Section 4.5

Genotoxicity

Second edition

(2020)

CONTENTS

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

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

List of abbreviations

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

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

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

List of contributors

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

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

Dr Riccardo Crebelli
Istituto Superiore di Sanità, Rome, Italy

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

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

Dr Susan Page Felter
Mason, Ohio, USA

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

Professor Andrea Hartwig
Karlsruhe, Germany

Dr Frank Le Curieux
European Chemicals Agency, Helsinki, Finland

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

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

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

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

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

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

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

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

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

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

4.5 Genotoxicity

4.5.1 Introduction

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

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

---

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

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

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

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

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

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

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

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

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

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

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

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

4.5.1.1 *Risk analysis context and problem formulation*

The identification of compounds to which exposure may lead to cancer (or other adverse effect) via a mutagenic MOA affects how these compounds are handled within regulatory paradigms. A distinction is often made between substances that require regulatory approval before use (e.g. pesticides, veterinary drugs, food additives) and those to which exposure is unavoidable (e.g. contaminants, natural constituents of the diet). In practice, this distinction affects the

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

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

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

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

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

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

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

Fig. 4.1. Decision-tree illustrating issues to be considered in assessing the mutagenic potential of different types of substances that can be found in food
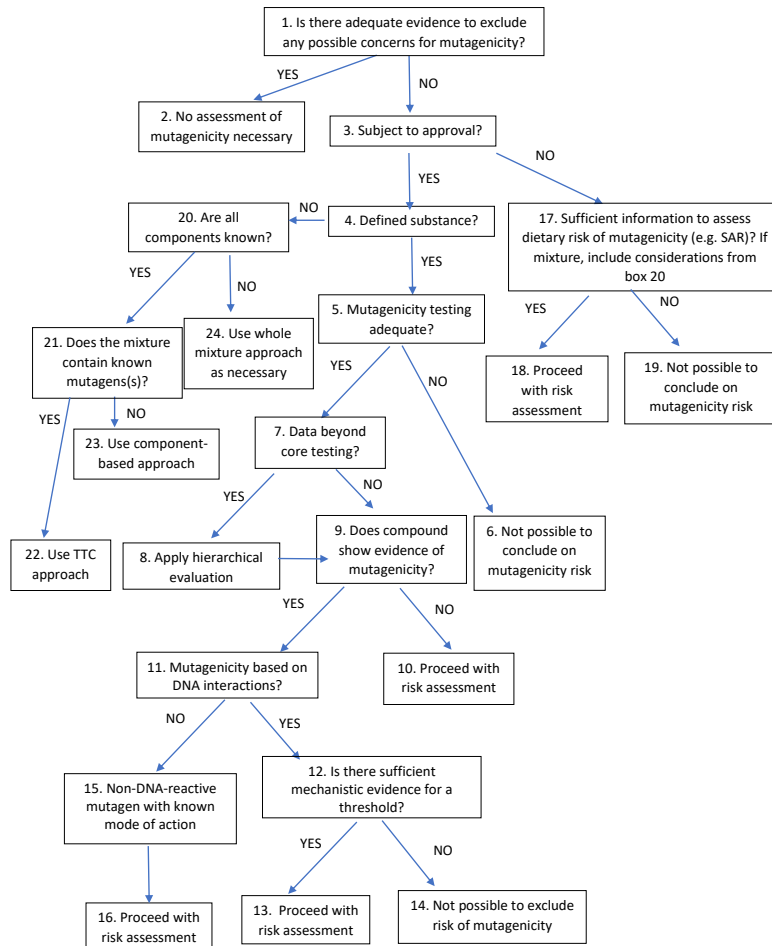
1. Is there adequate evidence to exclude any possible concerns for mutagenicity?

While it may be rare to exclude possible concerns for mutagenicity a priori, occasionally the nature of the substance or its production process may provide sufficient assurance that substance-

specific mutagenicity data are not necessary. One example is a natural constituent of the diet produced by a fully controlled process (e.g. invertase derived from *Saccharomyces cerevisiae* fermentation; FAO/WHO, 2002). [See section 4.5.6.4.]

2. No assessment of mutagenicity necessary

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

3. Subject to approval?

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

4. Defined substance?

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

5. Mutagenicity testing adequate?

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

6. Not possible to conclude on mutagenicity risk

If mutagenic potential has not been adequately tested (i.e. the answer to the question in box 5 is NO), it is not possible to conclude on the likelihood of mutagenic risk in vivo at dietary levels of

exposure. As such, it may be inappropriate to establish HBGVs that encompass potential mutagenicity. The main data gaps precluding a conclusion on mutagenic potential should be clearly articulated. [See section 4.5.4.5.]

7. Data beyond core testing?

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

8. Apply hierarchical evaluation

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

9. Does compound show evidence of mutagenicity?

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

10. Proceed with risk assessment

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

11. Mutagenicity based on DNA interactions?

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

12. Is there sufficient mechanistic evidence for a threshold?

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

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

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

14. Not possible to exclude risk of mutagenicity

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

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

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

16. Proceed with risk assessment

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

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

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

mutagenic potential of the substance. [See sections 4.5.5.1 and 4.5.5.3.]

18. Proceed with risk assessment

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

19. Not possible to conclude on mutagenicity risk

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

20. Are all components known?

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

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

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

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

22. Use TTC approach

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

23. Use component-based approach

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

24. Use whole mixture approach as necessary

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

4.5.2   *Tests for genotoxicity*

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

Table 4.1. Examples of assays for genotoxicity

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

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

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

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

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

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

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

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

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

Identification of germ cell mutagens is difficult, and studies in rodents to identify these agents historically required large numbers of animals. In contrast, identification of somatic cell mutagens can be accomplished in vitro or with fewer animals in vivo. To date, all identified germ cell mutagens are also somatic cell mutagens. Thus, in risk assessment, a default assumption is that a somatic cell mutagen may also be a germ cell mutagen. Regulatory decisions declaring that such hazards exist would not ordinarily have different consequences, unless there are demonstrated differences in potency between the

doses causing somatic versus germ cell mutagenicity, which, for example, may result in differential advice to pregnant women and the general population. For the majority of known germ/somatic cell mutagens, if the individual is protected from the genotoxic and carcinogenic effects of a substance, then that individual would also be protected from the heritable genetic effects. Although national regulatory authorities might take a different view, this is the practical viewpoint of JMPR and JECFA at this time, as information on developmental and reproductive toxicity is often available (particularly for chemicals subject to authorization in Member States).

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

4.5.2.1 Bacterial mutagenicity

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

4.5.2.2 In vitro mammalian cell mutagenicity

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

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

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

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

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

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

4.5.2.3   *In vivo mammalian cell mutagenicity*

(a)   Somatic cell assays

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

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

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

(b)   Germ cell assays

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

Rodent dominant lethal assay. The dominant lethal assay investigates whether a chemical induces mutations associated with

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

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

4.5.2.4 *In vitro chromosomal damage assays*

(a) Chromosomal aberration assay

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

(b)   Micronucleus (MN) assay

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

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

(c)   TK assay in mammalian cells

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

4.5.2.5   In vivo chromosomal damage assays

   (a)   Chromosomal aberration assay

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

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

   (b)   Micronucleus (MN) assay

   The in vivo MN test (OECD TG 474) is the most commonly used in vivo assay for chromosomal damage, as it can capture numerical and structural chromosomal changes, is not technically exacting and can be manually scored. It also lends itself to automation (flow cytometry), which speeds up data acquisition and increases the statistical power of the assay, as more cells can be readily counted (Torous et al., 2000; Dertinger et al., 2006, 2011b; MacGregor et al., 2006; Kissling et al., 2007). The standard assay evaluates MN formation in newly formed bone marrow erythrocytes of mice and rats. Modified versions of the assay can also be used in other tissues, such as the liver, spleen and colon (Morita, MacGregor & Hayashi, 2011). In most species, except mice, the spleen sequesters and destroys micronucleated erythrocytes entering the circulation, limiting the use of this assay in peripheral blood. However, this potential limitation has been overcome in a new flow cytometry version of the MN assay, which employs fluorescent dyes to identify

cell surface markers (transferrin receptors) specific to immature erythrocyte populations. This ability to distinguish erythrocytes by maturation stage allows the peripheral blood MN assay to be conducted in mice, rats and a variety of other species. MN are formed primarily by direct DNA damage, although formation through indirect mechanisms resulting from cytotoxicity and hypothermia can also occur. Positive results in in vivo chromosomal damage assays correlate with rodent (and human) carcinogenicity (Witt et al., 2000). However, the standard in vivo MN assay is limited to assessing events occurring in the rapidly dividing pro-erythrocyte population in the bone marrow, so negative results should be supported by evidence that this target cell population was adequately exposed to the putative reactive parent compound or metabolite (see subsection on "Relevance" in 4.5.4.1(b)).

4.5.2.6   In vitro DNA damage/repair assays

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

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

4.5.2.7   In vivo DNA damage/repair assays

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

The comet assay (OECD TG 489) detects DNA damage in the form of breaks that may occur endogenously through the normal action of enzymes involved in maintaining DNA integrity, such as DNA repair processes, or may be induced by exposure to DNA-damaging agents, either directly or indirectly (through the action of DNA repair processes on chemical-induced damage). The assay

detects overt double-strand and single-strand breaks as well as alkali-labile lesions (e.g. oxidized bases, alkylations, bulky adducts, crosslinks that can be converted to single-strand breaks under alkaline [pH > 13] conditions) that are visualized following electrophoresis. Furthermore, DNA strand break assays such as alkaline elution or alkaline unwinding in combination with specific DNA repair enzymes may be used to quantify specific DNA lesions, such as 8-oxoguanine. Some types of DNA breaks can be rapidly repaired, so tissues should be harvested shortly (usually 2–6 hours) after the last dose of chemical has been administered.

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

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

(b)   DNA adduct assays

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

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

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

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

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

4.5.3 Identification of relevant studies

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

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

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

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

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

---

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

"unscheduled DNA synthesis"   "DNA repair"

"sister chromatid exchange*"   "cell transformat*"

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

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

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

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

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

4.5.4   *Interpretation of test results*

Mutagenicity can be a hazard end-point of concern per se or a potential key event in the MOA for an adverse outcome such as carcinogenicity or developmental toxicity. Assessment of mutagenicity, both qualitatively and quantitatively, can therefore be of great value in interpreting the toxicological consequences of such adverse outcomes. Quantitatively, the potency of the response could

inform the nature of the overall dose–response relationship and the implications for establishing HBGVs based on these or other effects. Qualitatively, it can add to the WOE for mutagenicity as a key event in an adverse outcome, in different species, tissues, life stages, etc.

4.5.4.1 *Presentation and categorization of results*

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

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

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

Table 4.2 (continued)

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

Source: Modified from Amberg et al. (2016)

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

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

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

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

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

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

Specific aspects that should be considered for the evaluation of positive and negative findings in mutagenicity/genotoxicity studies have been addressed by the European Chemicals Agency (ECHA,

2017a). These are recommended for use in JECFA and JMPR assessments, as described below.

Particular considerations when evaluating positive results include:

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

Particular considerations when evaluating negative results include:

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

(b) Assessing data quality

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The type of document (e.g. published or unpublished study report) and TG or GLP conformance do not necessarily have an impact on reliability. Adequate data reporting is more relevant, recognizing that the quality of articles published in peer-reviewed journals is significantly higher than the quality of articles published in non-peer-reviewed journals. It is also recognized that for regulated

substances, such as food additives or pesticides, appropriate data can be requested from the petitioner or producer; this is not possible for substances such as food contaminants, for which the evaluation is performed based on available data and assessment approaches such as read-across from similar chemicals and (Q)SAR.

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

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

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

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

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

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

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

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

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

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

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

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

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

4.5.4.2   Weighting and integration of results

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

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

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

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

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

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

comparable methodology should be evaluated on a case-by-case basis using the weighting considerations outlined above.

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

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

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

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

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

4.5.4.3   Adequacy of the genotoxicity database

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

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

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

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

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

4.5.4.4 Mutagenic mode of action and adverse outcomes

The WOE conclusion on mutagenicity can be used to help interpret available data on specific adverse outcomes in humans or laboratory animals, particularly carcinogenicity and developmental toxicity. The default assumption in hazard and risk characterization has been that if the substance is mutagenic, then this is its MOA as a carcinogen. This policy decision has driven the manner in which mutagenic carcinogens are dealt with in national and international regulatory arenas and assumes that a single mutation in a single relevant gene (e.g. oncogene) could cause oncogenic transformation; therefore, it is reasoned, there can be no DNA damage threshold that is without consequence and, hence, no safe level of exposure to a mutagenic carcinogen. However, recent studies challenge this linear, non-threshold or "one-hit" theory of carcinogenesis, and

experimental thresholds have been observed for some DNA-reactive mutagenic carcinogens (Kobets & Williams, 2019). For example, studies on chromosomal damage and gene mutations in mice repeatedly exposed to the mutagen ethyl methanesulfonate demonstrated a clear, practical threshold or no-observed-genotoxic-effect level (NOGEL) (Pozniak et al., 2009). Thus, even for DNA-reactive mutagens, non-linear, threshold-type dose–response curves can be seen. For all mutagens, there may be a level of exposure below which chemical-induced mutation levels cannot be distinguished from background (spontaneous) mutation levels, which are tightly monitored by endogenous systems designed to control cellular perturbations, including DNA damage, caused by exogenous and endogenous stressors. In reaching a conclusion on the nature of the dose–response relationship and its linearity or otherwise, all relevant information on toxicokinetics and toxicodynamics should be considered, as described by Dearfield et al. (2002, 2011, 2017). In most cases, however, the available evidence is insufficient to enable a conclusion on the existence of a threshold, and the risk assessment should proceed as if there is no threshold. This is because even should a threshold exist, there would be considerable uncertainty, potentially by orders of magnitude, as to the dose at which it occurs.

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

Determining that a substance is mutagenic is not sufficient to conclude that it has a mutagenic MOA for an adverse outcome (Cimino, 2006). A WOE approach that applies various weights to different end-points or assays is recommended when evaluating

whether a substance is likely to act via a mutagenic MOA. The level of evidence is specific to the end-point that the assay is evaluating and thus needs to be considered along with all available evidence to conclude on the overall likelihood of a mutagenic MOA. Expert judgement is necessary with respect to the data quality described in section 4.5.4.2 (i.e. relevance, reliability, adequacy). For example, some factors that provide more weight include the following:

- The substance is mutagenic in the target organ or system in which the adverse outcome was observed.

- The substance is DNA reactive, or there is significant conversion to a DNA-reactive intermediate that is confirmed to be associated with the adverse outcome.

- There is evidence of substantial covalent binding to DNA, preferably in vivo in the target tissue or system.

- The substance is a multiroute, multisite and multispecies carcinogen in animal bioassays, particularly if tumours arise in tissues that do not have high spontaneous incidences or are not hormonally sensitive.

- There is evidence that the substance acts as an initiator in a well-conducted rodent tumour initiation:promotion assay.

- Highly similar structural analogues produce the same, or a pathologically closely related, adverse outcome via a mutagenic MOA; the WOE is increased if the substance contains structural alerts for DNA mutagenicity and reactivity.

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

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

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

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

4.5.4.5 Integration of carcinogenicity and mutagenicity

JECFA and JMPR integrate information on mutagenicity and carcinogenicity, together with all other relevant data, to reach an overall conclusion on carcinogenic risk. Similar to the standard phrases for mutagenic potential mentioned in section 4.5.4.3,

standard phrases with defined scenarios for chemicals with mutagenicity and carcinogenicity evaluations may include the following (adapted from WHO, 2015a, to reflect the updated guidance in this section of EHC 240). It should be noted that the wording for the conclusions on specific substances is taken from the respective meeting reports. It is anticipated that future conclusions of JMPR and JECFA will reflect the recommendations in this section of EHC 240:

[compound not carcinogenic or mutagenic]

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

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

or

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

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

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

or

[compound carcinogenic but not mutagenic]

In view of the lack of mutagenicity, the absence of carcinogenicity in [species] and the fact that only [tumours] were observed and that these were increased only in [sex] [species] at the highest dose tested,

it is concluded that [compound] is unlikely to pose a carcinogenic risk to humans from the diet. *[There is considerable flexibility in wording here.]*

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

or

[compound carcinogenic with positive in vitro mutagenicity]

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

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

or

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

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

or

[compound lacks carcinogenicity data]

If a compound lacks carcinogenicity data or has carcinogenicity data with major limitations, with or without adequate genotoxicity data, it should be noted that a conclusion on carcinogenic potential cannot be reached, and the major limitations of the existing database should be

specified. In such a case, establishment of an HBGV may not be appropriate if adequate genotoxicity data are available to support a WOE conclusion that the substance is mutagenic in vivo.

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

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

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

4.5.5 Approaches for evaluating data-poor substances

4.5.5.1 In silico approaches

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

---

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

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

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

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

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

(b)   Confidence in approaches

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

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

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

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

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

Table 4.5 (continued)

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

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

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

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

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

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

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

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

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

(c)   Mutagenicity assessment

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

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

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

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

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

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

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

Source: Modified from Myatt et al. (2018)

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

4.5.5.2 *Threshold of toxicological concern (TTC)*

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

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

the USFDA's threshold of regulation (USFDA, 1995), which was developed as a tool to facilitate the safety evaluation of food packaging materials, components of which have the potential to migrate into food at very low levels.

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

Chemicals are assigned to the "genotox tier" based on existing data (e.g. from mutagenicity assays) and evaluation of chemical structure. The latter is done based on the presence of structural alerts for DNA reactivity, which have been encoded in a number of software programs (e.g. Toxtree, OECD QSAR Toolbox, Derek Nexus; see section 4.5.5.1). Although this approach is generally

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

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

---

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

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

4.5.5.3 Grouping and read-across approaches

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

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

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

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

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

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

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

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

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

Read-across and (Q)SAR approaches are underpinned by the same principles and continuum of relating property or activity to a chemical structure, but boundaries between the two approaches are being challenged. (Q)SAR approaches are a more formal means of

characterizing the relationship, whereas read-across approaches tend to be more case by case, based on expert review and judgement.

4.5.6 Considerations for specific compounds

4.5.6.1 Mixtures

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

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

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

Mutagenicity testing of mixtures may apply the tiered approach recommended by EFSA (2019a). The mixture should be chemically characterized as far as possible, providing critical quantitative compositional data, including stability and batch-to-batch variability, to ensure that the test material is representative of the mixture added to food. Useful guidelines exist for the chemical characterization of botanicals (e.g. EFSA, 2009), novel foods (e.g. EFSA, 2016b) and herbal medicinal products (e.g. EMA, 2011; USFDA, 2016) and for assessing the combined exposure to multiple chemicals (e.g. Meek et al., 2011; OECD, 2018b; EFSA, 2019b). Analytical methods to identify and control mutagenic impurities and degradation products

of pharmaceuticals (e.g. Görög, 2018; Teasdale & Elder, 2018), although not directly applicable to food, could also be consulted.

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

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

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

---

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

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

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

4.5.6.2   *Flavouring agents*

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

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

manufacturing process (e.g. ethanol, edible oil, acetic acid), can be considered to be part of the flavouring.

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

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

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

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

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

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

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

- *Step 1*: Is toxicological information on the compound of interest available? If so, evaluate the available toxicological information to determine potency relative to that of the parent.

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

Fig. 4.2. Assessment scheme for the safety of plant and animal metabolites/degradation products
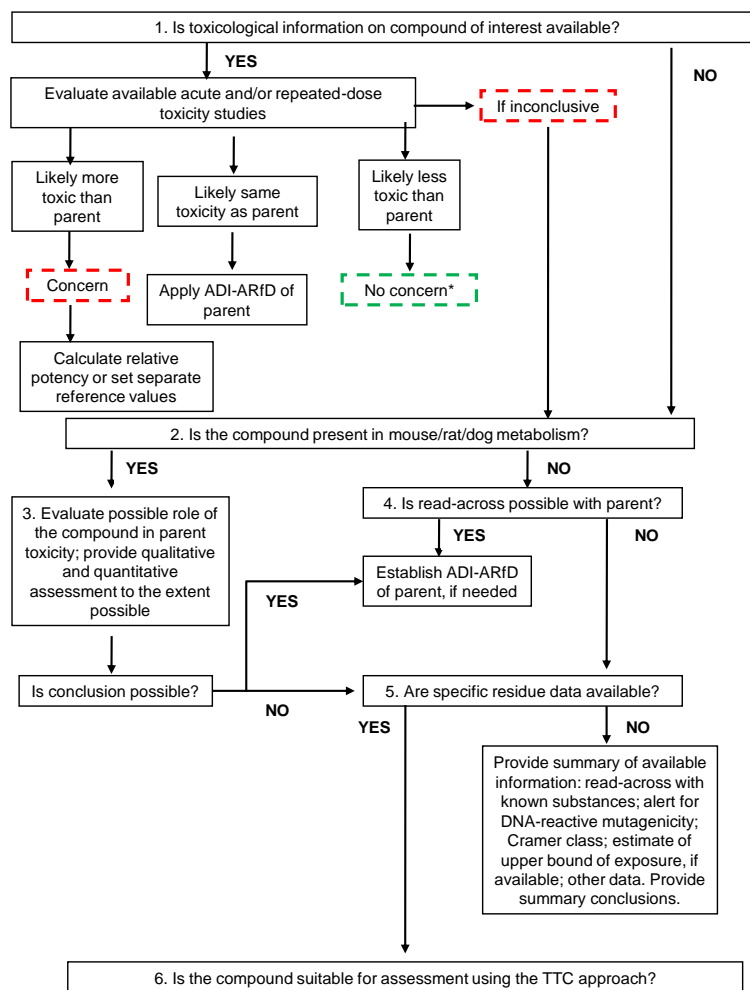
* Note: For compounds already included in residue definition.
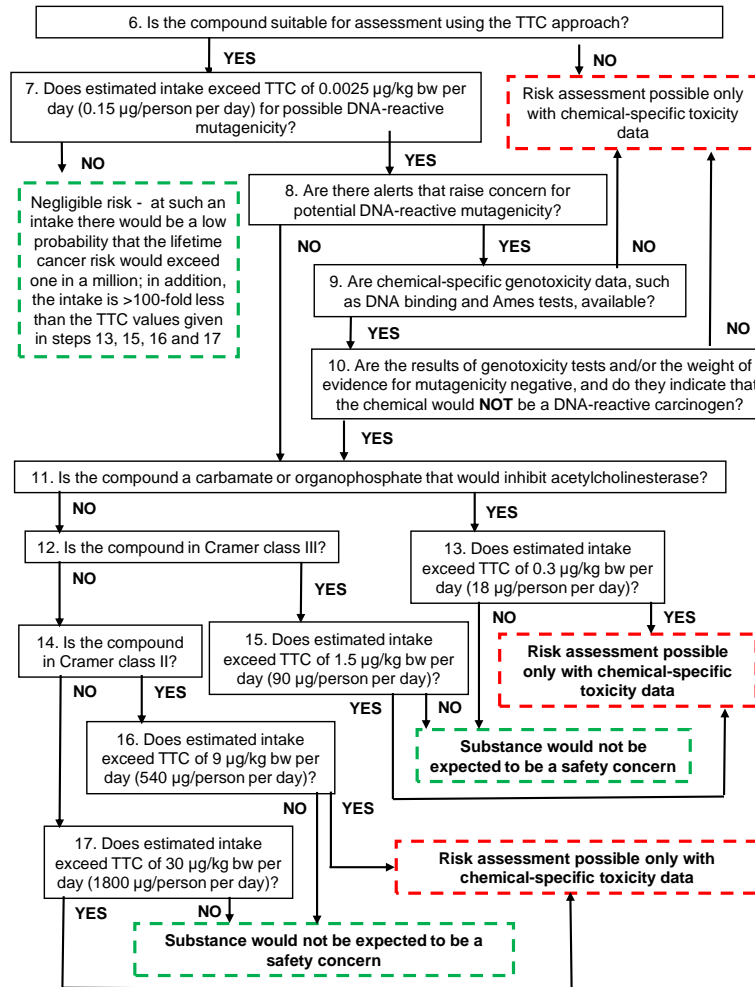
ADI: acceptable daily intake; ARfD: acute reference dose; bw: body weight; TTC: threshold of toxicological concern

Source: Adapted from WHO (2015a)

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

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

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

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

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

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

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

---

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

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

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

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

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

4.5.6.4 Secondary metabolites in enzyme preparations

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

The issue that is of interest from a safety assessment perspective is the presence of microorganism-derived secondary metabolites in

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

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

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

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

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

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

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

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

4.5.7   *Recent developments and future directions*

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

analysis that includes point of departure determinations for the interpretation of mutagenicity test data in the context of risk assessment (White & Johnson, 2016).

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

4.5.7.1 Novel in vivo genotoxicity approaches

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

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

4.5.7.2 Novel in vitro genotoxicity approaches

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

Table 4.7. Novel approaches for genotoxicity assessment

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

Table 4.7 (continued)

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

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

Table 4.7 (continued)

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

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

DDI: DNA damage–inducing; DNA: deoxyribonucleic acid; HTRF: Homogeneous Time-Resolved Fluorescence; MN: micronucleus; MOA: mode of action; RT-qPCR: reverse transcription quantitative polymerase chain reaction

noteworthy advantages related to the cost, throughput and information content of these assays, incorporation of realistic and effective xenobiotic metabolism is a concern. Nevertheless, high-throughput assays are now available to rapidly assess the induction of DNA damage and repair, gene mutations, chromosomal damage or DNA strand breaks (Table 4.7). As mutagenicity screening for regulatory purposes generally requires the assessment of gene mutations and chromosomal damage, assays that streamline detection of these end-points are particularly noteworthy. In vitro versions of the flow cytometric *Pig-a* gene mutation assay and the Transgenic Rodent Somatic and Germ Cell Mutation Assays (OECD TG 488) permit enumeration of mutations at a variety of endogenous and transgenic loci (e.g. *Pig-a*, *lacI*, *lacZ*, *cII*, *gpt*, Spi⁻). These assays do not require clonal selection and can measure mutagenicity more efficiently than, for example, traditional *Tk* and *Hprt* locus mutation assays.

Some of the high-throughput in vitro assays summarized in Table 4.7 exploit cellular pathways to rapidly measure biomarkers of DNA damage or repair; most are based on genetically engineered cell lines containing a promoter activated by genotoxic insult (e.g. p53 response element) fused to one or more reporter genes (e.g. β-lactamase). Reporter gene activation is visualized via, for example, automated micro-confocal imaging, fluorescent or luminescent readouts, or flow cytometry. Examples include the ToxTracker (Hendriks et al., 2012, 2016; Ates et al., 2016), GreenScreen (Hastwell et al., 2006; Simpson et al., 2013) and several reporter gene and antibody assays (e.g. p53RE, γH2AX, ATAD5) used by the United States Tox21 Program (https://ntp.niehs.nih.gov/whatwestudy/tox21/toolbox/index.html) or the USEPA's ToxCast Program (https://comptox.epa.gov/dashboard/chemical_lists/toxcast). Importantly, in addition to mutagenic hazard, the simultaneous or sequential examination of multiple end-points representing several distinct pathways permits delineation of the mutagenic MOA. Related assays, such as the MultiFlow DNA Damage assay, assess the presence and localization of proteins (e.g. γH2AX, nuclear p53, phospho-histone H3) indicative of DNA damage and alterations in chromosome structure or number (Bryce et al., 2016, 2017, 2018). Proteins are targeted by fluorescently labelled antibodies, and cellular phenotype is scored using flow cytometry. In addition to reporter-based approaches that track and quantify DNA damage response activation, gene expression–based strategies, such as DNA microarray, quantitative polymerase chain reaction (qPCR)

and RNA sequencing approaches, have been used as high-throughput approaches for measuring DNA damage signalling. For example, the TGx-DDI assay monitors genes involved in genomic stability (e.g. generalized stress responses, DNA repair, cell cycle control, apoptosis and mitotic signalling) to identify DNA damage–inducing (DDI) substances (Li et al., 2015, 2017; Williams et al., 2015; Yauk et al., 2016a; Corton, Williams & Yauk, 2018; Corton, Witt & Yauk, 2019). Similarly, a high-throughput real-time reverse transcription quantitative polymerase chain reaction (RT-qPCR) assay rapidly scores 95 genes active in maintaining genomic integrity (Fischer et al., 2016; Strauch et al., 2017). These reporter systems rapidly track DNA damage and repair as indirect measures of genotoxicity.

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

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

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

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

4.5.7.3   *Adverse outcome pathways for mutagenicity*

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

---

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

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

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

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

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

4.5.7.4 Quantitative approaches for safety assessment

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

---

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

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

4.5.8   References[1]

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

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

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

---

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Khanal S, Singh P, Avlasevich SL, Torous DK, Bemis JC, Dertinger SD (2018). Integration of liver and blood micronucleus and *Pig-a* gene mutation endpoints into rat 28-day repeat-treatment studies: proof-of-principle with diethylnitrosamine. Mutat Res. 828:30–5. doi:10.1016/j.mrgentox.2018.02. 005.

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

Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F et al. (2007). How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: report of an ECVAM workshop. Mutat Res. 628(1):31–55. doi:10.1016/j.mrgentox.2006.11.008.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Nohmi T, Tsuzuki T (2016). Possible mechanisms underlying genotoxic thresholds: DNA repair and translesion DNA synthesis. In: Nohmi T,

Fukushima S, editors. Thresholds of genotoxic carcinogens. From mechanisms to regulation. Tokyo: Academic Press; 49–66.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Petkov PI, Kuseva C, Kotov S, Honma M, Kitazawa A, Kulkarni S et al. (2019). Procedure for toxicological predictions based on mechanistic weight of

evidences: application to Ames mutagenicity. Comp Toxicol. 12:100009. doi:10.1016/j.comtox.2017.02.004.

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

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

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

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

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

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

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

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

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

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

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

Ritter L, Totman C, Krishnan K, Carrier R, Vézina A, Morisset V (2007). Deriving uncertainty factors for threshold chemical contaminants in drinking

water. J Toxicol Environ Health B Crit Rev. 10(7):527–57. doi:10.1080/15287390600975178.

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

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

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

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

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

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

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

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

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

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

Shah I, Liu J, Judson RS, Thomas RS, Patlewicz G (2016). Systematically evaluating read-across prediction and performance using a local validity

approach characterized by chemical structure and bioactivity information. Regul Toxicol Pharmacol. 79:12–24. doi:10.1016/j.yrtph.2016.05.008.

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

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

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

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

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

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

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

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

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

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

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

Teasdale A, Elder DP (2018). Analytical control strategies for mutagenic impurities: current challenges and future opportunities? Trends Anal Chem. 101:66–84. doi:10.1016/j.trac.2017.10.027.

Thougaard AV, Christiansen J, Mow T, Hornberg JJ (2014). Validation of a high throughput flow cytometric in vitro micronucleus assay including assessment of metabolic activation in TK6 cells. Environ Mol Mutagen. 55(9):704–18. doi:10.1002/em.21891.

Thybaud V, Dean S, Nohmi T, de Boer J, Douglas GR, Glickman BW et al. (2003). In vivo transgenic mutation assays. Mutat Res. 540(2):141–51. doi:10.1016/j.mrgentox.2003.07.004.

Ton ND, Nakagawa H, Ha NH, Duong NT, Nhung VP, Hien LTT et al. (2018). Whole genome sequencing and mutation rate analysis of trios with paternal dioxin exposure. Hum Mutat. 39(10):1384–92. doi:10.1002/humu.23585.

Torous DK, Dertinger SD, Hall NE, Tometsko CR (2000). Enumeration of micronucleated reticulocytes in rat peripheral blood: a flow cytometric study. Mutat Res. 465(1–2):91–9. doi:10.1016/s1383-5718(99)00216-8.

Tweats DJ, Johnson GE, Scandale I, Whitwell J, Evans DB (2016). Genotoxicity of flubendazole and its metabolites in vitro and the impact of a new formulation on in vivo aneugenicity. Mutagenesis. 31(3):309–21. doi:10.1093/mutage/gev070.

UKCOM (2010). Guidance statement: thresholds for in vivo mutagens. Chilton: United Kingdom Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/315698/assessment_of_threshold_for_in_vivo_mutagens.pdf).

UKCOM (2018). Statement on the quantitative approaches to the assessment of genotoxicity data. Chilton: United Kingdom Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment (Statement 2018/S1; https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/704462/COM_statement_on_quantitative_approaches.pdf).

Uno Y, Morita T, Luijten M, Beevers C, Hamada S, Itoh S et al. (2015a). Recommended protocols for the liver micronucleus test: report of the IWGT working group. Mutat Res. 783:13–8. doi:10.1016/j.mrgentox.2014.10.010.

Uno Y, Morita T, Luijten M, Beevers C, Hamada S, Itoh S et al. (2015b). Micronucleus test in rodent tissues other than liver or erythrocytes: report of the IWGT working group. Mutat Res. 783:19–22. doi:10.1016/j.mrgentox.2015.03.001.

USEPA (2019). OncoLogic™ – A computer system to evaluate the carcinogenic potential of chemicals. Washington (DC): United States Environmental Protection Agency (https://www.epa.gov/tsca-screening-tools/oncologictm-computer-system-evaluate-carcinogenic-potential-chemicals).

USFDA (1995). United States Food and Drug Administration; Food additives: threshold of regulation of substances used in food-contact articles: final rule. Fed Regist. 60(136):36582–96 (https://www.govinfo.gov/content/pkg/FR-1995-07-17/pdf/95-17435.pdf).

USFDA (2007). Toxicological principles for the safety assessment of food ingredients. Redbook 2000: IV.C.1. Short-term tests for genetic toxicity. Rockville (MD): United States Food and Drug Administration (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/redbook-2000-ivc1-short-term-tests-genetic-toxicity).

USFDA (2016). Botanical drug development: guidance for industry. Silver Spring (MD): United States Food and Drug Administration (https://www.fda.gov/downloads/Drugs/Guidances/UCM458484.pdf).

USNTP (2011). 12th report on carcinogens. Research Triangle Park (NC): United States Department of Health and Human Services, Public Health Service, National Toxicology Program (https://oehha.ca.gov/media/downloads/proposition-65/crnr/comments/12throc-complete.pdf).

Van Bossuyt M, Van Hoeck E, Raitano G, Vanhaecke T, Benfenati E, Mertens B et al. (2018). Performance of in silico models for mutagenicity prediction of food contact materials. Toxicol Sci. 163(2):632–8. doi:10.1093/toxsci/kfy057.

Verma JR, Rees BJ, Wilde EC, Thornton CA, Jenkins GJS, Doak SH et al. (2017). Evaluation of the automated MicroFlow® and Metafer™ platforms for high-throughput micronucleus scoring and dose response analysis in human lymphoblastoid TK6 cells. Arch Toxicol. 91(7):2689–98. doi:10.1007/s00204-016-1903-8.

Vuorinen A, Bellion P, Beilstein P (2017). Applicability of in silico genotoxicity models on food and feed ingredients. Regul Toxicol Pharmacol. 90:277–88. doi:10.1016/j.yrtph.2017.09.026.

Wang NCY, Zhao QJ, Wesselkamper SC, Lambert JC, Petersen D, Hess-Wilson JK (2012). Application of computational toxicological approaches in human health risk assessment. I. A tiered surrogate approach. Regul Toxicol Pharmacol. 63(1):10–9. doi:10.1016/j.yrtph.2012.02.006.

Wetmore BA (2015). Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. Toxicology. 332:94–101. doi:10.1016/j.tox.2014.05.012.

White PA, Johnson GE (2016). Genetic toxicology at the crossroads – from qualitative hazard evaluation to quantitative risk assessment. Mutagenesis. 31(3):233–7. doi:10.1093/mutage/gew011.

White PA, Luijten M, Mishima M, Cox JA, Hanna JN, Maertens RM et al. (2019). In vitro mammalian cell mutation assays based on transgenic reporters:

a report of the International Workshop on Genotoxicity Testing (IWGT). Mutat Res. 847:403039. doi:10.1016/j.mrgentox.2019.04.002.

WHO (2015a). Pesticide residues in food: Guidance document for WHO monographers and reviewers. Geneva: World Health Organization (https://www.who.int/foodsafety/publications/jmpr_guidance_document_1.pdf?ua=1).

WHO (2015b). Pesticide residues in food – 2014 evaluations. Part II – Toxicological. Geneva: World Health Organization (https://apps.who.int/iris/bitstream/handle/10665/164597/9789241665308_eng.pdf;jsessionid=65B92CA6D1D058F62646E2BCBBAB888D?sequence=1).

WHO (2016). Pesticide residues in food – 2015 evaluations. Part II – Toxicological. Geneva: World Health Organization (http://apps.who.int/iris/bitstream/10665/205165/1/9789241655316_eng.pdf?ua=1).

WHO (2017). Guidance to JECFA experts on systematic literature searches. Geneva: World Health Organization, JECFA Secretariat (https://www.who.int/foodsafety/chem/jecfa/Litertature_Search.pdf).

WHO (2019). Pesticide residues in food – 2018 evaluations. Part II – Toxicological. Geneva: World Health Organization (https://apps.who.int/iris/bitstream/handle/10665/325787/9789241655330-eng.pdf?ua=1).

Wichard J (2017). In silico prediction of genotoxicity. Food Chem Toxicol. 106(Part B):595–9. doi:10.1016/j.fct.2016.12.013.

Williams A, Buick JK, Moffat I, Swartz CD, Recio L, Hyduke DR et al. (2015). A predictive toxicogenomics signature to classify genotoxic versus non-genotoxic chemicals in human TK6 cells. Data Brief. 5:77–83. doi:10.1016/j.dib.2015.08.013.

Williams JG, Deschl U, Williams GM (2011). DNA damage in fetal liver cells of turkey and chicken eggs dosed with aflatoxin B_1. Arch Toxicol. 85:1167–72. doi:10.1007/s00204-011-0653-x.

Wills JW, Johnson GE, Doak SH, Soeteman-Hernández LG, Slob W, White PA (2016). Empirical analysis of BMD metrics in genetic toxicology. Part I: In vitro analyses to provide robust potency rankings and support MOA determinations. Mutagenesis. 31(3):255–63. doi:10.1093/mutage/gev085.

Witt KL, Knapton A, Wehr CM, Hook GJ, Mirsalis J, Shelby MD et al. (2000). Micronucleated erythrocyte frequency in peripheral blood of $B6C3F_1$ mice from short-term, prechronic, and chronic studies of the NTP carcinogenesis bioassay program. Environ Mol Mutagen. 36(3):163–94. doi:10.1002/1098-2280(2000)36:3<163::AID-EM1>3.0.CO;2-P.

Witt KL, Cunningham CK, Patterson KB, Kissling GE, Dertinger SD, Livingston E et al. (2007). Elevated frequencies of micronucleated erythrocytes in infants exposed to zidovudine in utero and postpartum to prevent mother-to-child

transmission of HIV. Environ Mol Mutagen. 48(3–4):322–9. doi:10.1002/em. 20266.

Witt KL, Livanos E, Kissling GE, Torous DK, Caspary W, Tice RR et al. (2008). Comparison of flow cytometry– and microscopy-based methods for measuring micronucleated reticulocyte frequencies in rodents treated with nongenotoxic and genotoxic chemicals. Mutat Res. 649(1–2):101–13. doi:10.1016/j. mrgentox.2007.08.004.

Witt KL, Hsieh JH, Smith-Roe SL, Xia M, Huang R, Zhao J et al. (2017). Assessment of the DNA damaging potential of environmental chemicals using a quantitative high-throughput screening approach to measure p53 activation. Environ Mol Mutagen. 58:494–507. doi:10.1002/em.22112.

Wolf T, Luepke N-P (1997). Formation of micronuclei in incubated hen's eggs as a measure of genotoxicity. Mutat Res. 394(1–3):163–75. doi:10.1016/ S1383-5718(97)00136-8.

Wolf T, Niehaus-Rolf C, Luepke N-P (2003). Investigating genotoxic and hematotoxic effects of *N*-nitrosodimethylamine, *N*-nitrosodiethylamine and *N*-nitrosodiethanolamine in the hen's egg micronucleus test (HET-MN). Food Chem Toxicol. 41(4):561–73. doi:10.1016/S0278-6915(02)00281-8.

Worth A, Fuart-Gatnik M, Lapenna S et al. (2010). Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Scientific report submitted to the European Food Safety Authority. Ispra: European Commission, Joint Research Centre, Institute for Health & Consumer Protection (https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/ sp.efsa.2010.EN-50).

Wu S, Blackburn K, Amburgey J, Jaworska J, Federle T (2010). A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. Regul Toxicol Pharmacol. 56(1):67–81. doi:10.1016/j.yrtph.2009.09.006.

Yamamoto KN, Hirota K, Kono K, Takeda S, Sakamuru S, Xia M et al. (2011). Characterization of environmental chemicals with potential or DNA damage using isogenic DNA repair–deficient chicken DT40 cell lines. Environ Mol Mutagen. 52(7):547–61. doi:10.1002/em.20656.

Yao C, Feng YL (2016). A nontargeted screening method for covalent DNA adducts and DNA modification selectivity using liquid chromatography–tandem mass spectrometry. Talanta. 159:93–102. doi:10.1016/j.talanta.2016.05.074.

Yauk CL (2004). Advances in the application of germline tandem repeat instability for in situ monitoring. Mutat Res. 566(2):169–82. doi:10.1016/ j.mrrev.2003.08.001.

Yauk CL, Buick JK, Williams A, Swartz CD, Recio L, Li HH et al. (2016a). Application of the TGx-28.65 transcriptomic biomarker to classify genotoxic and non-genotoxic chemicals in human TK6 cells in the presence of rat liver S9. Environ Mol Mutagen. 57(4):243–60. doi:10.100/Fem.22004.

Yauk C, Lambert I, Marchetti F, Douglas G (2016b). Adverse outcome pathway on alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations. Paris: OECD Publishing (OECD Series on Adverse Outcome Pathways, No. 3; https://doi.org/10.1787/5jlsvvxn1zjc-en).

Youssoufian H, Pyeritz RE (2002). Mechanisms and consequences of somatic mosaicism in humans. Nat Rev Genet. 3:748–58. doi:10.1038/nrg906.

Yu Y, Wang P, Cui Y, Wang Y (2018). Chemical analysis of DNA damage. Anal Chem. 90(1):556–76. doi:10.1021/acs.analchem.7b04247.

Zeiger E (2004). History and rationale of genetic toxicity testing: an impersonal, and sometimes personal, view. Environ Mol Mutagen. 44(5):363–71. doi: 10.1002/em.20062.

Zhu H, Bouhifd M, Donley E, Egnash L, Kleinstreuer N, Kroese DE et al. (2016). Supporting read-across using biological data. ALTEX. 33(2):167–82. doi:10.14573/altex.1601252.