined soybean oil is so low that there is no risk for persons with soy allergy).

In food allergy, risk assessment is based on data from challenge trials in food-allergic patients, intake data, levels of contamination with the allergenic food and, if possible, prevalence data. Most risk assessments have been done on a case-by-case basis, taking relevant information into account. The risk assessment concludes whether or not a level of allergen contamination will result in adverse reactions in food-allergic persons (EFSA, 2004). One of the big challenges for the risk assessor is that there is consensus that a threshold for food allergy reactions exists (Taylor et al., 2002), but it is not possible, based on current data, to set scientifically based thresholds for allergenic foods (EFSA, 2004).

Food allergy risk assessment is a relatively new discipline, and there is no general consensus on how it should be conducted. Three approaches have been suggested, using 1) NOAEL and uncertainty factors, 2) BMD and margin of exposure (MOE) and 3) probabilistic risk assessment (Madsen et al., 2009). The three approaches are described below (Madsen et al., 2009).

Risk assessment in food allergy using thresholds and uncertainty factors depends on the use of data from challenge trials that identify a NOAEL or a LOAEL. The relevant study that reports the lowest NOAEL (or LOAEL if a NOAEL cannot be identified) is used. The NOAEL can be based on either subjective or objective symptoms. The NOAEL is then divided by an uncertainty factor. There is no consensus on the use of uncertainty factors in food allergy, but it has been suggested that a factor of 10 be used to account for intraspecies differences and an additional factor of 10 to account for potential severity of reaction in the highly sensitive population (Buchanan et al., 2008). The advantage of this approach is that it is very simple and uses a methodology well known from toxicology. The disadvantage is that it is based on a single data point from a single study and may result in thresholds that are too low to be of practical use. For further discussion, see Madsen et al. (2009).

Instead of using a single data point from a single study, the use of mathematical modelling based on distribution of positive challenges from a single study or from a combination of challenge studies with the same allergenic food has been suggested. This allows the determination of a BMD (in food allergy, also called the eliciting dose) for this food based on all available relevant data (Crevel at al., 2007). A collection of data from peanut challenges of 185 patients from 12 studies was used to estimate the BMD using distribution models. The ED_{10} (i.e. the dose expected to give reaction in 10% of the peanut-allergic population) was found to be 17.6, 17.0 or 14.6 mg whole peanut, depending on the model used (Taylor et al., 2009).

The MOE approach generally uses the lower 95% confidence limit of the BMD. This is called the benchmark dose lower limit (BMDL) (see chapter 5).

The BMDL is divided by the estimated intake of the allergenic food, resulting in an MOE. Different intake scenarios can be compared as well as MOEs for different allergenic foods, in order to identify susceptible subgroups (e.g. high consumers) or to judge relative potencies of allergenic foods.

The advantage of the approach is that it uses all relevant data to establish a BMD. The disadvantage is that it does not describe the risk quantitatively. For examples and discussion, see Madsen et al. (2009).

The probabilistic risk assessment model calculates the most likely number of allergic reactions that might result from the accidental presence of an allergenic constituent in a food product. This calculation uses the distribution of positive challenges, together with those associated with variables determining the intake of the allergenic constituent. These include presence and concentration in the affected food, likelihood that an allergic person consumes the food and amount of the food consumed per eating occasion (Spanjersberg et al., 2007). The advantage of this approach is that it results in a quantitative estimate of a risk. The disadvantage is the demand not only for challenge data, but also for distribution of intake data.

As in other areas, a good risk assessment relies on the quality and suitability of the data used. In food allergy, the data used originate from humans, but there may be limitations in using existing data, because they were generated for other purposes. More and more threshold data on allergenic foods are being generated using standardized protocols with an extended range of doses, often starting at low microgram levels, generating NOAELs and LOAELs that can be used in risk assessment (Taylor et al., 2004; Flinterman et al., 2006; Ballmer-Weber et al., 2007; Crevel et al., 2008).

A reaction to a food allergen is analogous to an episode of acute poisoning rather than chronic toxicity in terms of dosimetry. Therefore, the relevant exposure assessments should be based on “meal/eating occasions” rather than exposure throughout the entire day or from a single food.

There has been much focus on the development and use of challenge data in food allergy risk assessment and much less focus on how intake data should be used. Both the MOE and the probabilistic approach use intake data, which, depending on how they are used, may influence the outcome of the risk assessment. For further discussion, see Madsen et al. (2009).

4.10.3.6 *Evaluating potential allergenicity of genetically modified food*

A part of the evaluation of the safety of genetically modified (GM) foods is to assess whether newly introduced proteins have allergenic potential. The purpose of this is 2-fold: 1) to protect food-allergic persons from exposure to the allergen and 2) to protect the population from introduction of new food allergens.

To predict the potential allergenicity of novel food proteins, two decision tree strategy approaches have been described (Metcalfe et al., 1996; FAO/WHO, 2001b).

The Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology (FAO/WHO, 2001b) proposed a decision tree for assessing the allergenic risks posed by novel proteins, which is an update of the original decision tree described in Metcalfe et al. (1996).

FAO/WHO (2001b) suggested that cross-reactivity between the expressed protein and a known allergen (as can be found in the protein databases) should be considered when there is either:

1) more than 35% identity in the amino acid sequence of the expressed protein (i.e. without the leader sequence, if any), using a window of 80 amino acids; or
2) identity of six contiguous amino acids.

As an identity of six contiguous amino acids between an allergen and a given protein sequence has a high probability of occurring by chance, verification of potential cross-reactivity would be warranted when criterion 1) is negative, but criterion 2) is positive. In this situation, suitable antibodies (from a human or animal source) would have to be tested to substantiate the potential for cross-reactivity.[1]

The decision tree suggested by FAO/WHO (2001b) shows that if a protein has an identity score that equals or exceeds 35%, the protein should be considered to be a likely allergen, and no further testing is suggested.

If there is no sequence homology between the novel protein and known allergens, the recommendation from the FAO/WHO (2001b) consultation is that the protein should be tested against patients' sera. In the case of a GM food, if the source of the gene is known to be allergenic, sera from patients allergic to the source should be tested in a so-called "specific serum screen". This indirectly identifies protein epitopes recognized by allergic patients' IgE, the presence of such epitopes conferring a risk of the novel protein triggering allergic reactions in individuals with a pre-existing sensitivity. If this specific serum screen is negative or if the source of the gene is not known to be allergenic, the protein should then undergo a "targeted serum screen". Thus, if the recombinant protein is derived from a monocotyledonous plant source, it is proposed that serum samples from patients with high levels of IgE antibodies to monocot allergens such as grass and rice be tested. Similarly, if the recombinant protein is derived from a dicotyledonous plant, serum samples from patients with high levels of IgE antibodies to dicot allergens such as tree pollen, weed pollen, celery, peanuts, tree nuts and latex should be used. A similar approach is suggested if the recombinant protein is derived from a mould, an invertebrate or a vertebrate. Such a screen should include 25 individual serum samples with high levels of IgE to the selected group of airborne allergens and (if applicable) 25 sera with IgE to the selected group of food allergens.

This targeted serum screen will determine whether the novel protein has IgE epitopes identical to those present in related inhalant or food allergens. This approach is pertinent, as a number of food allergies

[1] Using as few as six contiguous amino acids was later shown to be useless because of many false positives (Stadler & Stadler, 2003).

are caused by cross-reaction to inhalant allergens. However, with our current lack of knowledge regarding the mechanisms of food allergy, the positive predictability of the targeted serum screen is not known, making a risk assessment difficult.

The Codex Alimentarius Commission (CAC) later abandoned the decision tree strategy and described a risk assessment procedure based on a weight of evidence approach (FAO/WHO, 2003).

There are no validated animal models that can predict the allergenicity of an unknown protein. The risk assessment therefore relies on a combination of methods looking at protein structure, protein stability and binding properties to serum IgE from allergic patients.

The following elements are included in the Codex guideline (FAO/WHO, 2003):

- Identifying the source of the gene
  - Does it come from a known allergenic food?
    - If yes, screen with specific serum from allergic patients
- Sequence similarity with a known allergen
  - More than 35% identity in the amino acid sequence using a window of 80 amino acids
    - Screen with specific serum from allergic patients
- Resistance to pepsin digestion

It has been commonly accepted that for a protein to sensitize an individual and elicit an allergic reaction, it must survive the acidic and proteolytic environment of the gastrointestinal tract. Astwood et al. (1996) showed in a study comparing the in vitro stabilities of food allergens and non-allergenic proteins to simulated gastric fluid that there was an association between resistance to digestion and allergenic potential. This has led to pepsin resistance being used as a predictive parameter in the risk assessment of the allergenic potential of novel proteins, as suggested in all three approaches above. However, in recent years, the relationship between resistance to digestion and allergenic potential of a protein and the validity of taking this parameter into account in risk assessment have been questioned (Fu et al., 2002). It is still true that many allergens giving rise to class 1 food allergy are relatively resistant to digestion, but there are also important exceptions, such as the cupin superfamily, represented by the major peanut allergen Ara h 1 (Eiwegger et al., 2006), and the milk allergen casein, which is degraded relatively quickly by proteases (Wal, 2001). There are also examples of stable proteins that rarely cause allergy, such as thaumatin-like proteins from grape and apple (Vassilopoulou et al., 2006).

For further discussion of the scientific basis for allergenicity testing of GM food, see Goodman et al. (2008) and the European Food Safety Authority's draft scientific opinion on the assessment of allergenicity of GM foods (EFSA, 2009).

4.10.4 Non-IgE-mediated food allergy

4.10.4.1 Coeliac disease

The most well-described and prevalent non-IgE-mediated disorder caused by an immunological reaction to a food component is coeliac disease, also called gluten intolerance. It is a disease of the small intestine triggered by ingestion of gluten, a protein found in wheat, barley and rye. When a person with coeliac disease ingests gluten, an immunological reaction in the small intestine leads to flattening of the mucosa.

In the present text, coeliac disease is classified as a non-IgE-mediated food allergy. This definition is easy to communicate. Most people know about food allergy, and the treatment for coeliac disease, avoidance diet, is the same as for food allergy. Coeliac disease may also be seen as a multiorgan autoimmune disease, primarily as a gastrointestinal disease, but also with effects on the skeletal system, the peripheral and central nervous systems, the reproductive system and the cardiovascular system.

It is estimated that about 1% of the population has antibodies connected to coeliac disease. Wheat can also trigger IgE-mediated food allergy, although this is not as common as coeliac disease.

Coeliac disease was for many years diagnosed mainly in small children. Within months of starting a gluten-containing diet, susceptible children would present with chronic diarrhoea or loose stools, vomiting, a distended abdomen and failure to thrive. Similarly, diarrhoea, weight loss and general weakness are the most common symptoms in adults.

Today, we know that coeliac disease is a complex disorder with symptoms occurring not just in the gastrointestinal tract. Many symptoms and diseases are associated with coeliac disease. For example, the flattened mucosa caused by coeliac disease leads to poor absorption of nutrients in the intestine. Poor absorption of iron can lead to anaemia, poor absorption of vitamin B_{12} can lead to dementia, and poor absorption of vitamin D and calcium can affect bones and teeth. Coeliac disease is also often found in connection with other immunological diseases, such as diabetes and rheumatoid arthritis.

Coeliac disease is diagnosed on the basis of histological findings on a biopsy from the small intestine. In addition, symptoms should disappear on a gluten-free diet.

Patients with coeliac disease have IgA antibodies in serum against gluten as well as autoantibodies directed towards the enzyme tissue transglutaminase. Measurement of antibodies cannot be used as positive proof for the disease. A blood test can, however, help decide whether to take a biopsy from the small intestine.

About 10% of first-degree relatives to patients with coeliac disease also develop coeliac disease. The principal known determinants of genetic susceptibility are the highly variable human leukocyte antigen (HLA) genes located in the major histocompatibility gene complex. It has been demonstrated that the HLA-DQ2 and HLA-DQ8 class II protein molecules present gliadin peptides to T cells in the gut in a particularly efficient way. The HLA-DQ2 and HLA-DQ8 antigens are present in more than 95% of persons with coeliac disease (Troncone et al., 2008).

However, it is clear that additional factors are critical for the development of coeliac disease. Up to 30% of persons of North European ancestry, most of whom eat wheat, express HLA-DQ2, but coeliac disease develops in only a small proportion of these carriers. In Sweden, an epidemic of coeliac disease was started because of the early introduction of gluten-containing cereals (Ivarsson et al., 2000). Altered processing of gluten by gut enzymes and changes in the permeability of the gut may also be important factors (for more information, see the review by Troncone et al., 2008).

The only treatment for coeliac disease is avoiding gluten in the diet. Products with wheat, rye and barley must be avoided. Most patients tolerate products with oats as long as they are free from contamination with other cereals containing gluten. Once a coeliac patient is on a gluten-free diet, the flattened mucosa in the small intestine heals and the symptoms disappear.

(a) Risk assessment

To establish tolerable levels of gluten intake for patients with coeliac disease, it is necessary to challenge the patients over a period of time (e.g. 90 days). Adverse reactions are monitored by following serum antibodies as well as histological changes in the small intestine. A tolerable level of gluten has to be determined for the intake over a period of time and not as with IgE-mediated food allergy, where the dose at a single challenge occasion is the relevant intake scenario. Most patients with coeliac disease should ingest less than 50 mg of gluten per day (Hischenhuber et al., 2006; Catassi et al., 2007).

As opposed to food allergy, a regulatory threshold for gluten has been established. According to the Codex standards for food, gluten-free foods must adhere to a special standard for special dietary use for persons intolerant to gluten (FAO/WHO, 2008). Two standards for "gluten-free" food have recently been established (FAO/WHO, 2008):

1) "gluten-free" products contain gluten at concentrations below 20 mg/kg; and
2) products with "very low gluten content" may contain gluten at concentrations from 20 mg/kg to a maximum of 100 mg/kg.

According to CAC (FAO/WHO, 2008), gluten should be detected by an R5 ELISA method for gluten/gliadin. It is based on a monoclonal antibody reacting with the specific gliadin pentapeptide, QQPFP. This method shows a sensitivity and limit of detection for gliadin of 1.5 mg/kg (Mendez et al., 2005).

4.10.5 Non-immune-mediated food hypersensitivity

4.10.5.1 Metabolic disorders

Metabolic disorders describe those conditions where adverse reactions result from a genetic deficiency in the ability to metabolize some component of the consumed food. Common examples of metabolic food disorders include lactose intolerance, a deficiency of lactase. Lactose intolerance may be inborn (rare), but it mostly appears during adolescence or early adulthood. It is the normal condition in 75% of the human population, but it is relatively rare in northern Europeans, probably occurring in 3–6%. Lactose intolerance may be transient in connection with intestinal infections. Individuals with lactose intolerance are unable to digest lactose and experience adverse gastrointestinal effects associated with bacterial metabolism of lactose in the colon. Small portions of lactose rarely cause symptoms. This means that persons with lactose intolerance normally can eat cheese and smaller amounts of other dairy products.

Favism is a deficiency of erythrocyte glucose-6-phosphate dehydrogenase, with acute haemolytic anaemia resulting from oxidative damage to erythrocytes following the consumption of fava beans containing vicine and convicine.

4.10.5.2 Other

Hypersensitivity to food additives represents a condition for which a mechanism has not been determined; however, reactions are probably not based on an abnormal immune response.

There are few scientific investigations concerning food additives and hypersensitivity, probably because it is a difficult subject to investigate as a result of many different food additives and relatively few people who react to any individual substance. This means that most descriptions of food additive hypersensitivity are based on very few patients.

The one exception is sulfites. Hypersensitivity to sulfites is relatively well described, especially in people with asthma, and may also trigger skin reactions such as hives (urticaria) (Wüthrich, 1993; Taylor et al., 1997).

4.11 General principles of studies in humans

4.11.1 Introduction

The potential value of data from studies in humans has been recognized since the first meetings of JECFA and JMPR.

EHC 70 (IPCS, 1987) stated that JECFA "recognizes the value of human data, has sometimes requested such data, and has always used it in its evaluations when available", whereas EHC 104 (IPCS, 1990) stated that "All human data (accidental, occupational, and experimental exposures) are fundamental for the overall toxicological evaluation of pesticides and their residues in food". EHC 104 (IPCS, 1990) included the following three principles:

1) The submission of human data, with the aim of establishing dose–effect and dose–response relationships in humans, is strongly encouraged.

2) Studies on volunteers are of key relevance for extrapolating animal data to humans. However, attention to ethical issues is necessary.

3) The use of comparative metabolic data between humans and other animal species for the purpose of extrapolation is recommended.

The recent EHC on dose–response modelling (IPCS, 2009) also confirms the value of human data:

> In the evaluation of human health risks, sound human data, whenever available, are preferred to animal data. Animal and in vitro studies provide support and are used mainly to supply evidence missing from human studies. It is mandatory that research on human subjects is conducted in full accord with ethical principles, including the provisions of the Helsinki Declaration [see World Medical Association, 1997].

JMPR has repeatedly considered the use of human data in pesticide risk assessment, in particular when considering ARfDs (see chapter 5). Detailed considerations were given in the 2002 JMPR report (FAO/WHO, 2002a). JMPR noted that human data on a pesticide, whether from volunteer studies or from other investigations of human exposures in the workplace or environment, can be extremely valuable in placing the animal data in context and, when available, should always be evaluated, even when they are not used to derive an ARfD.

Evaluators should consider the following issues in determining whether to use a volunteer study in the derivation of an ARfD:

- The initial consideration should be the ethical acceptability of the study.

- The next consideration should be scientific merit. A poorly designed or conducted study in humans (as with experimental animals) should not be used for establishing an ARfD.
- The acceptable group size will depend on factors such as interindividual variation in response and the level of change considered not to be adverse. The studies should be assessed with particular consideration of their power to detect critical effects.
- The IPCS guidance for the use of CSAFs (IPCS, 2005) proposed a minimum group size of 5. Studies using small group sizes might be usable (e.g. by combining results from two or more dose levels or applying a higher safety factor).
- The critical end-points identified in animal studies should be investigated appropriately in human studies.
- If only one sex or a particular age group has been used, the general applicability of the results should be ascertained, if possible, using data from studies in animals.
- As recommended by the 1998 JMPR (FAO/WHO, 1999a), recent studies in humans should include clear statements that they were performed in accordance with internationally accepted ethical standards. For older studies, ethical considerations should take into account both current standards and the standards pertaining at the time the study was performed.
- Studies that have not been performed in accordance with ethical principles but are scientifically valid should be used only if the findings indicate that acceptable human exposure is lower than the level that would be determined without the use of such a study.

Information from humans is of potential importance in identifying and characterizing the hazards and evaluating the risks of macroingredients in foods and of substances such as food additives, contaminants and residues of veterinary drugs and pesticides. The information may come from:

- controlled experiments in human volunteers, usually related to specific end-points or toxicokinetics;
- surveillance studies, including post-marketing surveillance;
- epidemiological studies of populations with different levels of exposure, which may be particularly important for contaminants;

- experimental or epidemiological studies in specific subgroups of people; or
- clinical reports or case-series of individuals.

Investigations in humans may take the form of short-term experiments involving controlled exposure of a small number of intensively monitored subjects in a clinical laboratory, larger or longer-term and more loosely controlled studies of subjects living in the community but still receiving a controlled exposure, or epidemiological investigations of people in the community, leading a normal life and eating their ordinary diet.

End-points may include examination of safety or tolerance, nutritional and functional characteristics of foods or food components, the metabolism and toxicokinetics of the substance, mechanism or mode of action, possibly using biomarkers for effects identified in animal studies, and adverse health effects from unintentional exposures (e.g. to a contaminant).

The WHO Scientific Group on Procedures for Investigating Intentional and Unintentional Food Additives (WHO, 1967) highlighted

> the need, at a relatively early stage, to obtain information on the absorption, distribution, metabolism, and elimination of the chemical in human subjects, since this makes it possible to compare this information with that obtained in various animal species and to choose the species that are most likely to have a high predictive value for human responses.

Critical issues for any experimental study in humans are the ethical, professional and basic legal controls that govern whether a study in humans is necessary and the circumstances under which it may be properly performed (Royal College of Physicians, 1990a,b; USNRC, 2004). Consideration needs to be given to when the use of human tissues ex vivo or in vitro might be sufficient. Such data are likely to have increasing utility with the incorporation of human metabolic systems into in vivo and in vitro test systems. Prior to undertaking new in vivo experiments in humans, clinical information from other sources, such as investigation of any effects of exposure to the substance of interest in the workplace, reports of overdoses and accounts of human or veterinary medicinal usage of the same substance, should be analysed to determine the necessity of additional research. Increasing ethical concerns about the necessity and safety of studies in humans mean that in the future it may become increasingly difficult to justify and obtain ethics approval for in vivo studies involving the administration of a non-therapeutic substance to humans (see also section 4.11.5).

Of particular value for JECFA and JMPR in evaluating submitted experimental studies in humans are the guidelines developed by VICH (2000) for Good Clinical Practice (GCP). These guidelines include sections on the principles of VICH GCP, the institutional review board/independent ethics committee, the investigator, the sponsor, the clinical trial protocol and protocol amendments, the investigator's brochure and essential documents for the conduct of a clinical trial.

A helpful account of human studies of non-pharmaceuticals, such as pesticides and household products, has been published by Wilks (2001). It discusses the ethical and some of the practical problems and guiding principles that are applicable to items in the diet. Lessons learnt from human studies of pharmaceuticals are described below.

4.11.2 Lessons learnt from pharmaceutical development

Studies in humans are not a formal requirement for international or national safety assessment or regulatory approval of food additives or residues of veterinary drugs and pesticides. However, the information that can be obtained from humans is extremely valuable, and every opportunity should be taken to obtain worthwhile data both before and after a product becomes available for human consumption. In this respect, the regulatory assessment of substances in food differs from that of pharmaceuticals, such as prescription and other medicines, for which studies of efficacy and safety in humans are a data requirement for premarketing evaluation by regulatory authorities.

There are many similarities between the study of substances in food and the study of pharmaceutical compounds, because the basic physiological, pharmacological, immunological and biochemical processes that might be affected by exposure are similar. In addition, many metabolic and toxicokinetic processes of therapeutic drugs are also relevant to other low molecular weight "foreign" compounds, such as food additives, natural non-nutrients, contaminants and residues of veterinary drugs and pesticides.

Human experimental investigations of pharmaceuticals have been developed much further than the clinical evaluation of dietary components and have resulted in the principles and practices governing studies in humans. For that reason, the need for ethical review, professional obligations, laws and official guidelines developed for pharmaceuticals control the nature and circumstances of human studies.

The principles guiding studies in humans have been dominated by the objectives, needs and practices of pharmaceutical development. However, the investigation of drugs differs from some of the purposes, objectives and approaches appropriate to the study of non-pharmaceuticals, especially in the general area of substances in foods. Drug development generally focuses on treating identifiable diseases in population subgroups, often for short periods, and, where necessary, compares the potential benefit with the possible harm of the drug. In contrast, the diet (including food additives, natural non-nutrients, contaminants and residues of veterinary drugs and pesticides) is intended to be harmless and is consumed by all members of society throughout life. The conventional risk–benefit analysis applied to drugs and used to justify various investigations and trials in healthy humans and patients cannot be applied in the same way to studies of foods and dietary components. Ethics committee approval would require that any study on a food substance carries negligible risk to the participants. This leads to a much stricter evaluation of any potential for risk in clinical investigations, because there is no balancing "benefit" in the sense of relief from a disease.

Invaluable and up-to-date information about general and specific requirements for pharmaceuticals can be obtained by consulting the web site of the ICH (see http://www.ich.org/cache/compo/276-254-1.html). More local interpretations of the international guidelines can be obtained from the web sites of major agencies, such as the European Medicines Agency (see http://www.emea.europa.eu/) and those of France (Agence Française de Securité Sanitaire des Produits de Santé) (see http://www.afssaps.fr/), Germany (Bundesinstitut für Arzneimittel und Medizinprodukte) (see http://www.bfarm.de/EN/Home/homepage__node.html), the United Kingdom (Medicines and

Healthcare Products Regulatory Agency) (see http://www.mhra.gov.uk/home/idcplg?IdcService=SS_GET_PAGE&nodeId=5) and the USA (USFDA; see http://www.fda.gov/default.htm).

4.11.3 Types of studies in humans

The principal types of human studies are listed in Table 4.3.

The numbers of subjects entered into a study must be sufficient to realize the aims of the investigation. Ethics approval normally requires a calculation of the group size necessary to meet the study objectives, as it would be unethical to perform an underpowered study. One approach to deciding the size of the experimental groups is to consider normal variability in the end-point being examined and to employ standard statistical methods on the power of an experiment in order to calculate the number of subjects required to demonstrate a predefined magnitude of response. The numbers should include definition of the size of any control group and take into account the predicted drop-out rate. The drop-out rate will depend on various factors, including the nature of any effects produced (although for an ethical study on a food component, this should be minimal) and the overall convenience of the protocol for the subjects (of which duration will be an important consideration).

4.11.3.1 Short-term clinical laboratory studies

The key features of clinical laboratory studies are that 1) they are short term, 2) they are likely to involve relatively few subjects under close supervision, 3) the nature and extent of their exposure to the test material are strictly limited and 4) measures of general safety and tolerance are monitored intensively.

Examples include studies on the toxicokinetics of the substance and examination of any effects on physiological functions and processes, such as the absorption of dietary lipids, plasma cholesterol, uptake of calcium or iron, effects on or replacement of vitamins, actions on intestinal flora, etc.

For food additives, veterinary drugs and pesticides, the absorption, metabolism and excretion in humans can be defined by suitably designed, single-dose studies. The doses chosen would approximate

Table 4.3. Principal types of studies in humans relevant to JECFA and JMPR

Type of study	Principal features	Common reasons for considering it
Short-term	Control of exposure with the administration of low doses predicted to be non-toxic. Intensive monitoring of end-points, effect and safety. Usually in healthy volunteers. Special studies may be undertaken in population subgroups, such as diabetics taking intense sweeteners.	
Physiology	Functional effects on gastrointestinal tract or other body system.	Basic research. Effect of dietary component.
Pharmacology	Interference with normal functions.	Basic research. Potentially harmful effects of dietary components, such as inhibition of AChE. Identification and cause of intolerance.
Biochemistry	Mechanistic investigation of action on metabolic processes.	Basic research. Mechanism of potentially adverse effects, such as enzyme inhibition or enzyme induction. Identification and cause of intolerance.
Toxicokinetics	Absorption, disposition, metabolism and clearance of substance.	Identification of species differences to assist interspecies extrapolation. Identification of genotypic or phenotypic differences to assist identification of possibly vulnerable population subgroups. Validation of biomarkers of exposure.
Immunology	Effects on or via immune system.	Basic research. Potentially harmful effects of dietary components, such as allergic sensitization. Identification and cause of intolerance.

Table 4.3. ***(Continued)***

Type of study	Principal features	Common reasons for considering it
Nutrition	Effects on blood levels of essential nutrients or other biomarkers.	Interference with normal nutritional processes, such as the absorption of micronutrients.
Toxicology	Low exposure usually of limited duration. Sensitive indicator of minimal effect (biomarker).	Mechanistic investigations using reversible biomarkers of effect. Identification and cause of intolerance.
Long-term	In general populations or selected subgroups. Exposure via normal dietary matrix and conventional preparative methods.	
Epidemiology	Case-series, case–control or cohort studies, etc.	Identification and characterization of adverse effects, usually for inadvertent contaminants.
Toxicology	Tolerability	Assessment of general tolerability of an approved substance administered at or close to the health-based guidance value.

those likely to be established as a health-based guidance value based on the available toxicity data. Studies involving the uptake and disposition of labelled materials (e.g. radioactive or stable isotopes) are important in understanding the fate of the substance in the body.

Any immunological, pharmacological, physiological or pathophysiological actions of the substance might be studied using single doses or a small number of doses, but these should be selected so that only minimal and reversible effects would be predicted. Studies would normally involve readily reversible biomarkers of effect, rather than adverse health effects. Short-term studies could also be used to investigate any effect of the substance in food on normal physiological, nutritional, biochemical or other bodily processes, food palatability and taste.

Other short-term studies on the identified end-points of interest, whether biomarkers of kinetics or biomarkers of effect, might include experiments on volunteer patients suffering from a known disease, individuals taking prescription or proprietary medicines, individuals who are genotypically or phenotypically different when the data indicate that this could be a significant variable, and investigations on possible influences of dietary constituents.

It must be emphasized that any special study in a selected group of subjects would require the same justification and ethics approval as for a study in normal healthy volunteers.

The advent of food components prepared from GM organisms, such as enzymes that are evaluated by JECFA, has led to some interest, especially in Europe, in the place of clinical studies in evaluation of their acceptability. An assessment of how to undertake such studies and the criteria for their appropriateness and acceptability have been published by the United Kingdom Advisory Committee on Novel Foods and Processes (FSA, 2002). Most JECFA safety evaluations of food components and processing aids from GM organisms have been on the basis of 90-day studies in rodents.

4.11.3.2 *More prolonged clinical laboratory studies*

In principle, a dietary component might be administered to groups of healthy volunteers or patients for a period of days or even a few weeks, still in a controlled clinical laboratory setting. In reality, interference with normal human activities would mean that if the study were longer than a few days, the design would probably involve the subjects continuing the treatment while pursuing their normal lifestyle and returning to the laboratory periodically for measurements and investigations. This method can provide useful data to support the safety and tolerability of an approved food ingredient; a good example of this approach is the study on aspartame in 53 subjects given 75 mg/kg body weight per day for 26 weeks (Leon et al., 1989).

4.11.3.3 Post-marketing surveillance and epidemiological studies

These investigations involve studying exposure to the substance of interest and effects in people living in their normal communities for periods extending from weeks to months and occasionally longer. They require comparison of the end-points of interest, such as general health status, in groups with different levels of exposure. The different exposures in the groups included in the study often arise from lifestyle or geographical differences.

(a) Post-marketing surveillance

Post-marketing surveillance following the release of the substance in the diet requires that groups with different levels of exposure are identified. This could be a comparison between premarketing and post-marketing or following restricted marketing; for example, the mycoprotein Quorn was initially released in only part of the United Kingdom, which allowed a comparison of any general change in health status for different geographical regions. Obviously, such an approach would be very insensitive and could give only limited reassurance after the event.

The intakes of approved food substances show wide interindividual variations within a group of consumers, and it would be difficult to associate any reported effects with specific levels of intake. Nevertheless, useful insights may be obtained from collation of consumer complaints by the marketing company or the regulatory agency. The USFDA has collated and evaluated claims of adverse effects arising from the consumption of aspartame and the fat replacer olestra (Allgood et al., 2001). It should be recognized that the nature and frequency of anecdotal consumer complaints are likely to be highly influenced by the extent of media coverage of the subject matter.

The uncertainties in such data and the potential sources of unavoidable bias and error make definitive conclusions impossible. Anecdotal reports on individual patients have been historically important in identifying possible adverse effects of therapeutic drugs that were not detected by traditional toxicology testing. Therefore, anecdotal data from consumers should be evaluated to assess the possibility of a previously unrecognized effect from a substance in food.

(b) Epidemiological studies

Epidemiological studies comprise investigations on people in the community in relation to their exposure to the substance of interest. They have been of greatest value to JECFA and JMPR in relation to hazard identification and characterization of food contaminants.

An overview of epidemiological studies in relation to chemicals in the diet is given by Van den Brandt et al. (2002), and the place of and differences between epidemiological and other types of clinical investigation are considered by Duggan et al. (2002). Various guidelines for Good Epidemiological Practice (GEP) have been proposed. Information is available on the International Epidemiological Association web site (see http://www.dundee.ac.uk/iea/GEP07.htm).

In any survey, it is essential not to assume that an apparent association between two or more factors indicates a cause–effect relationship. There are many sources of confounding that may suggest an association that arises indirectly due to other, irrelevant processes and specious correlations; these sources of error are well discussed by Bradford Hill (1965) and in monographs on epidemiology (e.g. Bonita et al., 2006).

The central theme of any epidemiological investigation is the collection of information in such a way as to show whether there is a difference between groups of people exposed to the substance over a given period and an otherwise comparable group that had no exposure or was exposed to a lesser extent (Coggon et al., 1997). The studies are best performed prospectively but may be retrospective (including the use of biological samples collected and stored over many years).

Experience has led epidemiologists to classify ecological and case–control studies as "hypothesis generating"—i.e. the results may suggest that a substance has or lacks a particular action, but the evidence is inconclusive. They should be distinguished from prospective, cohort or intervention studies, which are capable of "hypothesis testing".

The different types of epidemiological studies are described briefly below:

- *Ecological studies or case-series*: These are simpler to undertake than other types of study, because they comprise the collection of a series of past cases of the target event combined with retrospective assessment of their exposure to the test substance for comparison with some local, national or even international data about occurrence of the target event. This type of study is very susceptible to unrecognized and uncontrollable biases and other confounding effects. The main value of such studies is in the recognition of possible associations, and they can act as a trigger for more definitive research.

- *Case–control studies*: These are a more powerful but still relatively simple type of formal epidemiological investigation; as with case-series, however, they have a limited ability to control or even assess many factors that may influence the result. The basis of the approach is a retrospective comparison of the exposure between two groups—patients with the adverse effect or disease of concern and unaffected controls; a higher exposure in the patient group would suggest a possible causative association. The basis of the approach is the collection of relevant information about exposure and perhaps other major factors in the "test" group—i.e. those who suffer from the effect of interest—and in a matched control group whose members do not suffer from the effect. In contrast to other types of epidemiological study, case–control studies can provide information only about the effect that was investigated. Dose–time and dose–response relationships may be suggested by the study results. Typical problems, especially as the data usually come from free-living individuals in the community, are the accuracy of information about exposure and the high possibility of recall bias if the subject matter of the exposure assessment is obvious from the exposure questionnaire.

- *Cohort studies*: These are inherently more precise and more powerful than case–control investigations, but they are more costly to perform, may last a long time and may be more intrusive for the subjects involved. The basis is comparison of the incidence of the target events between groups with different levels of exposure. In many cases, the development of health effects is monitored prospectively. The approach can also be applied retrospectively if the exposure data in the different groups relate to a period before the health assessments were undertaken. Cohort studies usually involve large group sizes and offer the opportunity for better analysis of confounding factors. Dose–response and time–response relationships can be examined, and cautious subset analyses can sometimes be done to indicate the role of other factors not originally considered. A common refinement of the method is to divide the total population studied into bands with different levels of exposure (e.g. tertiles, quintiles) in order to assess dose–response relationships. Cohort studies applied to occupational data may provide information at exposures that are much higher than would normally occur via the diet.

- *Analytical or interventional studies*: These are cohort studies in which the exposure of interest is controlled by the experimenter (i.e. subjects are asked to consume or to refrain from consuming sources of the substance of interest). They are really a large-scale variant of the controlled clinical trial, in this instance employing dietary intervention instead of administration of a medicine. Examples of formal dietary intervention trials include the Alpha-Tocopherol, Beta-Carotene (ATBC, 1994) and the Beta-Carotene and Retinol Efficacy Trial (CARET) (Omenn et al., 1996) studies on vitamins.

4.11.4 Other sources of information about effects in humans

4.11.4.1 Poisoning

Case-reports and case-series from surveillance of accidental or deliberate poisoning cases (e.g. from regional and national poison information centres) are further valuable indicators of the harm that very high doses of a substance can cause.

Like some occupational data, the reports must be interpreted with care in relation to more conventional, lower-dose exposure, but they can still be invaluable in indicating target organs and effects and toxic dose levels. Information about effective therapies can also be a useful guide to the mechanism of the toxic action and to the toxicokinetics of the substance in humans.

4.11.4.2 *Human tissues and other preparations in vitro*

Experiments on human cells or tissues or using other preparations containing or expressing human enzymes, receptors and other subcellular factors in vitro are fundamentally different from studies in people, because they bypass absorption, distribution, aspects of integrated metabolism and excretion. However, an advantage is that they permit mechanistic studies under controlled conditions not feasible in the clinic. Concentration–effect relationships need to be related to the toxicokinetics and possible blood and tissue concentrations of the substance in order to identify those in vitro effects that are feasible in vivo.

These techniques are of considerable value in suggesting metabolic pathways and response mechanisms that may be important in humans and may be worth monitoring as biomarkers of exposure or effect. A further important role of such in vitro experiments is to investigate similarities and differences between humans and test species in the metabolism and effects of xenobiotics that may provide information critical to the extrapolations normally used in risk assessment. In vitro studies are likely to be important in defining CSAFs for toxicodynamics (see chapter 5). They are also of potential value in investigations on the influence of genotypic and phenotypic differences on the metabolism and activities of compounds.

4.11.5 *Ethical, legal and regulatory issues*

Ethical, legal and regulatory issues have to be considered for any study involving humans or human tissues. Some are applicable throughout the world, and others are specific to the locale where the study is done. Associated factors affecting any study in humans are national laws about liability should any harm result from the exposure or the trial, any requirement for insurance coverage against that risk and the legal protection afforded to confidentiality of personal information.

Many of the requirements are mandatory, and non-compliance or breach of them may prevent the study from being done, or there may be legal sanctions and other penalties for all those involved, rejection by official bodies of the information obtained and refusal by editors to consider reports for publication in the biomedical press.

Experiments in humans are strictly controlled to ensure ethical, legal and medical protection of the subjects and the avoidance of foreseeable risks. It is mandatory, therefore, in planning clinical work to justify any proposal to do experimental investigations in humans, especially if it involves data to be used in risk assessment, which may imply uncertainty about risks to which the participants may be exposed. It is necessary to provide a clear, objective explanation as to why only results of experiments in people will provide information that is essential for risk assessment of the material or substance in question. It should be shown how findings from conventional, non-clinical experiments and in vitro and ex vivo studies using human tissues or preparations expressing human enzymes, receptors, etc. cannot give information of the same or similar value for risk assessment purposes.

The most important factor governing a study in healthy people is that a formal evaluation of any possibility of harm to participants and a documented judgement that there is no realistic likelihood of such a risk have been recorded. The fundamental assessment is the same in every type of human experiment, but the nature of the investigation has a considerable influence on the information required to support the evaluation of potential risk. Risk assessments on the proposed studies are an essential part of the evaluation by the institutional review board/independent ethics committee. Evaluation of studies on substances in food would be based on assessment of the likely overall value of the possible research findings and the lack of any predictable risk, based on appropriate non-clinical information.

4.12 Gastrointestinal tract considerations, including effects on the gut flora

4.12.1 General considerations

Interactions that may occur between chemicals in food, including food additives and residues of veterinary drugs, and the bacterial flora of the gastrointestinal tract should be considered in terms of the effects of the gut microflora on the chemical and the effects of the chemical on the gut microflora.

Because the gut microflora is important in the metabolic fate and toxicological activity of some chemicals, the safety assessment should consider the possibility that the chemical in food may affect the host microflora and thereby modify the host response to the chemical in food.

The gut microflora may influence the outcome of toxicity tests in a number of ways, reflecting their importance in relation to the nutritional status of the host animal, the metabolism of xenobiotics prior to absorption and the hydrolysis of biliary conjugation products. JECFA has recognized this and has drawn attention to the usefulness in toxicological evaluations of studies on metabolism involving the intestinal microflora (FAO/WHO, 1971).

4.12.1.1 *Effects of the gut microflora on the chemical*

The spectrum of metabolic activities performed by the gut flora contrasts markedly with that of the host tissues. Whereas hepatic metabolism of foreign compounds is predominantly by oxidation and conjugation reactions, the gut bacteria perform largely reductive and hydrolytic reactions, some of which appear to be unique to the gut flora. Typical reactions include 1) the hydrolysis of glycosides (including glucuronide conjugates), amides, sulfates and sulfamates, 2) the reduction of double bonds and functional groups and 3) the removal of functional groups, such as phenol and carboxylic acid moieties.

From a structural point of view, many chemicals present in food are potential substrates for microbial metabolism. Microbial metabolism of foreign compounds has the potential to convert the molecule into a more toxic form.

The gut bacterial flora is situated principally in the terminal parts of the intestinal tract in most host species and consists primarily of strict anaerobes. Thus, highly lipid-soluble compounds that are absorbed in the upper intestine will not undergo bacterial metabolism unless tissue metabolism produces conjugates that are excreted into the bile and delivered to the bacterial microflora. Clearly, the design of appropriate investigations with the gut microflora must be linked closely to in vivo studies on absorption and metabolism.

There are three primary in vivo methods for studying the role of the gut microflora in the metabolism of a compound:

1) parenteral administration of the compound, which should result in decreased microbial metabolism of poorly absorbed polar compounds, compared with oral dosing;
2) studies on animals in which the bacterial flora is reduced by the use of antibiotics; and
3) studies on germ-free animals and on (formerly) germ-free animals inoculated with known strains of bacteria (gnotobiotic animals).

In vitro incubation of the food additive or its metabolites with the bacteria of the caecum or faeces is a useful but difficult technique, with considerable potential for the generation of spurious data. Some of the pitfalls of prolonged incubations are that the use of a nutrient medium may allow the growth of a non-representative bacterial population and that the use of a non-nutrient medium may act as a powerful selective force for organisms able to use the additive as a source of carbon, nitrogen, sulfur or energy.

A number of factors may influence the metabolic activation of foreign chemicals by the host microflora:

- *Host species*: Species differences exist in the number and type of bacteria found in the gut and in their distribution along the gastrointestinal tract. In this respect, rats and mice are poor models for humans, because the higher pH of the stomach allows the presence of significant numbers of largely aerobic bacteria in the upper intestinal tract; this region is almost sterile in humans, dogs and rabbits, because ingested organisms do not survive the low gastric pH in these species. In addition, coprophagy occurs in rodents and rabbits, which may complicate the kinetics of poorly absorbed compounds and theoretically could enhance the potential for metabolic adaptation.

- *Individual variations*: There is wide variability between individuals within a species in the extent to which some compounds undergo metabolism by the gut flora. Interindividual variability in the hydrolysis of the sweetener cyclamate greatly exceeds the variability in foreign compound metabolism in the liver. Many of these variations probably arise from differences in the enzymatic capacity of the gut flora rather than in the delivery of the chemical to the lower intestine. Thus, if animal studies show that a chemical in food is metabolized by the gut flora to an entity of toxicological significance, it is essential that its metabolic fate is characterized in a sufficient number of humans to define the extent of any variability.

- *Diet*: The composition of the gut flora depends on the diet, which may influence the extent of microbial metabolism of a chemical in food.

- *Medication*: The widespread oral administration of medications, such as antibiotics and antacids, in the human population is a potential source of variation in metabolism by the gut microflora.

- *Metabolic adaptation*: The metabolic capacity of the gut flora is far more flexible than that of the host. Thus, long-term administration of foreign chemicals can lead to changes in both the pattern and extent of microbial metabolism of the chemical. Because prior exposure to the compound under test may significantly alter the metabolic potential of the gut microflora, metabolic studies should be performed not only on previously unexposed animals, but also on animals that have been exposed to the test compound for sufficient time to allow metabolic adaptation (a period of weeks rather than days). For the same reason, any in vitro studies should be performed with caecal contents that have been collected both prior to and during long-term animal feeding studies.

4.12.1.2 *Effects of the chemical on the gut microflora*

During high-dose animal feeding studies, the gut microflora may be affected in two ways:

1) *Antibacterial activity*: A weak antibacterial activity, shown by, for example, a food additive, may manifest after long-term intake of near-toxic doses either as an alteration in the numbers of bacteria present, which can be measured directly, or as an abnormal

microbial metabolic pattern. The latter can be studied by measurement of certain endogenous metabolites produced only by the gut flora, such as phenol and *p*-cresol, which provide indirect evidence of alterations in the gut flora. Such information may also be of value in the interpretation of other variables, such as nitrogen balance.

2) *Increased substrate for gut microflora*: The chemical may act directly as a substrate for bacterial growth. This can be readily illustrated by appropriate high-dose pharmacokinetic studies, coupled with in vitro metabolic studies on the gut flora. Alternatively, the chemical may inhibit digestion or absorption of other dietary components so that these become available to the bacteria in the lower intestine in increased amounts.

Increased amounts of substrates in the lower intestine provide an increased osmotic effect in the caecum, which may result in caecal enlargement. The reason for caecal enlargement must be studied before the significance of the lesion can be assessed, because it may be indicative of 1) abnormal osmotic balance with consequent changes in permeability to minerals in the caecum, which could lead to nephrocalcinosis; 2) microbial metabolism of nutrients, which could result in the formation of potentially toxic metabolites and abnormalities in the nitrogen balance; or 3) microbial metabolism of the chemical, which might lead to the formation of toxic products.

4.12.2 Decision tree approach for determining the potential adverse effects of residues of veterinary antimicrobial drugs on the human intestinal microflora

The potential for antibiotics in food to alter the intestinal flora is an important safety consideration. The only class of veterinary drugs to date that JECFA has evaluated for which the ADI is based on the selection of resistant bacterial strains is the tetracyclines (FAO/WHO, 1999b). At its fifty-second meeting, JECFA developed a decision tree for evaluating the potential effects of veterinary drug residues on human intestinal microflora (FAO/WHO, 2000a). This approach has been used subsequently by JECFA in several evaluations of residues of veterinary antimicrobial drugs (FAO/WHO, 2001a, 2002b, 2004).

At its fifty-second meeting (FAO/WHO, 2000a), JECFA proposed a comprehensive decision tree that takes account of all relevant data from model in vitro and in vivo test systems and includes minimum inhibitory concentrations (MICs) when setting an ADI. Similar approaches have been subsequently developed and used by several regulatory authorities. In the interest of harmonization of methods, VICH published a guideline entitled Studies to Evaluate the Safety of Residues of Veterinary Drugs in Human Food: General Approach to Establish a Microbiological ADI (VICH, 2004). This VICH guideline is a refinement of the JECFA approach. In recognition of the importance of international harmonization, JECFA, at its sixty-sixth meeting (FAO/WHO, 2006), agreed to incorporate the VICH guideline in future assessments to ensure consistency and transparency in the determination of microbiological ADIs.

A summary of the recommendations is given below:

1. Additional microbiological data are not required if there is evidence that:
   - the veterinary drug and its residues do not have antimicrobial properties, and/or
   - ingested residues do not enter the lower bowel, and/or
   - the ingested residues are transformed to inactive metabolites before entering the lower bowel, and/or
   - the ingested residues are transformed quantitatively to microbiologically inactive metabolites, and/or
   - data on the effects of the veterinary drug on gastrointestinal microflora in vitro and in vivo provide a basis for concluding that the ADI derived from toxicological data is sufficiently low to protect the intestinal microflora, and/or
   - clinical data show that the incidence of toxicological effects after therapeutic use of the drug in humans is substantially higher than that of any gastrointestinal side-effects due to the disruption of the microflora.

2. If none of the above can be demonstrated, additional studies were proposed for establishing an ADI (for detailed guidance, see FAO/WHO, 2000a):
   - The class of drug should be considered in order to determine whether the main concern is the emergence of resistance or

the disruption of the intestinal microflora. If effects on the barrier to colonization are a concern, the MIC of the veterinary drug against bacterial strains representative of relevant genera of the microflora in the gastrointestinal tract of healthy individuals can be used as the basis for a conservative estimate of the ADI.

– If disruption of the barrier to colonization is the concern and data are not available, information should be provided to show either that addition of the veterinary drug at concentrations covering the range expected in the colon from an ADI based on other effects does not cause disruption of the barrier to colonization or that oral administration of the veterinary drug to a monogastric animal (e.g. rat, mouse or other rodent inoculated with human flora), at a dose that would result in the concentrations expected in the human colon at an ADI, shows no effect on the barrier to colonization.
– If emergence of antimicrobial resistance due to consumption of residues is the concern, data to show that the expected residue concentrations in the colon do not change the antibiotic resistance or resident populations of *Escherichia coli* or other bacteria appropriate for the drug class should be provided.
– If the concern is change in a specific enzymatic activity that is directly linked to adverse effects on human health, in vitro or in vivo tests should be conducted to determine the concentration of the drug that does not alter that specific enzymatic activity.

4.13 References[1]

Adler I-D & Ashby J (1989) The present lack of evidence for unique rodent germ-cell mutagens. Mutat Res, **212**: 55–66.

Allgood GS, Kuter DJ, Roll KT, Taylor SL & Zorich NL (2001) Postmarketing surveillance of new food ingredients: results from the program with the fat replacer olestra. Regul Toxicol Pharmacol, **33**: 224–233.

[1] Internet links provided in these references were active as of the date of final editing.

Ames BN, Durston WE, Yamasaki E & Lee FD (1973) Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. Proc Natl Acad Sci USA, **70**: 2281–2285.

Ames BN, McCann J & Yamasaki E (1975) Methods for detecting carcinogens and mutagens with the *Salmonella*/microsome test: assay of 300 chemicals. Mutat Res, **31**: 347–364.

Arni P, Ashby J, Castellino S, Engelhardt G, Herbold BA, Priston RAJ & Bontinck WJ (1988) Assessment of the potential germ cell mutagenicity of industrial and plant protection chemicals as part of an integrated study of genotoxicity in vitro and in vivo. Mutat Res, **203**: 177–184.

Asero R, Mistrello G, Roncarolo D, de Vries SC, Gautier MF, Ciurana LF, Verbeek E, Mohammadi T, Knul-Brettlova V, Akkerdaas JH, Bulder I, Aalberse RC & Van Ree R (2000) Lipid transfer protein: a pan-allergen in plant-derived foods that is highly resistant to pepsin digestion. Int Arch Allergy Appl Immunol, **122**: 20–32.

Asero R, Ballmer-Weber BK, Beyer K, Conti A, Dubakiene R, Fernandez-Rivas M, Hoffmann-Sommergruber K, Lidholm J, Mustakov T, Oude Elberink JNG, Pumphrey RSH, Skov PS, Van Ree R, Vlieg-Boerstra BJ, Hiller R, Hourihane JO, Kowalski M, Papadopoulos NG, Wal J-M, Mills ENC & Vieths S (2007) IgE-mediated food allergy diagnosis: current status and new perspectives. Mol Nutr Food Res, **51**: 135–147.

Ashby J & Tennant RW (1991) Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the US NTP. Mutat Res, **257**: 229–306.

Astwood JD, Leach JN & Fuchs RL (1996) Stability of food allergens to digestion in vitro. Nat Biotechnol, **14**: 1269–1273.

ATBC (1994) The effect of vitamin E and beta-carotene on the incidence of lung cancer and other cancers in male smokers. The Alpha-Tocopherol Beta-Carotene (ATBC) Cancer Prevention Study Group. New Engl J Med, **330**: 1029–1035.

Auerbach C, Robson JM & Carr JG (1947) The chemical production of mutations. Science, **105**(2713): 243–247.

Auletta A & Ashby J (1988) Workshop on the relationship between short-term test information and carcinogenicity: Williamsburg, Virginia, January 20–23, 1987. Environ Mol Mutagen, **11**: 135–145.

Ballmer-Weber BK, Holzhauser T, Scibilia J, Mittag D, Zisa G, Ortolani C, Oesterballe M, Poulsen LK, Vieths S & Bindslev-Jensen C (2007) Clinical characteristics of soybean allergy in Europe: a double-blind, placebo-controlled food challenge study. J Allergy Clin Immunol, **119**(6): 1489–1496.

Barlow SM, Greig JB, Bridges JW, Carere A, Carpy AJ, Galli CL, Kleiner J, Knudsen I, Koeter HB, Levy LS, Madsen C, Mayer S, Narbonne JF, Pfannkuch F, Prodanchuk MG, Smith MR & Steinberg P (2002) Hazard identification by methods of animal-based toxicology. Food Chem Toxicol, **40**: 145–191.

Basketter DA, Evans P, Fielder RJ, Gerberick GF, Dearman RJ & Kimber I (2002) Local lymph node assay—validation, conduct and use in practice. Food Chem Toxicol, **40**(5): 593–598.

Becker WM & Reese G (2001) Immunological identification and characterization of individual food allergens. J Chromatogr B, **756**: 131–140.

Bindslev-Jensen C, Ballmer-Weber BK, Bengtsson U, Blanco C, Ebner C, Hourihane J, Knulst AC, Moneret-Vautrin DA, Nekam K, Niggemann B, Osterballe M, Ortolani C, Ring J, Schnopp C & Werfel T (2004) Standardization of food challenges in patients with immediate reactions to foods—Position paper from the European Academy of Allergology and Clinical Immunology. Allergy, **59**: 690–697.

Bock SA, Sampson HA, Atkins FM, Zeigler RS, Lehrer S, Sachs M, Bush RK & Metcalfe DD (1988) Double-blind, placebo-controlled food challenge (DBPCFC) as an office procedure: a manual. J Allergy Clin Immunol, **82**: 986–997.

Bøgh KL, Kroghsbo S, Dahl L, Barkholt V, Rigby NM, Mills ENC & Madsen CB (2008) Digested Ara h 1 retains its sensitising capacity in Brown Norway rats. Presented at the Nordic Symposium on Allergy, Sønderborg (abstract).

Bonita R, Beaglehole R & Kjellström T (2006) Basic epidemiology, 2nd ed. Geneva, World Health Organization.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, 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**: 781–792.

Boyano Martínez T, García-Ara C, Diaz-Pena JM, Muñoz FM, García Sánchez G & Esteban MM (2001) Validity of specific IgE antibodies in children with egg allergy. Clin Exp Allergy, **31**: 1464–1469.

Bradford Hill A (1965) The environment and disease: association or causation? Proc R Soc Med, **58**: 295–300.

Brandtzaeg P (2002) Current understanding of gastrointestinal immunoregulation and its relation to food allergy. Ann NY Acad Sci, **964**: 13–45.

Bredehorst R & David K (2001) What establishes a protein as an allergen? J Chromatogr B, **756**: 33–40.

Breiteneder H & Ebner C (2000) Molecular and biochemical classification of plant-derived food allergens. J Allergy Clin Immunol, **106**: 27–36.

Breiteneder H & Mills ENC (2005) Molecular properties of food allergens. J Allergy Clin Immunol, **115**: 14–23.

Broxup B (1991) Neuropathology as a screen for neurotoxicity assessment. J Am Coll Toxicol, **10**: 689–695.

Buchanan R, Dennis S, Gendel S, Acheson D, Assimon SA, Beru N, Bolger P, Carlson D, Carvajal R, Copp C, Falci K, Garber E, Harden E, Kane R, Kvenberg J, Luccioli S, Park D, Raybourne R, Troxell T & Vierk K; Threshold Working Group (2008) Approaches to establish thresholds for major food allergens and for gluten in food. J Food Prot, **71**: 1043–1088.

Buelke-Sam J, Kimmel CA & Adams J eds (1985) Design considerations in screening for behavioural teratogens: results of the collaborative behavioural teratology study. Neurobehav Toxicol Teratol, **7**(6): 537–789.

Burchiel SW, Kerkvliet NL, Gerberick GF, Lawrence DA & Ladics GS (1997) Assessment of immunotoxicity by multiparameter flow cytometry. Fundam Appl Toxicol, **38**(1): 38–54.

Burleson GR, Dean JH & Munson AE (1995) Methods in immunotoxicology. New York, NY, Wiley-Liss.

Bushnell P, Padilla S, Ward T, Pope C & Olszyk V (1991) Behavioral and neurochemical changes in rats dosed repeatedly with diisopropylfluorophosphate (DFP). J Pharmacol Exp Ther, **256**: 741–750.

Capen CC, Dybing E, Rice JM & Wilbourn JD (1999) Species differences in thyroid, kidney and urinary bladder carcinogenesis. Lyon, International Agency for Research on Cancer (IARC Scientific Publications, No. 147).

Catalano PJ, McDaniel KL & Moser VC (1997) The IPCS Collaborative Study on Neurobehavioral Screening Methods. VI. Agreement and reliability of the data. Neurotoxicology, **18**: 1057–1064.

Catassi C, Fabiani E, Iacono G, D'Agate C, Francavilla R, Biagi F, Volta U, Accomando S, Picarelli A, De Vitis I, Pianelli G, Gesuita R, Carle F, Mandolesi A, Bearzi I & Fasano A (2007) A prospective, double-blind, placebo-controlled trial to establish a safe gluten threshold for patients with celiac disease. Am J Clin Nutr, **85**(1): 160–166.

Chang LW & Slikker W Jr eds (1995) Neurotoxicology: approaches and methods. San Diego, CA, Academic Press.

Chapin RE & Sloane RA (1997) Reproductive assessment by continuous breeding: evolving study design and summaries of ninety studies. Environ Health Perspect, **105**(Suppl 1): 199–205.

Chernoff N, Roger JM & Kavlock RJ (1989) An overview of maternal toxicity and prenatal development: considerations for developmental toxicity hazard assessments. Toxicology, **59**: 111–125.

Chung K, Baker JR Jr, Baldwin JL & Chou A (2001) Identification of carmine allergens among three carmine allergy patients. Allergy, **56**(1): 73–77.

Coggon D, Rose G & Barker DJM (1997) Epidemiology for the uninitiated, 4th ed. London, BMJ Publishing Group (http://www.bmj.com/epidem/epid.html).

Cohen SM, Robinson D & MacDonald J (2001) Alternative models for carcinogenicity testing. Toxicol Sci, **64**(1): 14–19.

Coligan JE, Bierer BE, Margulies DH, Shevach EM, Strober W & Coico R (2005) Current protocols in immunology. Winston-Salem, NC, John Wiley & Sons.

Cooper JA, Saracci R & Cole P (1979) Describing the validity of carcinogen screening tests. Br J Cancer, **39**: 87–89.

Cooper RL, Lamb JC, Barlow SM, Bentley K, Brady AM, Doerrer NG, Eisenbrandt DL, Fenner-Crisp PA, Hines RN, Irvine LFH, Kimmel CA, Koeter H, Li AA, Makris SL, Sheets LP, Speijers GJA & Whitby KE (2006) A tiered approach to life stages testing for agricultural chemical safety assessment. Crit Rev Toxicol, **36**: 69–98.

Cory-Slechta DA, Crofton KM, Foran JA, Ross JF, Sheets LP, Weiss B & Mileson B (2001) Methods to identify and characterize developmental neurotoxicity for human health risk assessment. I: Behavioral effects. Environ Health Perspect, **109**: 79–91.

Crevel RWR, Briggs D, Hefle SL, Knulst AC & Taylor SL (2007) Hazard characterisation in food allergen risk assessment: the application of statistical approaches and the use of clinical data. Food Chem Toxicol, **45**: 691–701.

Crevel RW, Ballmer-Weber BK, Holzhauser T, Hourihane JO, Knulst AC, Mackie AR, Timmermans F & Taylor SL (2008) Thresholds for food allergens and their value to different stakeholders. Allergy, **63**: 597–609.

Crow JF (1968) Rate of genetic change under selection. Proc Natl Acad Sci USA, **59**: 655–668.

Dalal I, Binson I, Reifen R, Amitai Z, Shohat T, Rahmani S, Levine A, Ballin A & Somekh E (2002) Food allergy is a matter of geography after all: sesame as a major cause of severe IgE-mediated food allergic reactions among infants and young children in Israel. Allergy, **57**(4): 362–365.

Damstra T, Barlow S, Bergman A, Kavlock R & Van Der Kraak G eds (2002) Global assessment of the state-of-the-science of endocrine disruptors. Geneva, World Health Organization, International Programme on Chemical Safety, 180 pp (WHO/IPCS/EDC/02.2; http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/).

Daston GP (1994) Relationships between maternal and developmental toxicity. In: Kimmel CA & Buelke-Sam J eds. Developmental toxicology, 2nd ed. New York, NY, Raven Press, pp 189–214.

Dean JH, Twerdok LE, Tice RR, Sailstad DM, Hattan DG & Stokes WS (2001) ICCVAM evaluation of the murine local lymph node assay. Conclusions and recommendations of an independent scientific peer review panel. Regul Toxicol Pharmacol, **34**(3): 258–273.

Descotes J (2003) From clinical to human toxicology: linking animal research and risk assessment in man. Toxicol Lett, **140–141**: 3–10.

Deuel RK (1977) Determining sensory deficits in animals. In: Myers R ed. Methods in psychobiology. New York, NY, Academic Press, pp 99–125.

Donehower LA, Harvey M, Slagle BL, McArthur MJ, Montgomery CA Jr, Butel JS & Bradley A (1992) Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature, **356**: 215–221.

Dorman DC, Allen SL, Byczkowski JZ, Claudio L, Fisher JE, Fisher JW, Harry GJ, Li AA, Makris SL, Padilla S, Sultatos LG & Mileson BE (2001) Methods to identify and characterize developmental neurotoxicity for human health risk assessment. III: Pharmacokinetic and pharmacodynamic considerations. Environ Health Perspect, **109**: 101–111.

Duffus JH & Worth HGJ (2006) Fundamental toxicology. London, Royal Society of Chemistry.

Duggan C, Gannon J & Walker WA (2002) Protective nutrients and functional foods for the gastrointestinal tract. Am J Clin Nutr, **75**: 789–908.

Dybing E, Doe J, Groten J, Kleiner J, O'Brien J, Renwick AG, Schlatter J, Steinberg P, Tritscher A, Walker R & Younes M (2002) Hazard characterization of chemicals in food and diet: dose response, mechanisms and extrapolation issues. Food Chem Toxicol, **40**: 237–282.

Ebo DG & Stevens WJ (2001) IgE-mediated food allergy—Extensive review of the literature. Acta Clin Belg, **56**(4): 234–247.

EC (1996) Technical guidance document in support of Commission Directive 93/67/EEC on risk assessment for new notified substances and Commission Regulation EC/1488/94 on risk assessment for existing substances. Luxembourg, European Commission.

EC (1997) Criteria for the qualitative evaluation of human neurobehavioral studies of neurotoxicity. Luxembourg, European Commission, Office for Official Publications of the European Communities.

EC (1999) Report of the Working Group on Endocrine Disrupters of the Scientific Committee on Toxicity, Ecotoxicity and the Environment, Directorate-General for Consumer Policy and Consumer Health Protection. Brussels, European Commission.

EC (2006) European Union risk assessment report on phenol. CAS No. 108-95-2. EINECS No. 203-632-7. 1st Priority List, Volume 64. Revised edition. EUR 22522 EN/1. Luxembourg, European Commission, Directorate-General Joint Research Centre, European Chemicals Bureau; Office for Official Publications of the European Communities (http://ecb.jrc.it/DOCUMENTS/Existing-Chemicals/RISK_ASSESSMENT/REPORT/phenolreport060.pdf).

ECETOC (1992) Evaluation of the neurotoxic potential of chemicals. Brussels, European Centre for Ecotoxicology and Toxicology of Chemicals (Monograph No. 18).

ECETOC (1998) Organophosphorous pesticides and long-term effects on the nervous system. Brussels, European Centre for Ecotoxicology and Toxicology of Chemicals (Technical Report No. 75).

EDSTAC (1998) The Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) final report. Washington, DC, United States Environmental Protection Agency (http://www.epa.gov/endo/pubs/edspoverview/finalrpt.htm).

EFSA (2004) Opinion of the Scientific Panel on Dietetic Products, Nutrition and Allergies on a request from the Commission relating to the evaluation of allergenic foods for labelling purposes. EFSA J, **32**: 1–197.

EFSA (2009) Draft scientific opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. Parma, European Food Safety Authority, Panel on Genetically Modified Organisms (http://www.efsa.europa.eu/en/consultations/call/gmo091201-ax1.pdf).

Eisenbrandt DL, Allen SL, Berry PH, Classen W, Bury D, Mellert W, Millischer RJ, Schuh W & Bontinck WJ (1994) Evaluation of the neurotoxic potential of chemicals in animals. Food Chem Toxicol, **32**: 655–669.

Eiwegger T, Rigby N, Mondoulet L, Bernard H, Krauth MT, Boehm A, Dehlink E, Valent P, Wal JM, Mills ENC & Szepfalusi Z (2006) Gastro-duodenal digestion products of the major peanut allergen Ara h 1 retain an allergenic potential. Clin Exp Allergy, **36**: 1281–1288.

Enzmann H, Bomhard E, Iatropoulos M, Ahr HJ, Schlueter G & Williams GM (1998a) Short- and intermediate-term carcinogenicity testing—a review. Part 1: The prototypes mouse skin tumour assay and rat liver focus assay. Food Chem Toxicol, **36**: 979–995.

Enzmann H, Iatropoulos M, Brunnemann KD, Bomhard E, Ahr HJ, Schlueter G & Williams GM (1998b) Short- and intermediate-term carcinogenicity testing—a review. Part 2: Available experimental models. Food Chem Toxicol, **36**: 997–1013.

Faccini JM, Butler WR, Friedmann JC, Hess R, Reznik GK, Ito N, Hayashi Y & Williams GM (1992) IFSTP guidelines for the design and interpretation of the chronic rodent carcinogenicity bioassay. Exp Toxicol Pathol, **44**: 443–456.

FAO (1995) Food allergies. Report of the FAO Technical Consultation on Food Allergies, Rome, 13–14 November 1995. Rome, Food and Agriculture Organization of the United Nations.

FAO/WHO (1971) Evaluation of food additives: specifications for the identity and purity of food additives and their toxicological evaluation: some extraction solvents and certain other substances; and a review of the technological efficacy of some antimicrobial agents. Fourteenth report of the Joint FAO/WHO Expert Committee on Food Additives. Rome, Food and Agriculture Organization of the United Nations; Geneva, World Health Organization (FAO Nutrition Meetings Series, No. 48; WHO Technical Report Series, No. 462; http://whqlibdoc.who.int/trs/WHO_TRS_462.pdf).

FAO/WHO (1995) Evaluation of certain food additives and contaminants. Forty-fourth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 859; http://whqlibdoc.who.int/trs/WHO_TRS_859.pdf).

FAO/WHO (1997) Evaluation of certain food additives and contaminants. Forty-sixth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 868; http://whqlibdoc.who.int/trs/WHO_TRS_868.pdf).

FAO/WHO (1998) Abamectin (addendum). In: Pesticide residues in food—1997 evaluations. Geneva, World Health Organization (WHO/PCS/98.6; http://www.inchem.org/documents/jmpr/jmpmono/v097pr02.htm).

FAO/WHO (1999a) Procedures for estimating acute reference dose. In: Pesticide residues in food—1998. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome, Food and Agriculture Organization of the United Nations, pp 14–17 (FAO Plant Production and Protection Paper, No. 148).

FAO/WHO (1999b) Evaluation of certain veterinary drug residues in food. Fiftieth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 888; http://whqlibdoc.who.int/trs/WHO_TRS_888.pdf).

FAO/WHO (2000a) Evaluation of certain veterinary drug residues in food. Fifty-second report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 893; http://whqlibdoc.who.int/trs/WHO_TRS_893.pdf).

FAO/WHO (2000b) Evaluation of certain food additives and contaminants. Fifty-third report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 896; http://whqlibdoc.who.int/trs/WHO_TRS_896.pdf).

FAO/WHO (2001a) Evaluation of certain veterinary drug residues in food. Fifty-fourth report of the Joint FAO/WHO Meeting on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 900; http://whqlibdoc.who.int/trs/WHO_TRS_900.pdf).

FAO/WHO (2001b) Evaluation of allergenicity of genetically modified foods. Report of a Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology, Rome, 22–25 January 2001. Geneva, World Health Organization (http://www.who.int/foodsafety/publications/biotech/ec_jan2001/en/index.html).

FAO/WHO (2002a) Pesticide residues in food—2002. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome, Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 172; http://www.fao.org/ag/AGP/AGPP/Pesticid/JMPR/Download/2002_rep/2002JMPRReport2.pdf).

FAO/WHO (2002b) Evaluation of certain veterinary drug residues in food. Fifty-eighth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 911; http://whqlibdoc.who.int/trs/WHO_TRS_911.pdf).

FAO/WHO (2003) Appendix III, Draft guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants; Appendix IV, Proposed draft annex on the assessment of possible allergenicity of the draft guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants. In: Report of the third session of the Codex Ad Hoc Intergovernmental Task Force on Foods Derived from Biotechnology, Yokohama, 4–8 March 2002. Rome, Food and Agriculture Organization of the United Nations, Codex Alimentarius Commission, pp 47–60 (ALINORM 03/34; http://www.codexalimentarius.net/web/archives.jsp?year=03).

FAO/WHO (2004) Evaluation of certain veterinary drug residues in food. Sixty-second report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 925; http://whqlibdoc.who.int/trs/WHO_TRS_925.pdf).

FAO/WHO (2006) Evaluation of certain veterinary drug residues in food. Sixty-sixth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva, World Health Organization (WHO Technical Report Series, No. 939; http://whqlibdoc.who.int/publications/2006/9241209399_eng.pdf).

FAO/WHO (2008) Appendix III, Draft revised Codex standard for special dietary use for persons intolerant to gluten. In: Report of the twenty-ninth session of the Codex Committee on Nutrition and Foods for Special Dietary Purposes, Bad Neuenahr-Ahrweiler, 12–16 November 2007. Rome, Food and Agriculture Organization of the United Nations, Codex Alimentarius Commission (ALINORM 08/31/26; ftp://ftp.fao.org/codex/Alinorm08/al31_26e.pdf).

Fernandez-Rivas M, Bolhaar S, Gonzalez-Mancebo E, Asero R, Van Leeuwen A, Bohle B, Ma Y, Ebner C, Rigby N, Sancho AI, Miles S, Zuidmeer L, Knulst A, Breiteneder H, Mills C, Hoffmann-Sommergruber K & Van Ree R (2006) Apple allergy across Europe: how allergen sensitization profiles determine the clinical expression of allergies to plant foods. J Allergy Clin Immunol, **118**: 481–488.

Flammang TJ, Tungeln LS, Kadlubar FF & Fu PP (1997) Neonatal mouse assay for tumorigenicity: alternative to the chronic rodent bioassay. Regul Toxicol Pharmacol, **26**: 230–240.

Flinterman AE, Pasmans SG, Hoekstra MO, Meijer Y, Hoffen EV, Knol EF, Hefle SL, Bruijnzeel-Koomen CA & Knulst AC (2006) Determination of no-observed-adverse-effect levels and eliciting doses in a representative group of peanut-sensitized children. J Allergy Clin Immunol, **117**: 448–454.

Foote RH & Carney EW (2000) The rabbit as a model for reproductive and developmental toxicity studies. Reprod Toxicol, **14**: 77–93.

French J, Storer RD & Donehower LA (2001) The nature of the heterozygous Trp53 knockout model for identification of mutagenic carcinogens. Toxicol Pathol, **29** (Suppl): 24–29.

FSA (2002) Guidelines on the use of human studies in the pre-market safety assessment of novel foods. London, United Kingdom Food Standards Agency, Advisory Committee on Novel Foods and Processes (http://www.acnfp.gov.uk/acnfppapers/inforelatass/guidetastehuman/guidlinespre-market).

Fu TT, Abbott UR & Hatzos C (2002) Digestibility of food allergens and non-allergenic proteins in simulated gastric fluid and simulated intestinal fluid—A comparative study. J Agric Food Chem, **50**: 7154–7160.

Gad SC (1982) A neuromuscular screen for use in industrial toxicology. J Toxicol Environ Health, **9**: 691–704.

Gad SC (1989) Principles of screening in toxicology with special emphasis on applications to neurotoxicology. J Am Coll Toxicol, **8**: 21–27.

Garman RH, Fix AS, Jortner BS, Jensen KF, Hardisty JF, Claudio L & Ferenc S (2001) Methods to identify and characterize developmental neurotoxicity for human health risk assessment. II: Neuropathology. Environ Health Perspect, **109**: 93–100.

Gatzidou ET, Zira AN & Theocharis SE (2007) Toxicogenomics: a pivotal piece in the puzzle of toxicological research. J Appl Toxicol, **27**: 302–309.

Gerberick GF, Ryan CA, Dearman RJ & Kimber I (2007) Local lymph node assay (LLNA) for detection of sensitization capacity of chemicals. Methods, **41**(1): 54–60.

Germolec DR, Nyska A, Kashon M, Kuper CF, Portier C, Kommineni C, Johnson KA & Luster MI (2004a) Extended histopathology in immunotoxicity testing: interlaboratory validation studies. Toxicol Sci, **78**(1): 107–115.

Germolec DR, Kashon M, Nyska A, Kuper CF, Portier C, Kommineni C, Johnson KA & Luster MI (2004b) The accuracy of extended histopathology to detect immunotoxic chemicals. Toxicol Sci, **82**(2): 504–514.

Gold LS & Zeiger E (1996) Handbook of carcinogenic potency and genotoxicity data bases. Boca Raton, FL, CRC Press.

Goodman JI (2001) A perspective on current and future uses of alternative models for carcinogenicity testing. Toxicol Pathol, **29**(Suppl): 173–176.

Goodman RE, Vieths S, Sampson HA, Hill D, Ebisawa M, Taylor SL & Ree R (2008) Allergenicity assessment of genetically modified crops—what makes sense? Nat Biotechnol, **26**: 74–81.

Gray LE Jr, Ostby J, Wilson V, Lambright C, Bobseine K, Hartig P, Hotchkiss A, Wolf C, Furr J, Price M, Parks L, Cooper RL, Stoker TE, Laws SC, Degitz SJ, Jensen KM, Kahl MD, Korte JJ, Makynen EA, Tietge JE & Ankley GT (2002) Xenoendocrine disrupters—tiered screening and testing: filling key data gaps. Toxicology, **181–182**: 371–382.

Griffin JW (1990) Basic pathologic processes in the nervous system. Toxicol Pathol, **18**: 83–88.

Gulezian D, Jacobson-Kram D, McCullough CB, Olson HM, Recio L, Robinson DE, Storer RD, Tennant RW, Ward JM & Neumann DA (2000) Use of transgenic animals for carcinogenicity testing: consideration and implications for risk assessment. Toxicol Pathol, **28**: 482–499.

Gundert-Remy U, Dahl SG, Boobis A, Kremers P, Kopp-Schneider A, Oberemm A, Renwick A & Pelkonen O (2005) Molecular approaches to the identification of biomarkers of exposure and effect—report of an expert meeting organized by COST Action B15. November 28, 2003. Toxicol Lett, **156**(2): 227–240.

Haley P, Perry R, Ennulat D, Frame S, Johnson C, Lapointe JM, Nyska A, Snyder P, Walker D, Walter G; STP Immunotoxicology Working Group (2005) Best practice guideline for the routine pathology evaluation of the immune system. Toxicol Pathol, **33**(3): 404–406 (http://www.toxpath.org/Position_Papers/Immune_System.pdf).

Harleman JH (2000) Approaches to the identification and recording of findings in the lymphoreticular organs indicative for immunotoxicity in regulatory type toxicity studies. Toxicology, **142**(3): 213–219.

Harris MW, Chapin RE, Lockhart AC & Jokinen MP (1992) Assessment of short-term reproductive and development toxicity screen. Fundam Appl Toxicol, **19**: 186–196.

Harry GJ (1998) Developmental neurotoxicity. In: Korach KS ed. Reproductive and developmental toxicology. New York, NY, Marcel Dekker, pp 211–249.

Health Canada (1994) Human health risk assessment for priority substances. Ottawa, Ontario, Health Canada, Environmental Health Directorate (Publication No. En40-215/41E).

Hefle SL, Nordlee JA & Taylor SL (1996) Allergenic foods. Crit Rev Food Sci Nutr, **36**(Suppl): S69–S89.

Heilmann C, Grandjean P, Weihe P, Nielsen F & Budtz-Jørgensen E (2006) Reduced antibody responses to vaccinations in children exposed to polychlorinated biphenyls. PLoS Med, **3**(8): 1352–1359.

Heinje WH, Keinhuis AS, Van Ommen B, Stierum RH & Groten JP (2005) Systems toxicology: applications of toxicogenomics, transcriptomics, proteomics and metabolomics in toxicology. Expert Rev Proteomics, **2**: 767–780.

Hertel RF (1996) Outline on risk assessment of existing substances in the European Union. Environ Toxicol Pharmacol, **2**: 93–96.

Hill DJ, Hosking CS, Zhie CY, Leung R, Baratwidjaja K, Iikura Y, Iyngkaran N, Gonzalez-Andaya A, Wah LB & Hsieh KH (1997) The frequency of food allergy in Australia and Asia. Environ Toxicol Pharmacol, **4**: 101–110.

Hill RN, Crisp TM, Hurley PM, Rosenthal SL & Singh DV (1998) Risk assessment of thyroid follicular cell tumors. Environ Health Perspect, **106**: 447–457.

Hischenhuber C, Crevel R, Jarry B, Mäki M, Moneret-Vautrin DA, Romano A, Troncone R & Ward R (2006) Review article: Safe amounts of gluten for patients with wheat allergy or coeliac disease. Aliment Pharmacol Ther, **23**(5): 559–575.

Holden HE (1982) Comparison of somatic and germ cell models for cytogenetic screening. J Appl Toxicol, **2**: 196–200.

Holladay SD & Luster MI (1994) Developmental immunotoxicology. In: Kimmel CA & Buelke-Sam J eds. Developmental toxicology, 2nd ed. New York, NY, Raven Press, pp 93–118.

Hood R (2006) Developmental and reproductive toxicology—A practical approach. Boca Raton, FL, CRC Taylor & Francis.

Host A, Halken S, Jacobsen HP, Christensen AE, Herskind AM & Plesner K (2002) Clinical course of cow's milk protein allergy/intolerance and atopic diseases in childhood. Pediatr Allergy Immunol, **13**(Suppl 15): 23–28.

Hourihane JO (2002) Recent advances in peanut allergy. Curr Opin Allergy Clin Immunol, **2**: 227–231.

Hourihane JO, Kilburn SA, Nordlee JA, Hefle SL, Taylor SL & Warner JO (1997) An evaluation of sensitivity of subjects with peanut allergy to very low doses of peanut protein: a randomized, double-blind, placebo-controlled food challenge study. J Allergy Clin Immunol, **100**: 596–600.

Huby RDJ, Dearman RJ & Kimber I (2000) Why are some proteins allergens? Toxicol Sci, **55**: 235–246.

IARC (1995) Peroxisome proliferation and its role in carcinogenesis. Views and expert opinions of an IARC working group. Lyon, International Agency for Research on Cancer (IARC Technical Report No. 24).

Iatropoulos MJ & Williams GM (1996) Proliferation markers. Exp Toxicol Pathol, **48**: 175–181.

ICH (1994a) Note for guidance on toxicokinetics: The assessment of systemic exposure in toxicity studies. S3A. Geneva, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH Harmonised Tripartite Guideline; http://www.ich.org/LOB/media/MEDIA495.pdf).

ICH (1994b) Pharmacokinetics: Guidance for repeated dose tissue distribution studies. S3B. Geneva, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH Harmonised Tripartite Guideline; http://www.ich.org/LOB/media/MEDIA496.pdf).

ICH (1994c) Guideline for industry: Detection of toxicity to reproduction for medicinal products. ICH-S5A. Geneva, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (http://www.fda.gov/downloads/Drugs/GuidanceCompliance RegulatoryInformation/Guidances/ucm074950.pdf).

ICH (1997) Testing for carcinogenicity of pharmaceuticals. S1B. Geneva, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH Harmonised Tripartite Guideline; http://www.ich.org/LOB/media/MEDIA490.pdf).

ICH (2005) Immunotoxicity studies for human pharmaceuticals. S8. Geneva, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH Harmonised Tripartite Guideline; http://www.ich.org/LOB/media/MEDIA1706.pdf).

ICICIS Group Investigators (1998) Report of validation study of assessment of direct immunotoxicity in the rat. Toxicology, **125**(2–3): 183–201.

ICME (1994) A guide to neurotoxicity: international regulation and testing strategies. Ottawa, Ontario, International Council on Metals and the Environment.

ILSI (2001) ILSI/HESI alternatives to carcinogenicity testing project. Washington, DC, International Life Sciences Institute. Toxicol Pathol, **29**(Suppl): 1–322.

Immamura T, Kanagawa Y & Ebisawa M (2008) A survey of patients with self-reported severe food allergies in Japan. Pediatr Allergy Immunol, **19**: 270–274.

IPCS (1986a) Principles of toxicokinetic studies. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 57; http://www.inchem.org/documents/ehc/ehc/ehc57.htm).

IPCS (1986b) Principles and methods for the assessment of neurotoxicity associated with exposure to chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 60; http://www.inchem.org/documents/ehc/ehc/ehc060.htm).

IPCS (1987) Principles for the safety assessment of food additives and contaminants in food. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 70; http://www.inchem.org/documents/ehc/ehc/ehc70.htm).

IPCS (1990) Principles for the toxicological assessment of pesticide residues in food. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 104; http://www.inchem.org/documents/ehc/ehc/ehc104.htm).

IPCS (1996) Principles and methods for assessing direct immunotoxicity associated with exposure to chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 180; http://www.inchem.org/documents/ehc/ehc/ehc180.htm).

IPCS (1999) Principles and methods for assessing allergic hypersensitization associated with exposure to chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 212; http://www.inchem.org/documents/ehc/ehc/ehc212.htm).

IPCS (2001a) Neurotoxicity risk assessment for human health: principles and approaches. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 223; http://www.inchem.org/documents/ehc/ehc/ehc223.htm).

IPCS (2001b) Principles for evaluating health risks to reproduction associated with exposure to chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 225; http://www.inchem.org/documents/ehc/ehc/ehc225.htm).

IPCS (2001c) Biomarkers in risk assessment: validity and validation. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 222; http://www.inchem.org/documents/ehc/ehc/ehc222.htm).

IPCS (2003) Toxicogenomics and the risk assessment of chemicals for the protection of human health. Workshop report. Summary. Geneva, World Health Organization (IPCS/Toxicogenomics/03/1; http://www.who.int/ipcs/methods/en/toxicogenomicssummaryreport.pdf).

IPCS (2004) IPCS risk assessment terminology. Geneva, World Health Organization, International Programme on Chemical Safety (Harmonization Project Document, No. 1; http://www.who.int/ipcs/methods/harmonization/areas/ipcsterminologyparts1and2.pdf).

IPCS (2005) Chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration–response assessment (Harmonization Project Document, No. 2; http://whqlibdoc.who.int/publications/2005/9241546786_eng.pdf).

IPCS (2006a) Principles and methods for assessing autoimmunity associated with exposure to chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 236; http://www.who.int/ipcs/publications/ehc/ehc236.pdf).

IPCS (2006b) Principles for evaluating health risks in children associated with exposure to chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 237; http://www.who.int/ipcs/publications/ehc/ehc237.pdf).

IPCS (2009) Principles for modelling dose–response for the risk assessment of chemicals. Geneva, World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 239; http://whqlibdoc.who.int/publications/2009/9789241572392_eng.pdf).

Isaacson RL & Jensen KF (1992a) The vulnerable brain and environmental risks. Vol. 1. Malnutrition and hazard assessment. New York, NY, Plenum Press.

Isaacson RL & Jensen KF (1992b) The vulnerable brain and environmental risks. Vol. 2. Toxins in food. New York, NY, Plenum Press.

Ivarsson A, Persson LA & Nystrom L (2000) Epidemic of celiac disease in Swedish children. Acta Paediatr, **89**: 165–171.

Jacobson-Kram D & Keller KA eds (2006) Toxicological testing handbook: Principles, applications, and data interpretation, 2nd ed. New York, NY, Informa Healthcare USA, Inc.

Jenkins JA, Griffiths-Jones S, Shewry PR, Breiteneder H & Mills ENC (2005) Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis. J Allergy Clin Immunol, **115**: 163–170.

Jenkins JA, Breiteneder H & Mills EN (2007) Evolutionary distance from human homologs reflects allergenicity of animal food proteins. J Allergy Clin Immunol, **120**: 1399–1405.

Johansson SGO, Hourihane JO, Bousquet J, Bruijnzeel-Koomen C, Dreborg S, Haahtela T, Kowalski ML, Mygind N, Ring J, Van Cauwenberge P, Van Hage-Hamsten M, Wüthrich B; EAACI (the European Academy of Allergology and Clinical Immunology) nomenclature task force (2001) A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force. Allergy, **56**: 813–824 [cited in Asero et al., 2007].

Johansson SGO, Bieber T, Dahl R, Friedmann PS, Lanier BQ, Lockey RF, Motala C, Ortega Martell JA, Platts-Mills TA, Ring J, Thien F, Van Cauwenberge

P & Williams HC (2004) Revised nomenclature for allergy for global use: report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. J Allergy Clin Immunol, **113**(5): 832–836.

Johnsen H, Lund SP, Matikainen E, Midtgard U, Simonsen L & Wennberg A (1992) Occupational neurotoxicity: criteria document for evaluation of existing data. Copenhagen, Nordic Council of Ministers and National Institute of Occupational Health.

Johnson MK (1990) Organophosphates and delayed neuropathy. Is NTE alive and well? Toxicol Appl Pharmacol, **102**: 385–399.

Kägi MK, Wüthrich B & Johansson SG (1994) Campari-Orange anaphylaxis due to carmine allergy. Lancet, **344**: 60–61.

Kanny G, Moneret-Vautrin D-A, Flabbee J, Beaudouin E, Morisset M & Thevenin F (2001) Population study of food allergy in France. Allergy, **108**: 133–140.

Kavlock RJ, Logsdon T & Gray JA (1993) Fetal development in the rat following disruption of maternal renal function during pregnancy. Teratology, **48**: 247–258.

Keen CL, Clegg MS, Hanna LA, Lanoue L, Rogers JM, Daston GP, Oteiza P & Uriu-Adams JY (2003) The plausibility of micronutrient deficiencies being a significant contributing factor to the occurrence of pregnancy complications. J Nutr, **133**(5 Suppl 2): 1597S–1605S.

Kemper RA, Hayes JA & Bogdanffy MS (2007) Metabolism: a determinant of toxicity. In: Hayes AW ed. Principles and methods of toxicology, 5th ed. Boca Raton, FL, CRC Press, pp 103–178.

Kimmel CA & Makris S (2001) Recent developments in regulatory requirements for developmental toxicity. Toxicol Lett, **120**: 73–82.

Kimmel CA, Rees DC & Francis EZ (1990) Qualitative and quantitative comparability of human and animal developmental neurotoxicology. Teratology, **12**: 175–292.

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**: 1–256.

Kitchin KT (1999) Carcinogenicity: testing, predicting, and interpreting chemical effects. New York, NY, Marcel Dekker.

Kitchin KT & Brown JL (1994) Dose–response relationship for rat liver DNA damage caused by 49 rodent carcinogens. Toxicology, **88**: 31–49.

Klaunig JE, Babich MA, Baetcke KP, Cook JC, Corton JC, David RM, DeLuca JG, Lai DY, McKee RH, Peters JM, Roberts RA & Fenner-Crisp PA (2003)

PPARα agonist-induced rodent tumors: modes of action and human relevance. Crit Rev Toxicol, **33**(6): 655–780.

Krinke GJ (1989) Neuropathologic screening in rodent and other species. J Am Coll Toxicol, **8**: 141–155.

Krishnan K & Andersen ME (2007) Physiologically-based pharmacokinetic and toxicokinetic modeling. In: Wallace Hayes A ed. Principles and methods in toxicology. New York, NY, Taylor & Francis, pp 232–291.

Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, Van Schothorst F, Vos JG & Wurtzen G (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem Toxicol, **42**: 65–83.

Kulig BM (1996) Comprehensive neurotoxicity assessment. Environ Health Perspect, **104**(Suppl 2): 317–322.

Kuper CF, Harleman JH, Richter-Reichelm HB & Vos JG (2000) Histopathologic approaches to detect changes indicative of immunotoxicity. Toxicol Pathol, **28**(3): 454–466.

Lack G, Fox D, Northstone K & Golding J; Avon Longitudinal Study of Parents and Children Study Team (2003) Factors associated with the development of peanut allergy in childhood. New Engl J Med, **348**: 977–985.

Ladefoged O, Lam HR, Ostergaard G & Nielsen E (1995) Neurotoxicology: review of definitions, methodology and criteria. Copenhagen, National Food Agency of Denmark and Danish Environmental Protection Agency (Miljoproject nr. 282).

Lau C & Kavlock RJ (1994) Functional toxicity of the heart, lungs and kidneys. In: Kimmel CA & Buelke-Sam J eds. Developmental toxicology, 2nd ed. New York, NY, Raven Press, pp 119–188.

LeBoeuf RA, Kerckaert KA, Aardema MJ & Isfort RJ (1999) Use of Syrian hamster embryo and BALB/c 3T3 cell transformation for assessing the carcinogenic potential of chemicals. In McGregor DB, Rice JM & Venitt S eds. The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. Lyon, International Agency for Research on Cancer, pp 409–425 (IARC Scientific Publications, No. 146).

Leder A, Kuo A, Cardiff RD, Sinn E & Leder P (1990) v-Ha-*ras* transgene abrogates the initiation step in mouse skin tumorigenesis: effects of phorbol esters and retinoic acid. Proc Natl Acad Sci USA, **87**: 9178–9182.

Lee W & Kim RB (2004) Transporters and renal drug elimination. Annu Rev Pharmacol Toxicol, **44**: 137–166.

Lehrer SB, Horner WE & Reese G (1996) Why are some proteins allergenic? Implications for biotechnology. Crit Rev Food Sci Nutr, **36**(6): 553–564.

Leon AS, Hunninghake DB, Bell C, Rassin DK & Tephly TR (1989) Safety of long-term large doses of aspartame. Arch Intern Med, **149**: 2318–2324.

Lipscomb JC & Ohanian EV eds (2007) Toxicokinetics and risk assessment. New York, NY, Informa Healthcare USA, Inc., 361 pp.

Luebke RW, Chen D, Dietert R, Yang Y, King M & Luster MI; Immunotoxicology Working Group (2006) The comparative immunotoxicity of five selected compounds following developmental or adult exposure. J Toxicol Environ Health B Crit Rev, **9**(1): 1–26.

Luster MI, Portier C, Pait DG, White KL Jr, Gennings C, Munson AE & Rosenthal GJ (1992) Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. Fundam Appl Toxicol, **18**(2): 200–210.

Luster MI, Portier C, Pait DG, Rosenthal GJ, Germolec DR, Corsini E, Blaylock BL, Pollock P, Kouchi Y, Craig W, White KL, Munson AE & Comment CE (1993) Risk assessment in immunotoxicology. II. Relationships between immune and host resistance tests. Fundam Appl Toxicol, **21**(1): 71–82.

Luster MI, Germolec DR, Parks C, Blanciforti L, Kashon M & Luebke RW (2005) Associating changes in the immune system with clinical diseases for interpretation in risk assessment. In: Tryphonas H, Fournier M, Blakley BR, Smits JEG & Brousseau P eds. Investigative immunotoxicology. New York, NY, Taylor & Francis, pp 165–182.

Luster MI, Dietert RR, Germolec DR, Luebke RW & Makris SL (2008) Developmental immunotoxicology. In: Sonawane B & Brown R eds. Encyclopedia of environmental health. Oxford, Elsevier.

Madsen CB, Hattersley S, Buck J, Gendel SM, Houben GF, Hourihane JO, Mackie A, Mills EN, Nørhede P, Taylor SL & Crevel RW (2009) Approaches to risk assessment in food allergy: report from a workshop “Developing a framework for assessing the risk from allergenic foods”. Food Chem Toxicol, **47**(2): 480–489.

Mahler JF, Flagler ND, Malarkey DE, Mann PC, Haseman JK & Eastin W (1998) Spontaneous and chemically induced proliferative lesions in Tg.AC transgenic and p53-heterozygous mice. Toxicol Pathol, **26**: 501–511.

Matthews EJ, Spalding JW & Tennant RW (1993) Transformation of BALB/c 3T3 cells: transformation responses of 168 chemicals compared with mutagenicity in *Salmonella* and carcinogenicity in rodent bioassays. Environ Health Perspect, **101**(Suppl 2): 347–482.

McClain RM (1993) Growth factors and tumor promotion, implications for risk assessment. Proceedings of the seventh conference on carcinogenesis and risk assessment. Austin, TX, Wiley-Liss.

McClain RM, Keller D, Casciano D, Fu P, MacDonald J, Popp J & Sagartz J (2001) Neonatal mouse model: review of methods and results. Toxicol Pathol, **29**(Suppl): 128–137.

McGregor DB (1996) A review of some properties of ethylene glycol ethers relevant to their carcinogenic evaluation. Occup Hyg, **2**: 213–235.

McGregor D (2006) Genetic toxicology. In: Duffus JH, Worth HGJ eds. Fundamental toxicology. Cambridge, RSC Publishing.

McGregor D & Anderson D (1999) DNA damage and repair in mammalian cells in vitro and in vivo as indicators of exposure to carcinogens. In: McGregor DB, Rice JM & Venitt S eds. The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. Lyon, International Agency for Research on Cancer, pp 309–354 (IARC Scientific Publications, No. 146).

McGregor DB, Rice JM & Venitt S eds (1999) Consensus report. In: The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. Lyon, International Agency for Research on Cancer, pp 1–18 (IARC Scientific Publications, No. 146).

McMillan DE (1987) Risk assessment for neurobehavioral toxicity. Environ Health Perspect, **76**: 155–161.

Mendez E, Vela C, Immer U & Janssen FW (2005) Report of a collaborative trial to investigate the performance of the R5 enzyme linked immunoassay to determine gliadin in gluten-free food. Eur J Gastroenterol Hepatol, **17**: 1053–1063.

Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL & Fuchs RL (1996) Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Crit Rev Food Sci Nutr, **36**(Suppl): S165–S186.

Mileson BE & Ferenc SA (2001) Methods to identify and characterize developmental neurotoxicity for human health risk assessment: overview. Environ Health Perspect, **109**: 77–78.

Miller EC & Miller JA (1979) Milestones in chemical carcinogenesis. Semin Oncol, **6**: 445–460.

Miller TJ, Honchel R, Espandiari P, Knapton A, Zhang J, Sistare FD & Hanig JP (2008) The utility of the K6/ODC transgenic mouse as an alternative short term dermal model for carcinogenicity testing of pharmaceuticals. Regul Toxicol Pharmacol, **50**(1): 87–97.

Moreno FJ, Mellon FA, Wickham MSJ, Bottrill AR & Mills ENC (2005) Stability of the major allergen Brazil nut 2S albumin (Ber e 1) to physiologically relevant in vitro gastrointestinal digestion. FEBS J, **272**: 341–352.

Morton D, Alden CL, Roth AJ & Usui T (2002) The Tg rasH2 mouse in cancer hazard identification. Toxicol Pathol, **30**(1): 139–146.

Moser VC (1989) Screening approaches to neurotoxicity: a functional observational battery. J Am Coll Toxicol, **8**: 85–93.

Moser VC (1995) Comparisons of the acute effects of cholinesterase inhibitors using a neurobehavioural screening battery in rats. Neurotoxicol Teratol, **17**: 617–625.

Moser VC (1997) Behavioural screening for neurotoxicity. In: Lowndes HE & Reuhl KR eds. Comprehensive toxicology. Vol. 11. Nervous system and behavioural toxicology. New York, NY, Pergamon Press, pp 351–362.

Moser VC, Tilson HA, MacPhail RC, Becking GC, Cuomo V, Frantik E, Kulig BM & Winneke G (1997a) The IPCS collaborative study on neurobehavioral screening methods: II. Protocol design and testing procedures. Neurotoxicology, **18**: 929–938.

Moser VC, Becking GC, Cuomo V, Frantik E, Kulig BM, MacPhail RC, Tilson HA, Winneke G, Brightwell WS, Cagiano R, Gill MW, Haggerty GC, Hornychova M, Lammers J, Larsen J-J, McDaniel KL, Nelson BK & Ostergaard G (1997b) The IPCS collaborative study on neurobehavioral screening methods: III. Results of proficiency studies. Neurotoxicology, **18**: 939–946.

Moser VC, Becking GC, Cuomo V, Frantik E, Kulig BM, MacPhail RC, Tilson HA, Winneke G, Brightwell WS, De Salvia MA, Gill MW, Haggerty GC, Hornychova M, Lammers J, Larsen J-J, McDaniel KL, Nelson BK & Ostergaard G (1997c) The IPCS collaborative study on neurobehavioral screening methods: IV. Control data. Neurotoxicology, **18**: 947–967.

Moser VC, Becking GC, Cuomo V, Frantik E, Kulig BM, MacPhail RC, Tilson HA, Winneke G, Brightwell WS, De Salvia MA, Gill MW, Haggerty GC, Hornychova M, Lammers J, Larsen J-J, McDaniel KL, Nelson BK & Ostergaard G (1997d) The IPCS collaborative study on neurobehavioral screening methods: V. Results of chemical testing. Neurotoxicology, **18**: 969–1055.

Müller HJ (1927) Artificial transmutation of the gene. Science, **66**: 84–87.

Munro IC, Ford RA, Kennepohl E & Sprenger JG (1996) Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. Food Chem Toxicol, **34**(9): 829–867.

Nelson BK (1991) Selecting exposure parameters in developmental neurotoxicity assessments. Neurotoxicol Teratol, **13**: 569–573.

O'Donoghue JL (1989) Screening for neurotoxicity using a neurologically based examination and neuropathology. J Am Coll Toxicol, **8**: 97–115.

O'Donoghue JL (1994) Defining what is neurotoxic. In: Weiss B & O'Donoghue JL eds. Neurobehavioural toxicity: analysis and interpretation. New York, NY, Raven Press, pp 19–33.

OECD (1981a) Carcinogenicity studies. OECD Guidelines for Testing of Chemicals: Test Guideline No. 451. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s36/p1).

OECD (1981b) Chronic toxicity studies. OECD Guidelines for Testing of Chemicals: Test Guideline No. 452. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s37/p1).

OECD (1981c) Combined chronic toxicity/carcinogenicity studies. OECD Guidelines for Testing of Chemicals: Test Guideline No. 453. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s38/p1).

OECD (1983) One-generation reproduction toxicity study. OECD Guidelines for Testing of Chemicals: Test Guideline No. 415. Paris, Organization for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s16/p1).

OECD (1995a) Repeated dose 28-day oral toxicity study in rodents. OECD Guidelines for Testing of Chemicals: Test Guideline No. 407. Paris, Organisation for Economic Co-operation and Development.

OECD (1995b) Delayed neurotoxicity of organophosphorus substances following acute exposure. OECD Guidelines for Testing of Chemicals: Test Guideline No. 418. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s19/p1).

OECD (1995c) Delayed neurotoxicity of organophosphorus substances: 28-day repeated dose study. OECD Guidelines for Testing of Chemicals: Test Guideline No. 419. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s20/p1).

OECD (1995d) Reproduction/developmental toxicity screening test. OECD Guidelines for Testing of Chemicals: Test Guideline No. 421. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s22/p1).

OECD (1996) Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test. OECD Guidelines for Testing of Chemicals: Test Guideline No. 422. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s23/p1).

OECD (1997) Neurotoxicity study in rodents. OECD Guidelines for Testing of Chemicals: Test Guideline No. 424. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s25/p1).

OECD (1998a) Repeated dose 90-day oral toxicity study in rodents. OECD Guidelines for Testing of Chemicals: Test Guideline No. 408. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s9/p1).

OECD (1998b) Repeated dose 90-day oral toxicity study in non-rodents. OECD Guidelines for Testing of Chemicals: Test Guideline No. 409. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s10/p1).

OECD (2001a) Prenatal developmental toxicity study. OECD Guidelines for Testing of Chemicals: Test Guideline No. 414. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s15/p1).

OECD (2001b) Two-generation reproduction toxicity study. OECD Guidelines for Testing of Chemicals: Test Guideline No. 416. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s17/p1).

OECD (2007) Developmental neurotoxicity study. OECD Guidelines for Testing of Chemicals: Test Guideline No. 426. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s27/p1).

OECD (2008) Repeated dose 28-day oral toxicity study in rodents. OECD Guidelines for Testing of Chemicals: Test Guideline No. 407. Paris, Organisation for Economic Co-operation and Development (http://masetto.sourceoecd.org/vl=1992568/cl=12/nw=1/rpsv/ij/oecdjournals/1607310x/v1n4/s8/p1).

Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL Jr, Valanis B, Williams JH Jr, Barnhart S, Cherniack MG, Brodkin CA & Hammar S (1996) Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. J Natl Cancer Inst, **88**: 1550–1559.

Osterballe M, Hansen TK, Mortz CG, Host A & Bindslev-Jensen C (2005) The prevalence of food hypersensitivity in an unselected population of children and adults. Pediatr Allergy Immunol, **16**: 567–573.

Padilla S, Wilson VZ & Bushnell PJ (1994) Studies on the correlation between blood cholinesterase inhibition and “tissue” inhibition in pesticide-treated rats. Toxicology, **92**: 11–25.

Parry JM, Fielder RJ & McDonald A (1994) Thresholds for aneuploidy-inducing chemicals. Advisory Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment of the UK Department of Health. Mutagenesis, **9**: 503–504.

Pepelko WE (1987) Feasibility of route extrapolation in risk assessment. Br J Ind Med, **44**(10): 649–651.

Peto R, Pike MC, Day NE, Gray RG, Lee PN, Parish S, Peto J, Richards S & Wahrendorf J (1980) Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. Lyon, International Agency for Research on Cancer.

Pumphrey RSH & Gowland H (2007) Further fatal allergic reactions to foods in the United Kingdom, 1999–2006. J Allergy Clin Immuol, **119**: 1018–1019.

Quillardet P & McGregor D (1999) Identification of carcinogenic substances by means of short-term tests in bacteria. In: McGregor DB, Rice JM & Venitt S eds. The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. Lyon, International Agency for Research on Cancer, pp 487–498 (IARC Scientific Publications, No. 146).

Ray DE (1999) Chronic effects of low level exposure to anticholinesterases—a mechanistic review. Toxicol Lett, **103**: 527–533.

Rees DC, Francis EZ & Kimmel CA (1990) Qualitative and quantitative comparability of human and animal developmental neurotoxicity. Neurotoxicol Teratol, **12**: 171–292.

Renwick AG (2008) Toxicokinetics. In: Hayes AW ed. Principles and methods of toxicology, 5th ed. Boca Raton, FL, CRC Press, pp 179–230.

Rice JM, Baan RA, Blettner M, Genevois-Charmeau C, Grosse Y, McGregor DB, Partensky C & Wilbourn JD (1999) Rodent tumors of urinary bladder, renal cortex, and thyroid gland in IARC Monographs evaluations of carcinogenic risk to humans. Toxicol Sci, **49**: 166–171.

Richardson RJ (1995) Assessment of the neurotoxic potential of chlorpyrifos relative to other organophosphorous compounds: a critical review of the literature. J Toxicol Environ Health, **44**: 135–165.

Robaire B & Hales BF eds (2003) Advances in male mediated developmental toxicity. In: Advances in experimental medicine and biology. Vol. 518. New York, NY, Kluwer Academic/Plenum Publishers, 300 pp.

Robinson DE & MacDonald JS (2001) Background and framework for ILSI's collaborative evaluation program on alternative models for carcinogenicity assessment. International Life Sciences Institute. Toxicol Pathol, **29**(Suppl): 13–19.

Rodier PM (1990) Developmental neurotoxicology. Toxicol Pathol, **18**: 89–95.

Rodier PM, Cohen IP & Buelke-Sam J (1994). Developmental neurotoxicity: neuroendocrine manifestations of CNS insults. In: Kimmel CA & Buelke-Sam J eds. Developmental toxicology, 2nd ed. New York, NY, Raven Press, pp 65–92.

Rodricks JV, Gaylor DW & Turnbull D (2007) Quantitative extrapolations in toxicology. In: Hayes AW ed. Principles and methods of toxicology, 5th ed. Boca Raton, FL, CRC Press, pp 453–474.

Rona RJ, Keil T, Summers C, Gislason D, Zuidmeer L, Sodergren E, Sigurdardottir ST, Lindner T, Goldhahn K, Dahlstrom J, McBride D & Madsen C (2007)

The prevalence of food allergy: a meta-analysis. J Allergy Clin Immunol, **120**(3): 638–646.

Royal College of Physicians (1990a) Guidelines on the practice of ethics committees in medical research involving human subjects, 2nd ed. London, Royal College of Physicians.

Royal College of Physicians (1990b) Research involving patients. Summary and recommendations of a report of the UK Royal College of Physicians. J R Coll Physicians Lond, **24**: 10–14.

Sampson HA (1999) Food allergy. Part 1: Immunopathogenesis and clinical disorders. J Allergy Clin Immunol, **103**: 717–728.

Sampson HA (2004) Update on food allergy. J Allergy Clin Immunol, **113**: 805–819.

Schardein JL (2000) Chemically induced birth defects, 3rd ed. New York, NY, Marcel Dekker, pp 1–65.

Scheurer S, Lauer I, Foetisch K, Moncin MSM, Retzek M, Hartz C, Enrique E, Lidholm J, Cistero-Bahima A & Vieths S (2004) Strong allergenicity of Pru av 3, the lipid transfer protein from cherry, is related to high stability against thermal processing and digestion. J Allergy Clin Immunol, **114**: 900–907.

Schulze GE & Boysen BG (1991) A neurotoxicity screening battery for use in safety evaluation: effects of acrylamide and 3,3-iminodipropionitrile. Fundam Appl Toxicol, **16**: 602–615.

Sen M, Kopper R, Pons L, Abraham EC, Burks AW & Bannon GA (2002) Protein structure plays a critical role in peanut allergen stability and may determine immunodominant IgE-binding epitopes. J Immunol, **169**: 882–887.

Sette WF & MacPhail RC (1992) Qualitative and quantitative issues in assessment of neurotoxic effects. In: Tilson H & Mitchel C eds. Neurotoxicology. New York, NY, Raven Press, pp 345–361.

SGOMSEC (1996) Risk assessment for neurobehavioural toxicity. Scientific Group on Methodologies for the Safety Evaluation of Chemicals. Environ Health Perspect, **104**(Suppl 4): 171–212.

Shah E & Pongracic J (2008) Food-induced anaphylaxis: who, what, why, and where? Pediatr Ann, **37**: 536–541.

Shelby MD, Bishop JB, Mason JM & Tindall KR (1993) Fertility, reproduction, and genetic disease: studies on the mutagenic effects of environmental agents on mammalian germ cells. Environ Health Perspect, **100**: 283–291.

Skolnick HS, Conover-Walker MK, Koerner CB, Sampson HA, Burks W & Wood RA (2001) The natural history of peanut allergy. J Allergy Clin Immunol, **107**: 367–374.

Sleijffers A, Yucesoy B, Kashon M, Garssen J, De Gruijl FR, Boland GJ, Van Hattum J, Luster MI & Van Loveren H (2003) Cytokine polymorphisms play a role in susceptibility to ultraviolet B–induced modulation of immune responses after hepatitis B vaccination. J Immunol, **170**(6): 3423–3428.

Slikker W (1997) The developing nervous system. In: Lowndes HE & Reuhl KR eds. Comprehensive toxicology. Vol. 11. Nervous system and behavioural toxicology. New York, NY, Pergamon Press, pp 295–308.

Solecki R, Bürgin H, Buschmann J, Clark R, Duverger M, Fialkowski O, Guittin P, Hazelden K, Hellwig J, Hoffmann E, Hofmann T, Hübel U, Khalil S, Lingk W, Mantovani A, Moxon M, Müller S, Parkinson M, Paul M, Paumgartten F & Pfeil R (2001) Harmonisation of rat fetal skeletal terminology and classification. Report of the third workshop on the terminology in developmental toxicology: Berlin, 14–16 September 2000. Reprod Toxicol, **15**: 713–721.

Solecki R, Davies L, Dellarco V, Dewhurst I, Van Raaij M & Tritscher A (2005) Guidance on setting of acute reference dose (ARfD) for pesticides. Food Chem Toxicol, **43**: 1569–1593.

Sonich-Mullin C, Fielder R, Wiltse K, Baetcke K, Dempsey J, Fenner-Crisp P, Grant D, Hartley M, Knaap A, Kroese D, Mangelsdorf I, Meek E, Rice JM & Younes M (2001) IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul Toxicol Pharmacol, **34**: 146–152.

Spanjersberg MQ, Kruizinga AG, Rennen MA & Houben GF (2007) Risk assessment and food allergy: the probabilistic model applied to allergens. Food Chem Toxicol, **45**(1): 49–54.

Spencer PS & Schaumburg HH (1985) Organic solvent neurotoxicity. Facts and research needs. Scand J Work Environ Health, **11**(Suppl 1): 53–60.

Spencer PS, Bischoff MC & Schaumberg HH (1980) Neuropathological methods for the detection of neurotoxic disease. In: Spencer PS & Schaumberg HH eds. Experimental and clinical neurotoxicology. Baltimore, MD, Williams and Wilkins Press, pp 743–757.

Spencer PS, Schaumburg HH & Ludolph AC (2000) Experimental and clinical neurotoxicology, 2nd ed. New York, NY, Oxford University Press.

Stadler MB & Stadler BM (2003) Allergenicity prediction by protein sequence. FASEB J, **17**: 1141–1143 [cited in Goodman et al., 2008].

Stevenson DE (1990) Mouse liver carcinogenesis: mechanisms and species comparisons. Austin, TX, Wiley-Liss.

Storer RD, McKelvey TW, Kraynak AR, Elia MC, Barnum JE, Harmon LS, Nichols WW & DeLuca JG (1996) Revalidation of the in vitro alkaline elution/rat hepatocytes assay for DNA damage: improved criteria for assessment of cytotoxicity and genotoxicity and results for 81 compounds. Mutat Res, **368**: 59–101.

Storer RD, French JE, Haseman J, Hajian G, LeGrand EK, Long GG, Mixson LA, Ochoa R, Sagartz JE & Soper KA (2001) P53+/− hemizygous knockout mouse: overview of available data. Toxicol Pathol, **29**(Suppl): 30–50.

Strahl BD & Allis CD (2000) The language of covalent histone modifications. Nature, **403**: 41–45.

Tandon P, Padilla S, Barone S, Pope CN & Tilson HA (1994) Fenthion produces a persistent decrease in muscarinic receptor function in the adult rat retina. Toxicol Appl Pharmacol, **125**: 271–280.

Taylor SL & Hefle SL (2001) Food allergies and other food sensitivities. Food Technol, **55**(9): 68–83.

Taylor SL & Lehrer SB (1996) Principles and characteristics of food allergens. Crit Rev Food Sci Nutr, **36**(Suppl): S91–S118.

Taylor SL, Bush RK & Nordlee JA (1997) Sulfites. In: Metcalfe DD, Sampson HA & Simon RA eds. Food allergy, 2nd ed. Cambridge, MA, Blackwell Science.

Taylor SL, Hefle SL, Bindslev-Jensen C, Burks AW, Christie L, Hill DJ, Host A, Hourihane JO, Lack G, Metcalfe DD, Moneret-Vautrin DA, Vadas PA, Rance F, Skrypec DJ, Trautman TA, Yman IM & Zeigler RS (2002) Factors affecting the determination of threshold doses for allergenic foods: how much is too much? J Allergy Clin Immunol, **109**: 24–30.

Taylor SL, Hefle SL, Bindslev-Jensen C, Atkins FM, Andre C, Bruijnzeel-Koomen C, Burks AW, Bush RK, Ebisawa M, Eigenmann PA, Host A, Hourihane JO, Isolauri E, Hill DJ, Knulst A, Lack G, Sampson HA, Moneret-Vautrin DA, Rance F, Vadas PA, Yunginger JW, Zeigler RS, Salminen JW, Madsen C & Abbott P (2004) A consensus protocol for the determination of the threshold doses for allergenic foods: how much is too much? Clin Exp Allergy, **34**: 689–695.

Taylor SL, Crevel RWR, Sheffield D, Kabourek J & Baumert J (2009) Threshold dose for peanut: risk characterization based upon published results from challenges of peanut-allergic individuals. Food Chem Toxicol, **47**(6): 1198–1204.

Tennant RW, Margolin BH, Shelby MD, Zeiger E, Haseman JK, Spalding J, Caspary W, Resnick M, Stasiewicz S, Anderson B & Minor B (1987) Prediction of chemical carcinogenicity in rodents from in vitro genetic toxicity assays. Science, **236**: 933–941.

Tennant RW, Stasiewicz S, Eastin WC, Mennear JH & Spalding JW (2001) The Tg.AC (v-Ha-ras) transgenic mouse: nature of the model. Toxicol Pathol, **29**(Suppl): 51–59.

Thake DC, Iatropoulos MJ, Hard GC, Hotz KJ, Wang CX, Williams GM & Wilson AG (1995) A study of the mechanism of butachlor-associated gastric neoplasms in Sprague-Dawley rats. Exp Toxicol Pathol, **47**: 107–116.

Thomas GA & Williams ED (1991) Evidence for and possible mechanisms of non-genotoxic carcinogenesis in the rodent thyroid. Mutat Res, **248**: 357–370.

Thomas K, Herouet-Guicheney C, Ladics G, Bannon G, Cockburn A, Crevel R, Fitzpatrick J, Mills C, Privalle L & Vieths S (2007) Evaluating the effect of food processing on the potential human allergenicity of novel proteins: international workshop report. Food Chem Toxicol, **45**: 1116–1122.

Thompson ED (1986) Comparison of in vivo and in vitro cytogenetic assay results. Environ Mutagen, **8**: 753–768.

Thorgeirsson UP, Dalgard DW, Reeves J & Adamson RH (1994) Tumor incidence in a chemical carcinogenesis study of nonhuman primates. Regul Toxicol Pharmacol, **19**: 130–151.

Tilson HA (1990a) Behavioural indices of neurotoxicity. Toxicol Pathol, **18**: 96–104.

Tilson HA (1990b) Neurotoxicology in the 1990s. Neurotoxicol Teratol, **12**: 293–300.

Tilson HA (1998) Developmental neurotoxicology of endocrine disruptors and pesticides: identification of information gaps and research needs. Environ Health Perspect, **106**: 807–811.

Tilson HA & Moser VC (1992) Comparison of screening approaches. Neurotoxicology, **13**: 1–14.

Tilson HA, MacPhail RC, Moser VC, Becking GC, Cuomo V, Frantik E, Kulig BM & Winneke G (1997) The IPCS collaborative study on neurobehavioural screening methods: VII. Summary and conclusions. Neurotoxicology, **18**: 1065–1070.

Toews AD & Morell P (1999) Molecular biological approaches in neurotoxicology. In: Tilson HA & Harry GJ eds. Neurotoxicology. New York, NY, Taylor & Francis, pp 1–35.

Troncone R, Ivarsson A, Szajewska H & Mearin ML; Members of European Multistakeholder Platform on CD (CDEUSSA) (2008) Review article: Future research on coeliac disease—a position report from the European multistakeholder platform on coeliac disease (CDEUSSA). Aliment Pharmacol Ther, **27**: 1030–1043.

Tuch K, Ockert D, Hauschke D & Christ B (1992) Comparison of the ECL-cell frequency in the stomachs of 3 different rat strains. Pathol Res Pract, **188**: 672–675.

Tugwood JD & Elcombe CR (1999) Predicting carcinogenicity: peroxisome proliferators. In: Kitchin KT ed. Carcinogenicity. New York, NY, Marcel Dekker, pp 337–360.

Tupper DE & Wallace RB (1980) Utility of the neurological examination in rats. Acta Neurobiol Exp, **40**: 999–1003.

United States Congress, Office of Technology Assessment (1990) Neurotoxicity: identifying and controlling poisons of the nervous system. Washington, DC, United States Government Printing Office (OTA-BA-436).

USEPA (1991a) Pesticide assessment guidelines, subdivision F, hazard evaluation: human and domestic animals. Addendum 10: Neurotoxicity, series 81, 82, and 83. Washington, DC, United States Environmental Protection Agency, Office of Prevention, Pesticides and Toxic Substances (EPA No. 540/09-91-123).

USEPA (1991b) Guidelines for developmental toxicity risk assessment. Washington, DC, United States Environmental Protection Agency. Fed Regist, **56**: 63798–63826 (EPA/600/FR-91/001; http://www.epa.gov/NCEA/raf/pdfs/devtox.pdf).

USEPA (1991c) Neurotoxicity testing guidelines. Washington, DC, United States Environmental Protection Agency.

USEPA (1991d) Alpha 2μ-globulin: association with chemically induced renal toxicity and neoplasia in the male rat. Washington, DC, United States Environmental Protection Agency, Risk Assessment Forum.

USEPA (1996) Guidelines for reproductive toxicity risk assessment. Washington, DC, United States Environmental Protection Agency (http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=2838).

USEPA (1998a) Guidelines for neurotoxicity risk assessment. Washington, DC, United States Environmental Protection Agency. Fed Regist, **63**: 26926–26951.

USEPA (1998b) Health effects test guidelines. OPPTS 870.3800. Reproduction and fertility effects. Washington, DC, United States Environmental Protection Agency (Document No. EPA 712-C-98-208; http://www.epa.gov/opptsfrs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3800.pdf).

USEPA (1998c) Health effects test guidelines. OPPTS 870.3700. Prenatal developmental toxicity study. Washington, DC, United States Environmental Protection Agency (Document No. EPA 712-C-98-207; http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3700.pdf).

USEPA (1998d) Health effects test guidelines. OPPTS 870.3150. 90-day oral toxicity in nonrodents. Washington, DC, United States Environmental Protection Agency (Document No. EPA 712-C-98-2200; http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3150.pdf).

USEPA (1998e) Health effects test guidelines. OPPTS 870.3100. 90-day oral toxicity in rodents. Washington, DC, United States Environmental Protection Agency (Document No. EPA 712-C-98-199; http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3100.pdf).

USEPA (1998f) Health effects test guidelines. OPPTS 870.4100. Chronic toxicity. Washington, DC, United States Environmental Protection Agency (Document No. EPA 712-C-98-210; http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-4100.pdf).

USEPA (2000) Health effects test guidelines. OPPTS 870.3050. Repeated dose 28-day oral toxicity in rodents. Washington, DC, United States Environmental Protection Agency (Document No. EPA 712-C-00-366; http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3050.pdf).

USEPA (2002) A review of the reference dose and reference concentration processes. May 2002 external review draft. Washington, DC, United States Environmental Protection Agency (EPA/630/P-02/002A; http://www.epa.gov/ncea/raf/pdfs/RfDRfC/rfdrfcextrevdrft.pdf).

USFDA (2000) Toxicological principles for the safety assessment of direct food additives and color additives used in food (Redbook 2000). Washington, DC, United States Food and Drug Administration, Center for Food Safety and Applied Nutrition (http://www.cfsan.fda.gov/~redbook/red-toca.html).

USNRC (1992) Environmental neurotoxicology. Prepared by the Committee on Neurotoxicology and Models for Assessing Risk, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council. Washington, DC, National Academy Press (http://books.nap.edu/openbook.php?record_id=1801&page=9).

USNRC (1993) Pesticides in the diets of infants and children. Prepared by the Committee on Pesticides in the Diets of Infants and Children, Board on Agriculture and Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council. Washington, DC, National Academy Press (http://www.nap.edu/openbook.php?isbn=0309048753).

USNRC (1999) Hormonally active agents in the environment. Prepared by the Committee on Hormonally Active Agents in the Environment, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council. Washington, DC, National Academy Press (http://www.nap.edu/openbook.php?isbn=0309064198).

USNRC (2004) Intentional human dosing studies for EPA regulatory purposes: scientific and ethical issues. Prepared by the Committee on the Use of Third Party Toxicity Research with Human Research Participants, Science and Technology Law Program, Policy and Global Affairs Division, National Research Council. Washington, DC, National Academies Press (http://www.nap.edu/openbook.php?isbn=0309091721).

USNRC (2006) Health risks from dioxin and related compounds: evaluation of the EPA reassessment. Prepared by the Committee on EPA's Exposure and Human Health Reassessment of TCDD and Related Compounds, Board on

Environmental Studies and Toxicology, National Research Council of the National Academies. Washington, DC, National Academies Press (http://www.nap.edu/catalog.php?record_id=11688#toc).

Vaino H, Magee PN, McGregor DB & McMichael AJ eds (1992) Consensus report. In: Mechanisms of carcinogenesis in risk identification. Lyon, International Agency for Research on Cancer, pp 1–14 (IARC Scientific Publications, No. 116).

Van den Brandt P, Voorrips L, Hertz-Picciotto I, Shuker D, Boeing H, Speijers G, Guittard C, Kleiner J, Knowles M, Wolk A & Goldbohm A (2002) The contribution of epidemiology to risk assessment of chemicals in food and diet. Food Chem Toxicol, **40**: 387–424.

Van Loveren H (1995) Host resistance models. Hum Exp Toxicol, **14**: 137–140.

Van Loveren H, Cockshott A, Gebel T, Gundert-Remy U, de Jong WH, Matheson J, McGarry H, Musset L, Selgrade MK & Vickers C (2008) Skin sensitization in chemical risk assessment: report of a WHO/IPCS international workshop focusing on dose–response assessment. Regul Toxicol Pharmacol, **50**(2): 155–199.

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.

Van Steeg H, de Vries A, Van Oostrom CT, Van Benthem J, Beems RB & Van Kreijl CF (2001) DNA repair-deficient Xpa and Xpa/p53+/− knock-out mice: nature of the models. Toxicol Pathol, **29**(Suppl): 109–116.

Vassilopoulou E, Rigby N, Moreno FJ, Zuidmeer L, Akkerdaas J, Tassios I, Papadopoulos NG, Saxoni-Papageorgiou P, Van Ree R & Mills C (2006) Effect of in vitro gastric and duodenal digestion on the allergenicity of grape lipid transfer protein. J Allergy Clin Immunol, **118**: 473–480.

VICH (2000) Good Clinical Practice. Brussels, International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products (VICH GL9; http://www.vichsec.org/pdf/2000/Gl09_st7.pdf).

VICH (2002) Studies to evaluate the safety of residues of veterinary drugs in human food: carcinogenicity testing. Brussels, International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products (VICH GL28; http://www.vichsec.org/pdf/11_2002/Gl28_st7.pdf).

VICH (2004) Studies to evaluate the safety of residues of veterinary drugs in human food: general approach to establish a microbiological ADI. Brussels, International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products (VICH GL36; http://www.vichsec.org/pdf/05_2004/Gl36_st7_F_rev.pdf).

Vohr H-W (2005) Encyclopedic reference of immunotoxicology. Berlin, Springer.

Vorhees CV (1987) Reliability, sensitivity and validity of indices of neurotoxicity. Neurotoxicol Teratol, **9**: 445–464.

Vos JG & Van Loveren H (1998) Experimental studies on immunosuppression: how do they predict for man? Toxicology, **129**(1): 13–26.

Wal JM (2001) Structure and function of milk allergens. Allergy, **56**: 35–38.

Weisglas-Kuperus N, Sas TC, Koopman-Esseboom C, Van der Zwan CW, De Ridder MA, Beishuizen A, Hooijkaas H & Sauer PJ (1995) Immunologic effects of background prenatal and postnatal exposure to dioxins and polychlorinated biphenyls in Dutch infants. Pediatr Res, **38**(3): 404–410.

Weisglas-Kuperus N, Patandin S, Berbers GA, Sas TC, Mulder PG, Sauer PJ & Hooijkaas H (2000) Immunologic effects of background exposure to polychlorinated biphenyls and dioxins in Dutch preschool children. Environ Health Perspect, **108**(12): 1203–1207.

Wensing M, Penninks AH, Hefle SL, Koppelman SJ, Bruijinzeel-Koomen CAFM & Knulst AC (2002) The distribution of individual threshold doses eliciting allergic reactions in a population with peanut allergy. J Allergy Clin Immunol, **110**: 915–920.

WHO (1967) Procedures for investigating intentional and unintentional food additives. Report of a WHO Scientific Group. Geneva, World Health Organization (WHO Technical Report Series, No. 348).

Wier PJ, Guerriero FJ & Walker RF (1989) Implantation of a primary screen for developmental neurotoxicity. Fundam Appl Toxicol, **13**: 118–136.

Wilks MF (2001) Human in vivo studies of non-pharmaceutical products. Toxicol Lett, **120**: 125–130.

Williams GM (1992) DNA reactive and epigenetic carcinogens. Exp Toxicol Pathol, **44**: 457–464.

Williams GM (1997) Chemicals with carcinogenic activity in the rodent liver; mechanistic evaluation of human risk. Cancer Lett, **117**: 175–188.

Williams GM (1999) Chemically induced preneoplastic lesions in rodents as indicators of carcinogenic activity. Lyon, International Agency for Research on Cancer, pp 185–202 (IARC Scientific Publications, No. 146).

Williams GM & Iatropoulos MJ (2001) Principles of testing for carcinogenic activity. In: Hayes W ed. Principles & methods of toxicology. Philadelphia, PA, Taylor & Francis, pp 959–1000.

Williams GM & Iatropoulos MJ (2002) Alteration of liver cell function and proliferation: differentiation between adaptation and toxicity. Toxicol Pathol, **30**: 41–53.

Williams GM & Perrone C (1996) Mechanism-based risk assessment of peroxisome proliferating rodent hepatocarcinogens. In: Reddy JK, Suga T, Mannaerts GP, Lazarow PB & Subramani S eds. Peroxisomes: biology and role in toxicology and disease. Ann NY Acad Sci, **804**: 554–572.

Wise LD, Beck SL, Beltrame D, Beyer BK, Chahoud I, Clark RL, Clark R, Druga AM, Feuston MH, Guittin P, Henwood SM, Kimmel CA, Lindstrom P, Palmer AK, Petrere JA, Solomon HM, Yasuda M & York RG (1997) Terminology of developmental abnormalities in common laboratory mammals (version 1). Teratology, **55**: 249–292.

World Medical Association (1997) World Medical Association Declaration of Helsinki: Recommendations guiding medical doctors in biomedical research involving human subjects. As revised at the 48th General Assembly of the World Medical Association, Somerset West, South Africa, 1996. JAMA, **277**: 925–926.

Wüthrich B (1993) Adverse reaction to food additives. Ann Allergy, **71**: 379–384.

Wüthrich B, Kägi MK & Stücker W (1997) Anaphylactic reactions to ingested carmine (E120). Allergy, **52**: 1133–1137.

Youssef AF, Borellini F, Jacobson-Kram D & Fort FL (2001) Analysis for loss of heterozygosity (LOH) of *p53* allele in tumors derived from p53+/− and CD-1 mice following repeated subcutaneous injections of solutions containing antioxidants. Environ Mol Mutagen, **37**: 27–30.

Zuberbier T, Edenharter G, Worm M, Ehlers I, Reimann S, Hantke T, Roehr CC, Bergmann KE & Niggemann B (2004) Prevalence of adverse reactions to food in Germany—a population study. Allergy, **59**: 338–345.

Zuidmeer L, Goldhahn K, Rona RJ, Gislason D, Madsen C, Summers C, Sodergren E, Dahlstrom J, Lindner T, Sigurdardottir ST, McBride D & Keil T (2008) The prevalence of plant food allergies: a systematic review. J Allergy Clin Immunol, **121**: 1210–1218.