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

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

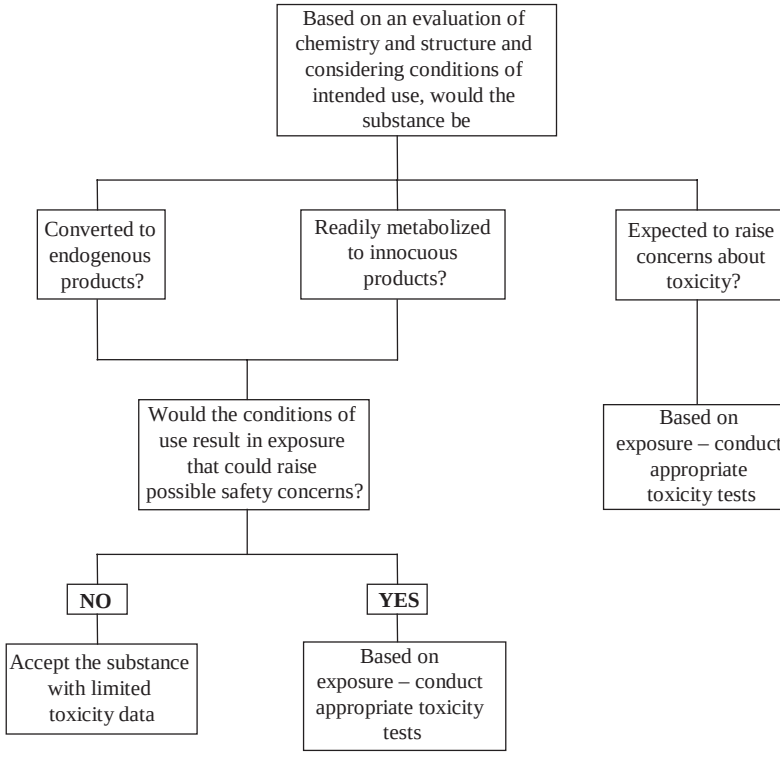

Fig. 4.1. A stepwise approach to assessing toxicity testing needs

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 [$AUC_{oral} \times dose_{iv} / AUC_{iv} \times dose_{oral}$] 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.

4.5 Genotoxicity[1]

4.5.1 Introduction

Genetic toxicology—the study of toxic effects on the inherited genetic material in cells—originated with the experiments of Müller (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, the first scientific publication dating from 1947, when Auerbach and co-workers, using Müller's fruit fly model, described mutations arising from exposure to sulfur mustards (Auerbach et al., 1947). Deep concern over mutagenesis was first expressed in the mid-1960s with the discovery of "supermutagens", as exemplified by chemicals such as the heterocyclic nitrogen mustard ICR-170, AF-2, hycanthone and β-propiolactone, which induce high levels of mutation at high levels of survival. Several leading geneticists were concerned that supermutagens might be widely distributed (Crow, 1968) because either they had passed through traditional toxicity screens without showing adverse effects or they had never been tested at all. In spite of these concerns, the major impetus given to mutagenesis as a toxicological topic came from the belief that carcinogenic activity was predictable by examining the potential of a chemical to interact with DNA. Thus, in 1973, Ames and co-workers pronounced that "carcinogens are mutagens" (Ames et al., 1973).

[1] The text in section 4.5 has been published in similar form in McGregor (2006), because the author developed the two documents in parallel.

Although this may have been the spark that became a blaze of activity in developing and validating new tests for genetic toxicity, the concern that the human germline should be well protected in its own right also benefited. The result is that tests for genetic toxicity serve to identify not only potential somatic cell mutagens that may lead to cancer via a genotoxic mode of action, but also potential human germ cell mutagens as well.

The importance of including germ cell effects specifically in genotoxicity testing is, however, questionable. Identification of substances that are germ cell mutagens, even under experimental conditions in mammals, is difficult, and quantitative studies can require large numbers of animals. In contrast, identification of potential somatic cell mutagens can be done in vitro, or with fewer animals in vivo, and to date there is no evidence of any substances that are germ cell mutagens but not somatic cell mutagens (see section 4.5.4.2). Thus, in risk assessment, a default assumption can be made that a somatic cell mutagen may also be a potential germ cell mutagen. Knowledge that a substance is a germ cell mutagen does not mean that it should be treated differently from a substance that is a somatic cell mutagen but has not been tested for germ cell mutagenicity. Regulatory decisions declaring that such hazards exist should not have different consequences in these cases. If the individual is protected from the genotoxic and carcinogenic effects of a substance, then so is the population from the heritable genetic effects. Although national regulatory authorities might take a different view, this is the practical standpoint of JMPR and JECFA at this time, and there is no clear scientific reason why this should be changed.

4.5.2 Tests for genetic toxicity

4.5.2.1 Test categories

To address the need for identifying all aspects of genetic toxicity, more than 100 different in vitro and in vivo test methods have been developed. However, only a few are commonly used. It is the diversity of potentially damaging events that has encouraged this development, and for this reason it should not be expected that there will always be consistency among the results of different classes of assays. The tests can be grouped as:

- genetic toxicity tests that measure types of DNA damage known to be precursors of genetic alterations (e.g. formation of DNA adducts or DNA strand breaks) or cellular responses to DNA damage (e.g. unscheduled DNA synthesis); and
- mutagenicity tests that measure expressed genetic damage (e.g. gene mutations, chromosomal rearrangements or deletions, and loss or gain of chromosomal segments or of whole chromosomes, the last also known as aneuploidy).

4.5.2.2 *Commonly used tests*

Commonly used tests (Table 4.1) include those for:

- gene mutation in bacteria;
- gene mutation in mammalian cell lines;
- chromosomal aberrations (including micronuclei) and aneuploidy in cultured mammalian cells;
- DNA damage in primary cultures of mammalian cells (commonly rat hepatocytes);
- in vivo tests for DNA damage (such as DNA binding, unscheduled DNA synthesis in the liver or the comet assay in a number of tissues);
- in vivo tests for chromosomal damage (including micronuclei) using mammalian haematopoietic cells; and
- in vivo tests for gene mutations.

Less commonly used testing methods, with more limited validation, make use of yeast, moulds and insects (*Drosophila*) as test organisms.

Some useful information can also be provided by tests for cell transformation in vitro. Positive results obtained with such tests, however, are not necessarily indicative of genetic toxicity in the form of reactivity with DNA; they may also represent a consequence of an epigenetic event (any heritable influence in the progeny of cells or of individuals on chromosome or gene function that is not accompanied by a change in DNA nucleotide sequence).

4.5.3 *Testing strategy*

For comprehensive coverage of the potential mutagenicity of a substance, information on the ability to induce gene mutations, structural

Table 4.1. Some assays for genetic toxicity

DNA damage/repair	Gene mutation	Chromosomal damage
In vitro assays DNA adduct measurement using cell cultures Unscheduled DNA synthesis using primary cultures (often hepatocytes) DNA strand breakage and alkali-labile sites monitored by single-cell gel electrophoresis (comet assay) or by sucrose gradient or filter elution, using cell cultures **In vivo assays** DNA adduct measurement Unscheduled DNA synthesis (usually in liver) Strand breakage and alkali-labile sites monitored by single-cell gel electrophoresis (comet assay) or by sucrose gradient or filter elution in tissue DNA	**In vitro assays** *Microbial tests* Reversion to a specific nutrient independence using: • *Salmonella typhimurium* • *Escherichia coli* Forward mutation to resistance to a specific anti-metabolite using: • *Salmonella typhimurium* • *Escherichia coli* *Mammalian tests* Forward mutation at the hypoxanthine-guanine phosphoribosyl transferase (*hprt*) gene using cell lines such as: • Chinese hamster ovary (CHO) • Chinese hamster lung (V79) • Human lymphocytes Forward mutation at the thymidine kinase (*tk*) gene using cell lines such as: • Mouse lymphoma L5178Y • *cII* and *lacI* transgenes in cultured Big Blue® mouse and rat embryonic fibroblasts **In vivo assays** Somatic cell assays: • *LacZ* (Muta™Mouse) or *lacI or cII* (Big Blue® mouse or rat) Germline cell assays: • Specific locus test in mice	**In vitro assays** Sister chromatid exchange (SCE) Chromosomal aberrations, micronuclei and aneuploidy using: • CHO and V79 cell lines and human lymphocytes **In vivo assays** Somatic cell assays: • SCE Chromosomal aberrations, micronuclei and aneuploidy using: • Bone marrow cells and lymphocytes (rodent) Germline cell assays: • Chromosomal aberrations, heritable translocations and dominant lethals using mice and rats

chromosomal aberrations and aneuploidy is required. Usually a small number of well-validated in vitro assays are selected to cover different genetic end-points. Commonly used test batteries include a gene mutation test in bacteria (i.e. the *Salmonella*/microsome assay) and one or two tests in mammalian cells detecting point mutations or chromosome damage (clastogenicity/aneugenicity).

Completely negative results in the in vitro test battery are normally considered sufficient to conclude that a substance is devoid of genotoxic potential, unless there are reasons for special concern (e.g. high or sustained human exposure, structural considerations). Conversely, one or more positive in vitro tests normally require follow-up by in vivo testing. The choice of the appropriate in vivo test is made case by case taking into account the results of in vitro assays and information on the toxicokinetics and toxicodynamics of the substance. For substances with adequate systemic availability, tests on rodent erythropoietic or liver cells are generally performed. For directly reactive, short-lived substances, tests on tissues at the initial site of contact are selected. If the first in vivo test is negative, the need for further in vivo tests is decided case by case taking into account the quality of available data, the evidence of target tissue exposure and any other relevant information.

4.5.4 Data assessment

Given the variety of test methods applied, which are designed to cover different genetic end-points, a weight of evidence approach should be used to decide whether a substance is genotoxic. A clear positive result at a single mutagenicity end-point is generally sufficient to classify a substance as positive, even when multiple negative results at other end-points are reported. On the other hand, contrasting results at the same end-point, or in the same test method, should be evaluated case by case with consideration of study design, reproducibility and biological plausibility of the results. With so many different types of assay, so many standard protocols and so many protocol variations possible for special studies, it is not practicable, other than in very general terms, to describe the process of data assessment. This process should, of course, include a judgement of the standards under which the experiments were conducted. Guideline protocols for many of these assays have been published

by the OECD (see http://masetto.sourceoecd.org/vl=2781582/cl=14/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/v1n4/contp1-1.htm).

Following an initial quality assessment of individual studies, the second scientific requirement is that any observation that is made should be reproducible. Although this principle applies to any kind of study, the resource and animal welfare constraints that hinder replication of many mammalian toxicity studies are less of a hindrance in genetic toxicology, particularly for in vitro studies. The strength of a finding is increased if it can be demonstrated in a number of laboratories. Indeed, where an observation is made in a single laboratory—even if made on a number of occasions—it is generally viewed with suspicion if other laboratories fail to achieve the same result. This suspicion may not be justified in some cases, but it is nevertheless an understandable view in data evaluation.

The third step is to look for a plausible pattern in the hierarchy of results. Such hierarchical patterns can be used only as general guides, because there can always be exceptions. It is expected that a substance that is clastogenic in vivo will also be clastogenic in vitro; and that, in vivo, a germline cell clastogen will also be clastogenic to somatic cells. Deviations from this pattern may occasionally occur, but these should be scrutinized with special care. The basis for suggesting this procedure for cytogenetic assays is given below. Unfortunately, apart from the data in section 4.5.4.3, a similar basis cannot be presented for induction of gene mutation because of the current paucity of in vivo tests for gene mutation. This situation may be expected to change as more data accumulate from the increasing use of transgenic mouse models.

4.5.4.1 Cytogenetic assays in vivo and in vitro

Thompson (1986) reviewed the literature and found 216 chemicals that had been tested both in vitro and in rodent bone marrow tests for clastogenicity. Definitive results were obtained with 181 of them, among which there was concordance between in vivo and in vitro results for 126 chemicals. Of the 55 chemicals for which there were discordant results, 53 were positive in vitro and negative in vivo. Only D-ascorbic acid and ethinylestradiol were negative in vitro while inducing significant clastogenicity in vivo in bone marrow. This leads to the conclusion that a chemical that fails to induce a significant response in

an in vitro clastogenicity assay is unlikely to be clastogenic in in vivo bone marrow assays.

4.5.4.2 Germline and somatic cell in vivo cytogenetic assays

Holden (1982) reviewed the literature and found 76 compounds that had been tested for chromosomal effects in vivo in both somatic and germline cells. Of these, concordant results were obtained for 58 chemicals. The remaining 18 chemicals for which there were discordant results were all positive (i.e. induced damage) in somatic cells only. At that time, therefore, the available evidence suggested that a negative somatic cell response is highly predictive of a negative germline cell response. Subsequently, it was suggested by a USEPA Gene-Tox workshop that six chemicals could be uniquely germline cell mutagens (Auletta & Ashby, 1988), but a re-evaluation of the Gene-Tox Program literature on these chemicals indicated that they had been misclassified (Adler & Ashby, 1989). Thus, as of 1989, there was no reason to change the presumption that all germ cell clastogens are also somatic cell clastogens. This kind of thinking led the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) to propose a testing strategy in which agents without somatic cell genotoxicity in vivo could be assumed to have no potential for germline cell genotoxicity (Arni et al., 1988).

4.5.4.3 In vivo gene mutation assays in germline cells

Data on the mouse specific locus test from the National Toxicology Program (NTP) of the United States of America (USA), reviewed by Shelby et al. (1993), identified only 6 chemicals out of 12 tested as being clearly positive. The chemicals identified as germ cell mutagens are all highly reactive and induce a wide variety of toxic effects in humans and other animals; they are also clastogenic in germ cells.

4.5.5 Genetic toxicity in relation to carcinogenicity

4.5.5.1 Validation of genetic toxicity tests for the prediction of carcinogenicity

The outcomes of different in vitro genetic toxicity tests have been compared with experimental results from rodent carcinogenicity bioassays, such as those from the NTP and its predecessor at the United States National Cancer Institute. The comparisons are expressed in

terms of sensitivity, specificity, concordance, positive predictivity and negative predictivity, as defined by Cooper et al. (1979). The first two aspects are considered to be more important than the last three, and together they provide an adequate description of assay performance. (The proportion of carcinogens that give a positive result in the test is termed the sensitivity, and the proportion of non-carcinogens that give a negative result in the test is termed the specificity.) However, determination of whether or not a compound is genotoxic should be based on an overview assessment of all the available data (see section 4.5.4); therefore, comparisons of the results from carcinogenicity bioassays with the results from a single genotoxicity testing system reflect the relationship only to that test and not to the outcome of a comprehensive testing strategy.

There were hopes in the early 1970s that it would be possible to predict whether or not a chemical was carcinogenic on the basis of a relatively simple bacterial gene mutation test. Using a database consisting of 283 chemicals (prevalence of carcinogens 62%), it was found initially that the *Salmonella*/microsome test had a sensitivity of 90% and a specificity of 87% (Ames et al., 1975). However, a subsequent analysis based on data for 301 chemicals tested by the NTP for carcinogenic activity (prevalence of carcinogens 62%) found that the sensitivity was 56% and the specificity was 75% (Ashby & Tennant, 1991).

Similar results were found in an NTP validation study of four in vitro tests using 73 chemicals (Tennant et al., 1987). The tests investigated were the *Salmonella*/microsome test, the Chinese hamster ovary (CHO) cell test for sister chromatid exchange (SCE), the CHO test for chromosomal aberrations and the mouse lymphoma L5178Y cell $tk^{+/-}$ locus test for gene mutations. Similar test performance evaluations have been made on the basis of published data (McGregor, 1996; McGregor & Anderson, 1999; Quillardet & McGregor, 1999) and on the basis of experience within single laboratories for four additional assays (Matthews et al., 1993; Kitchin & Brown, 1994; Storer et al., 1996; LeBoeuf et al., 1999). For all of these tests, the concordance was about 60%, with sensitivities ranging from 45% to 73% and specificities ranging from 43% to 86%. Because of the heterogeneity of the carcinogenicity test reporting and the variations in genetic test protocols used, the rigor of the Tennant et al. (1987) study is lacking in

some of these analyses. Overall, these data suggest that none of the results of the short-term tests considered can alone provide a reliable prediction of whether or not a chemical is a carcinogen in rodents.

More recently, the ability of a battery of three of the most commonly used in vitro genotoxicity assays (*Salmonella*/microsome assay, mouse lymphoma $tk^{+/-}$ test and chromosomal aberrations/micronucleus tests) to discriminate between rodent carcinogens and rodent non-carcinogens from a large database of over 700 chemicals has been evaluated (Kirkland et al., 2005). All test batteries displayed high sensitivity but insufficient specificity, mainly due to the high incidence of false-positive results produced by mammalian cell systems. Thus, an understanding of the mechanism of action and consideration of the weight of evidence are recommended to assess the carcinogenic risk from genotoxicity test results.

4.5.5.2 *Evidence of mode of action*

The dogma that cancer is primarily a genetic disease has led to a default assumption in the evaluation of chemicals for hazard and risk—namely, if the substance is mutagenic or clastogenic, then this is its mode of action as a carcinogen. This assumption has driven the manner in which carcinogenic chemicals are dealt with in regulatory arenas at the national and supranational levels. Its origins are simple to understand: all studied neoplasms contain mutations of one type or another; there is a single copy of DNA in every cell; therefore, it is reasoned, there can be no threshold of damage below which DNA damage has no consequence; hence, there can be no safe level of exposure to a carcinogen that is genotoxic. These considerations do not apply to substances that do not react with DNA, such as those that affect spindle function and organization, inducing aneuploidy, or that affect chromosome integrity through topoisomerase inhibition. For these compounds, which interact with redundant cellular targets, threshold-based mechanisms are assumed.

The assumption that genotoxic carcinogens act through direct DNA modification is probably useful only as a guide, as there are indications that epigenetic events are more important than hitherto believed, in both experimental animal and human cancer. Many of the schemes of epigenetic modes of action that have been developed from observations on rodents and humans involve disturbance of hormonal regulating

networks, whereas others involve enzyme induction, enzyme inhibition or other expressions of toxicity. It is also clear that a property common to the proposed modes of action is the persistent stimulation of cell populations to divide as a hyperplastic response either to toxicity or to mitogenesis; inhibition of apoptosis may also play a role. In many examples, the mode of action has been reasonably well established, up to the point where there is a cell population increase. Why hyperplasia should ever result in neoplasia is not well established, but the assumed mechanism until now has been that a shortened cell cycle time reduces the time for repair of "background" damage to DNA and increases the probability of mutation.

Whereas intermitotic DNA damage and inefficient repair may be parts of the process, it is known that relatively short cell cycle times are also characteristic of several normal cell populations (e.g. dermis, intestinal epithelium, haematopoietic tissues and, not least, throughout embryos and early fetuses). Another possibility is emerging from recent studies in molecular biology and protein chemistry. If a chemical reacts with DNA, then it is highly probable that it is also reacting with various amino acids in proteins and peptides that, with DNA and many other types of molecules, constitute chromatin. These proteins, including histones, are often involved in normal gene regulatory function. It has been proposed that there is a "histone code", based upon patterns of acetylation, methylation, phosphorylation and ubiquitylation of basic amino acid histone tails protruding from the nucleosomes, which enables other proteins to recognize specific regions of the genome (Strahl & Allis, 2000). These patterns determine whether particular genes are expressed or not. Interference with the activity or function of methyl transferases, acetylases and deacetylases, etc., may very well cause inappropriate phenotypic changes in histones, which could include the silencing of repair genes or of tumour suppressor genes.

Genetic and epigenetic mechanisms can cooperate in chemical carcinogenesis. Consequently, the International Agency for Research on Cancer (IARC) has recommended that there should be no genetic toxicity associated with a substance if an epigenetic mechanism of carcinogenesis is to be accepted. Thus, because of its implications in risk characterization and for the definition of a health-based guidance value, the elucidation of the genotoxic potential of a chemical carcinogen plays an important role in risk assessment.

4.5.6 Conclusions

The use of genetic toxicology in hazard identification, both for effects on somatic cells that may lead to cancer and for heritable effects on the germline, has been accepted in academic, industrial and regulatory circles. The tests that are used provide evidence that the chemical under study can react with biologically important molecules, either directly or after metabolism. Validation studies, using rodent carcinogenicity data as the yardstick, have shown that any prediction of carcinogenic hazard will be imperfect, but certain tests perform better than others in this respect. Similarly, evidence of a particular mode of carcinogenic action that might be derived from positive results of genetic toxicity tests will always have an element of uncertainty about it. As long as these weaknesses are not forgotten and the strengths are not overemphasized, the results can provide useful guidance in chemical risk assessment.

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 semiquantitative 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

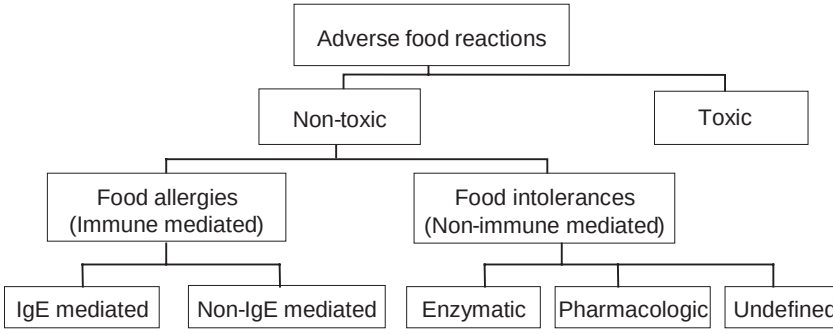

Fig. 4.2. Classification according to the European Academy of Allergology and Clinical Immunology nomenclature task force (adapted from Johansson et al., 2001)

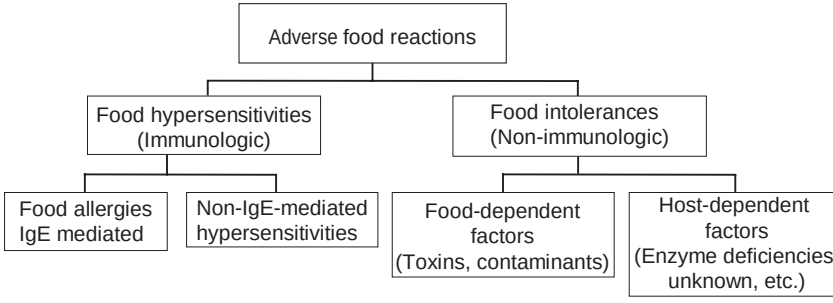

Fig. 4.3. Classification adapted from Sampson (1999)

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/GuidanceComplianceRegulatoryInformation/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.