WORLD HEALTH ORGANIZATION
ORGANISATION MONDIALE DE LA SANTE

EHC240: Principles and Methods for the Risk Assessment of Chemicals in Food

SUBCHAPTER 4.5. Genotoxicity

Draft 12/12/2019

Deadline for comments 31/01/2020

The contents of this restricted document may not be divulged to persons other than those to whom it has been originally addressed. It may not be further distributed nor reproduced in any manner and should not be referenced in bibliographical matter or cited.

Le contenu du présent document à distribution restreinte ne doit pas être divulgué à des personnes autres que celles à qui il était initialement destiné. Il ne saurait faire l'objet d'une redistribution ou d'une reproduction quelconque et ne doit pas figurer dans une bibliographie ou être cité.

4.5 Genotoxicity 3

4.5.1 Introduction 3
4.5.1.1 Risk Analysis Context and Problem Formulation 5
4.5.2 Tests for genetic toxicity 14
4.5.2.2 Bacterial mutagenicity 18
4.5.2.2 In vitro mammalian cell mutagenicity 18
4.5.2.3 In vivo mammalian cell mutagenicity 20
4.5.2.4 In vitro chromosomal damage assays 22
4.5.2.5 In vivo chromosomal damage assays 23
4.5.2.6 In vitro DNA damage/repair assays 24
4.5.2.7 In vivo DNA damage/repair assays 25
4.5.3 Interpretation of test results 26
4.5.3.1 Identification of relevant studies 27
4.5.3.2 Presentation and categorization of results 30
4.5.3.3 Weighting and integration of results 36
4.5.3.4 Adequacy of the genotoxicity database 39
4.5.3.5 Mutagenic mode of action and carcinogenicity 40
4.5.3.6 Integration of carcinogenicity and genotoxicity 43
4.5.4 Special Considerations 45
4.5.4.1 In silico approaches for data-poor substances 45
4.5.4.2 Threshold of Toxicological Concern 52
4.5.4.3 Grouping and read-across approaches 54
4.5.5 Considerations for Specific Compounds 57
4.5.5.1 Mixtures 57
4.5.5.2 Flavouring Agents 59
4.5.5.3 Minor Constituents 61
4.5.5.4 Secondary Metabolites in Enzyme Preparations 66
4.5.6 Recent Developments and Future Directions 68
4.5.6.1 Novel in vivo genotoxicity approaches 69
4.5.6.2 Novel in vitro genotoxicity approaches 69
4.5.6.3 Adverse outcome pathways for genotoxicity 75
4.5.6.4 Quantitative utility for safety assessment 76
4.5.7 References 78

4.5 Genotoxicity

4.5.1 Introduction

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

The purpose of genotoxicity testing is to identify substances that can cause genetic alterations in somatic and/or germ cells and this information is used in regulatory decisions (OECD, 2016). 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 4.5.3.4). A conclusion on the genotoxic potential of a chemical, and more specifically and perhaps importantly, a mutagenic MOA for carcinogenicity, can be made on the basis of only a few specific types of evidence from properly conducted and well-reported studies. Moreover, a chemical could be acknowledged as having genotoxic potential but low concern for a mutagenic MOA in its carcinogenicity.

Some agencies, such as those within the United States, Canada, United Kingdom, Japan and the European Union consider heritable mutation a regulatory endpoint. Mutations in germ cells may be inherited by future generations and may contribute to genetic disease. Germline or somatic cell mutations are implicated in the etiology of disease states, such as, cancer, sickle cell anemia, and neurological diseases (Youssoufian and Pyertitz, 2002; Lupsky, 2013). Inherited mutations

linked to human diseases are compiled in the Human Gene Mutation Database (HGDB, 2017).

The overview presented in this chapter focuses on the identification of mutagens and genotoxic carcinogens, consistent with the WHO/IPCS harmonized scheme for mutagenicity testing (Eastmond et al., 2009). The term 'mutation' refers to permanent changes in the structure and/or amount of the genetic material of an organism that can lead to heritable changes in its function, and these mutations include gene mutations as well as structural and numerical chromosome alterations. The term "mutagens" refers to chemicals that induce heritable genetic changes, most commonly through interaction with DNA[1]. The broader term of 'genotoxicity' includes mutagenicity but also includes DNA damage which may be reversed by DNA repair 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, 2015). Thus, genotoxicity tests also include those that measure the capability of substances to damage DNA and/or cellular components regulating the fidelity of the genome—such as the spindle apparatus, topoisomerases, DNA repair systems and DNA polymerases—and includes a broad range of adverse effects on genetic components of the cell. Therefore, the broader term "genotoxicant" refers to chemicals that induce adverse effects on genetic components via a variety of mechanisms including mutation. Testing for genotoxicity should utilize internationally-recognized protocols, where they exist. For example, mutagenicity (gene mutation and chromosomal aberration/damage assays) 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 genotoxicity are generally based on a combination of tests to assess three major endpoints of genetic damage associated with human disease:

1. Gene mutation (i.e. point mutations or deletions/insertions that affect single or blocks of genes)

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

2. Clastogenicity (i.e. structural chromosome changes)
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; whereas there are scenarios such as minor plant and animal metabolites of pesticides or veterinary drugs that often lack empirical data, or enzyme preparations used in food production that are not single chemicals, but rather are mixtures that include proteins and one or more low molecular weight chemicals. Special considerations related to these scenarios, including the genotoxicity evaluation of food extracts obtained from natural sources, which are often complex botanical mixtures that may not be fully characterized, are also discussed in this chapter.

4.5.17.1 Risk Analysis Context and Problem Formulation

Identification of compounds that may lead to cancer 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 where exposure is unavoidable (e.g. contaminants, natural constituents of the diet). In practice, this impacts the nature of information provided to risk managers. For substances intentionally added/used that require regulatory approval, key outputs of the hazard characterisation are health-based guidance values (HBGV) (e.g. ADI, 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 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 cancer. Consequently, for carcinogens considered to act through a mutagenic MOA, it is often not possible to establish with confidence a HBGV below which concern is considered negligible. In the context of the work of JECFA and JMPR, it is generally understood that it is unacceptable to intentionally add low doses of

mutagenic carcinogens to food, and hence it is inappropriate to establish a HBGV for these substances. 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 for carcinogenicity and, if so, is there a concern at estimated dietary exposures?

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

- *in silico* and (quantitative)structure-activity relationship [(Q)SAR] models (section 4.5.4.1)
- Threshold of Toxicological Concern (TTC; section 4.5.4.2)
- Read-across (section 4.5.4.3)

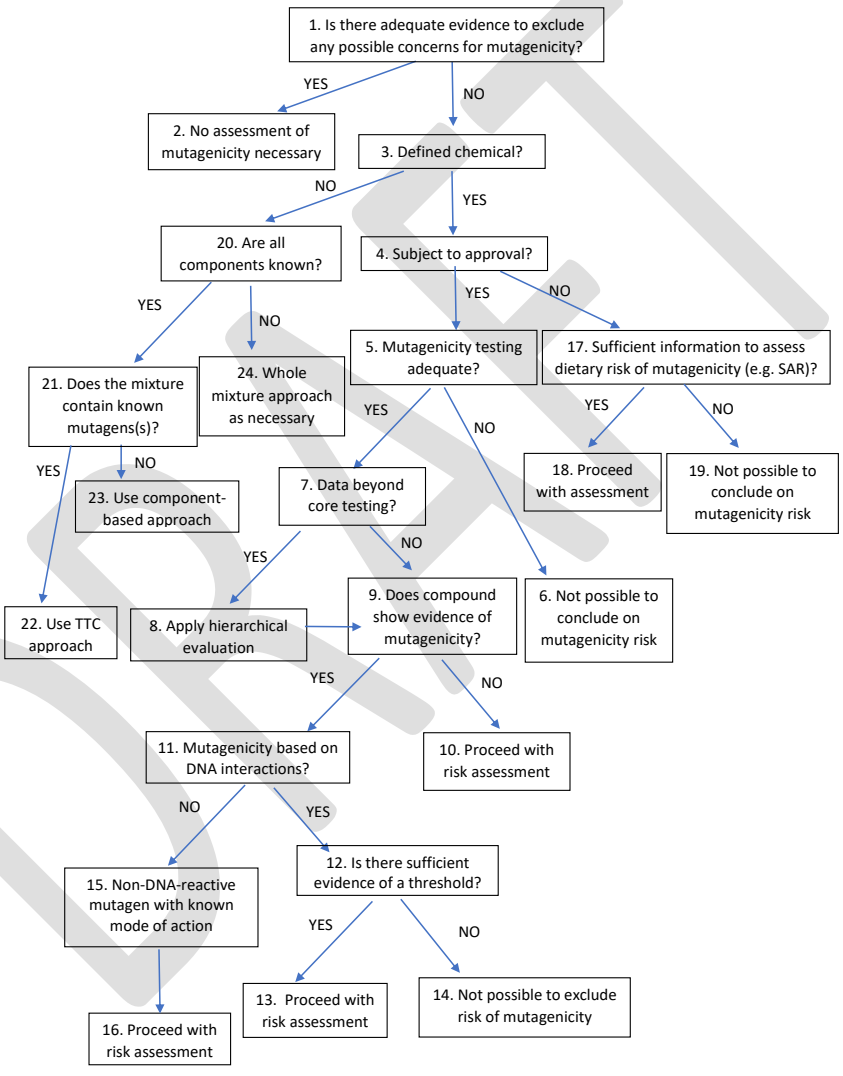
In the future, quantitative dose-response approaches for genotoxicity may also be appropriate for unavoidable contaminants (section 4.5.7). However, since this is a deviation from current practice, acceptability of such approaches should be indicated in the problem formulation (see, for example UK COM, 2018; MacGregor et al., 2015a, b).

The joint expert committees of FAO and WHO (e.g. JECFA, JMPR) do not set data requirements for their food additive and pesticide residue evaluations, although there is a minimum data set expected on which to conduct an assessment. In the case of genotoxicity, the nature of and guidance to interpret the information is described in this chapter. In

general, the joint committees evaluate the available data, most often generated in support of regulatory submissions elsewhere. Data requirements can vary substantially depending upon the use and potential for substantial human exposure. Figure 4.1 is a decision tree illustrating issues to be considered in assessing the genotoxic potential of different types of substances that can be found in food. Subsequent sections of this chapter will describe the process of identifying relevant and reliable genotoxicity data, and depending on the regulatory jurisdiction, determining whether the data and weight of evidence (WOE) are adequate to conclude on genotoxicity potential. If a substance is shown to possess genotoxic potential, the process of discerning the likelihood of a mutagenic MOA for carcinogenicity is also discussed, in conjunction with cancer bioassay data, if available.

DRAFT

Figure 4.1 Decision tree illustrating issues to be considered in assessing the mutagenic potential (as defined in this document) of different types of substances that can be found in food



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

While it may be rare to exclude possible concerns for genotoxicity *a priori*, occasionally the nature of the substance or its production process may provide sufficient assurance that substance-specific genotoxicity data are not necessary. One example is a natural constituent of the diet produced by a fully-controlled process, e.g. invertase derived from *S. cerevisiae* fermentation (JECFA, 2002). [See section 4.5.5.1]

2. ***No assessment of genotoxicity necessary***

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

3. ***Defined chemical?***

If the answer to question 1 is NO, does the substance comprise a single or a small number (e.g. stereoisomers) of chemicals of known structure, i.e. 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 a chemical that nominally is a single substance is so impure that it is considered a mixture with uncharacterized constituents [e.g., <90% purity; See section 4.5.5.1].

4. ***Subject to approval?***

If the substance is chemically defined (the answer to Q3 is YES), does it require regulatory approval in member countries prior to uses that could knowingly result in its presence in food (i.e. pesticides, veterinary drugs, food additives, food flavourings)? Excluded are contaminants and natural constituents of the diet, e.g. mycotoxins. [See section 4.5.1.1].

5. ***Genotoxicity testing adequate?***

For substances subject to regulatory approval in some jurisdictions (the answer to Q4 is YES), determine if available data are adequate to conclude whether the substance is likely to pose a genotoxic risk *in vivo* at dietary levels of exposure. [See sections 4.5.2 and 4.5.3.5].

6. *Not possible to conclude on genotoxicity risk*

If genotoxic potential has not been adequately tested (the answer to Q5 is NO), it is not possible to conclude on the likelihood of genotoxic risk *in vivo* at dietary levels of exposure. As such, it may be inappropriate to establish health-based guidance values that encompass potential genotoxicity. The main data gaps precluding a conclusion on genotoxic potential should be clearly articulated. [See section 4.5.3.6].

7. Data beyond core testing?

For some compounds, particularly newer ones, genotoxicity testing may be adequate (the answer to Q5 is YES) based on available data from a small range of relevant and reliable "standard" genotoxicity tests. However, for others, particularly those in use for some time and/or about which there are specific concerns, the available data may be much more extensive and include 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 4.5.3.1].

8. Apply hierarchical evaluation

When the genotoxicity database is complex and/or contradictory (the answer to Q7 is YES), a WOE approach that considers factors such as results of *in vivo* versus *in vitro* testing, relevance of the test/endpoint to humans and relevance of the route of exposure and dose, is used to weight the studies. [See sections 4.5.3.2 and 4.5.3.3].

9. Does compound show evidence of genotoxicity?

Regardless of how extensive the database is (the answer to Q7 is NO or after application of Q8), a WOE conclusion should be reached on whether the substance shows evidence of genotoxicity for relevant endpoint(s). For example, an isolated positive result at high, cytotoxic concentrations *in vitro* without evidence of genotoxicity in numerous guidelines studies conducted to an appropriate standard is insufficient to conclude that overall there is concern for genotoxicity. 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 4.5.3.1, 4.5.3.2, 4.5.3.3].

10. *Proceed with risk assessment*

If the WOE does not suggest genotoxicity (the answer to Q9 is NO), no further consideration of the genotoxic potential of the substance is necessary and risk assessment of non-genotoxic effects can proceed. [See other sections of chapter 4].

11. *Genotoxicity based on DNA interactions?*

If there is evidence of genotoxicity (the answer to Q9 is YES), the nature of the genotoxicity should be determined, specifically whether the genotoxicity is based on the parent or a metabolite interacting with DNA, thereby resulting in heritable DNA changes. This evidence should come primarily from appropriate tests for gene mutation and clastogenicty/ aneugenicity 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 evidence of a threshold?*

For a genotoxic chemical (the answer to Q11 is YES), the relevance of the dose/concentration used in testing to the estimated dietary exposure should be considered. For the vast majority of mutagens, there may be or no evidence for an effect threshold. Hence, lacking such evidence it is assumed that high-dose effects are relevant for assessing genotoxic potential in humans. However, for a few substances, there is clear evidence *in vitro* and *in vivo* for a threshold. Hence, in theory, it may be possible to discount effects seen only at doses that are irrelevant to conceivable human exposure (or even a multiple of that exposure). [See section 4.5.5.5].

13. *There is sufficient evidence for a threshold for genotoxicity*

If it is concluded that the mutagenicity observed experimentally is not relevant considering conceivable human exposure levels (the answer to Q12 is YES), risk assessment based on the critical effect(s) can proceed. [See other sections of chapter 4].

.

DRAFT

14. ***There is insufficient evidence for a threshold for genotoxicity***

If it is concluded that the mutagenicity observed experimentally is relevant considering conceivable human exposure levels (the answer to Q12 is NO), it will be inappropriate to establish health-based guidance values [See section 4.5.3.6].

15. ***Non-DNA-reactive genotoxicant with known mode of action***

For genotoxic compounds in which a DNA-reactive MOA can be excluded (the answer to Q11 is NO), the nature of the genotoxicity, 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 genotoxicity secondary to inflammation that generates reactive oxygen species. [See section 4.5.3.4].

16. ***Proceed with risk assesment***

The output of the genotoxic hazard characterisation (output from Q15) can be used in the risk assessment, as appropriate. For example, if genotoxicity is considered to exhibit a threshold, the 'normal' approach to establishing health-based guidance values and to risk characterisation can be applied. In many cases, this would mean that the critical effect was other than genotoxicity, as it occurred at lower exposure levels. A concluding statement regarding potential risk of genotoxicity *in vivo* at dietary levels of exposure should be provided. [See section 4.5.3.6].

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

For substances not subject to regulatory approval (the answer to Q4 is NO) that have unavoidable dietary exposure, such as contaminants or natural dietary constituents (e.g. mycotoxins), assess whether there is sufficient information to reach a conclusion about potential genotoxicity. When insufficient empirical genotoxicity data exist to reach a conclusion, additional information from related analogues (i.e. read-across) and/or from *in silico* approaches, such as (Q)SARs, should also be considered in an overall WOE for the genotoxic potential of the substance. [See sections 4.5.5.1, 4.5.4.3].

18. ***Proceed with risk assessment***

Where sufficient information is available to conclude on the genotoxic potential of the substance (the answer to Q17 is YES), a risk assessment can proceed. This may justify establishing a health-based guidance value, such as a tolerable daily intake, or the use of a margin of exposure (MOE) approach. Where exposures are likely to be very low, and the compound is a potential mutagen, the threshold of toxicological concern (TTC) approach can be used. If exposure is below the genotoxicity TTC tier (0.0025 ug/kg/day for chemicals with structural alerts for DNA reactivity), there is low concern for effects on human health [See sections 4.5.4.2 and other sections of chapter 4].

19. ***Not possible to conclude on genotoxicity risk***

Where it is not possible to conclude on potential genotoxicity (the answer to Q17 is NO), advice should be provided on the assumption that the substance might be a mutagen. Hence, the TTC for such compounds could be used, recognizing the health-protective nature and considerable uncertainty in the assessment. 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 comprised of one (or a small number of) defined chemical entities (the answer to Q3 is NO), a 'whole mixture' approach should be applied tailored to whether the mixture is simple ((i.e., all components are known and have established chemical structures and concentrations) or complex (a significant fraction of components are of unknown structure and/or concentration). [See section 4.5.6.1].

While there is no explicit question in the decision tree as to whether mixtures are subject to approval, a number of the considerations for defined chemicals will apply to mixtures. That is, for those mixtures subject to approval, consideration will need to be given to the adequacy of genotoxicity 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.

DRAFT

21. ***Does the mixture contain known genotoxicant(s)?***

Where all of the components in a mixture above a certain level are known (the answer to Q20 is YES), each component should be assessed for its genotoxicity, on the basis of prior knowledge. If one or more known mutagens are present, these should be assessed before considering the potential genotoxicity of other components.

22. ***Use TTC approach***

For mutagenic substances present in a defined mixture (the answer to Q21 is YES), the TTC approach can be applied. If estimated human exposure is below the genotoxicity TTC, there is low concern for genotoxicity in exposed individuals from these substances and the remaining components can then be assessed individually, as described under the component-based approach. If the estimated exposure exceeds the genotoxicity TTC, additional information will be needed to determine if there is concern for possible genotoxicity in exposed individuals. [See section 4.5.5.3].

23. ***Use component-based approach***

For a "simple mixture' in which all of the components are known (the answer to Q21 is NO), each component should be assessed for potential genotoxicity, as described for defined chemicals. [See section 4.5.5.1].

24. ***Use whole mixture approach as necessary***

For a 'complex mixture' in which a significant fraction of a mixture is unknown (the answer to Q20 is NO), extracts, sub-fractions or the whole mixture should be tested for genotoxicity, depending on the nature of the mixture, information available, and its intended use. [See section 4.5.5].

4.5.2 Tests for genetic toxicity

More than 100 different *in vitro* and *in vivo* test methods exist. Given the high degree of overlap, a much smaller number of formally validated methods with OECD test guidelines are commonly used (Table 4.1) and can be grouped according to the test system (e.g. *in vitro* or *in vivo*) and the genetic endpoint assessed for genetic damage:

<u>Gene mutations:</u>

- Gene mutation in bacteria;

- Gene mutation in mammalian cell lines;
- Gene mutation in rodents *in vivo* using constitutive or transfected genes

Clastogenicity and aneuploidy:

- Chromosomal aberrations in cultured mammalian cells (to assess structural changes);
- Micronucleus (MN) induction in cultured mammalian cells (to assess chromosome breaks and aneuploidy);
- Chromosomal aberration *in vivo* in mammalian haematopoietic cells (to assess structural and numerical chromosome changes)
- 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)
- Endpoints 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, unscheduled DNA synthesis in liver cells).

Complete consistency among the results of different classes of assays is generally not expected, as they measure different endpoints. In addition to the commonly used tests in Table 4.1, there are numerous methods with more limited validation, e.g., where yeast, molds, and insects (*Drosophila*) are used as test organisms. Tests for cell transformation *in vitro* may also be useful, however, positive results are not necessarily indicative of DNA-reactive
genotoxicity, but, rather, may represent epigenetic events (any heritable influence in the progeny of cells or of individuals on chromosome or gene function that does not change DNA nucleotide sequence).

Identification of germ cell mutagens is difficult, and quantitative studies in rodents require 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 and vice

DRAFT

versa. 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. 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 standpoint of JMPR and JECFA at this time, as information on developmental and reproductive toxicity would often be available (particularly for chemicals subject to authorization in member states).

DRAFT

Table 4.1 Examples of assays for genetic toxicity

Gene mutation	Chromosome damage	DNA damage/repair
***In vitro* assays** *Bacterial tests* Reversion to a specific nutrient independence in: • *Salmonella typhimurium and Escherichia coli* (OECD 471) *Mammalian tests* Forward mutation at the hypoxanthine-guanine phosphoribosyl transferase (*hprt*) gene (OECD 476) in cell lines such as: • Chinese hamster ovary (CHO) • Chinese hamster lung (V79) • Human lymphocytes Forward mutation at the thymidine kinase (*tk*) gene (OECD 490) in cell lines such as: • Mouse lymphoma L5178Y • Human TK6 ***In vivo* assays** Somatic cell assays: • *gpt,* Spi$^-$ (*gpt* delta mouse or rat), *LacZ or cII* (Muta™Mouse) or *lacI or cII* (Big Blue® mouse or rat) (OECD 488) • *Pig-a* gene mutation assay Germline cell assays: • Specific locus test in mice • *gpt,* Spi$^-$ (*gpt* delta mouse or rat), *LacZ or cII* (Muta™Mouse) or *lacI or cII* (Big Blue® mouse or rat) (OECD 488)	***In vitro* assays** Sister chromatid exchange (SCE) (OECD TG 479)* Chromosomal aberrations (OECD 473); Micronuclei (resulting from clastogenicity and aneuploidy) (OECD 487) in: • CHO or V79 cell lines and human cells (lymphocytes; TK6) ***In vivo* assays** Somatic cell assays: • SCE (OECD TG 482)* in bone marrow (rodent) • Chromosomal aberrations (OECD 475); Micronuclei (resulting from clastogenicity and aneuploidy) (OECD 474) in erythrocytes or lymphocytes (rodent) Germline cell assays: • Chromosomal aberrations (OECD 483), heritable translocations (OECD 485) and dominant lethal mutations (OECD 478) in mice or rats	***In vitro* assays** • DNA adduct measurement in cell cultures • Unscheduled DNA synthesis in primary cultures (often hepatocytes; OECD 482)* • 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 • Upregulation or stabilization of DNA damage responses (e.g. p53, ATAD5pH2AX) ***In vivo* assays** • DNA adduct measurement • Unscheduled DNA synthesis (usually in liver; OECD 486) • Strand breakage and alkali-labile sites monitored by single-cell gel electrophoresis (comet assay) in nuclear DNA of any tissue type (OECD 489)

*OECD Test Guidelines 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

DRAFT

4.5.2.2 Bacterial mutagenicity

As one of the original genotoxicity assays 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 *S. typhimurium* that carry different mutations in various genes of the histidine operon, and *E. 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 frame-shift 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, e.g., liver S9-mix from Aroclor 1254 or phenobarbital/5,6-benzoflavone-induced rats.

Negative results in the assay do not necessarily demonstrate non-carcinogenicity, as cancer can also be induced by other mechanisms such as clastogenicity, epigenetic or non-genotoxic modes of action. Conversely, positive results do not always predict carcinogenicity, as some bacterial mutagens are negative in *in vitro* mammalian mutagenicity tests and in carcinogenicity bioassays (Kirkland et al., 2014).

4.5.2.2 In vitro mammalian cell mutagenicity

Currently, two *in vitro* assays for mammalian gene mutation induction have formal OECD Test Guidelines:

In Vitro Mammalian Cell Gene Mutation Tests Using the Thymidine Kinase Gene

The mouse lymphoma TK assay (OECD TG 490) detects mutagenic and clastogenic events at the thymidine kinase (*Tk*) locus of L5178Y mouse lymphoma $Tk^{+/-}$ cells (Lloyd and 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 (TFT), which becomes incorporated into DNA during cell replication, but the analogue cannot be incorporated into the DNA of mutated $Tk^{-/-}$ cells, which survive and form colonies: large colonies indicate gene mutation (point mutations or base deletions that do not affect the rate of cell doubling), whereas small colonies indicate chromosomal mutation (chromosomal rearrangements or translocations that result in extended cell doubling times). Similarly, $TK^{-\backslash -}$ mutants in TK6 cells can be selected with TFT and early appearing and late appearing colonies indicate gene mutation and chromosome mutation, respectively.

In Vitro Mammalian Cell 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, while $HPRT^{+}$ 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., CHO, V79, L5178Y, TK6), while CHO-derived AS52 cells containing the gpt transgene (and having the Hprt gene deleted) are used for the XPRT test (OECD 476), either 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 mutation frequencies at the named genes in mammalian cells following chemical exposure, but each genetic target detects a different spectrum of mutational events. Mutational frequency is measured by counting mutant colonies arising on plates with selective medium. The mouse lymphoma assay (TG 490) is

DRAFT

used more commonly than the Hprt/Xprt assay (TG 476) since it can detect a broader range of mutagenic events.

4.5.2.3 *In vivo mammalian cell mutagenicity*

Somatic cell assays

***gpt* delta mouse or rat, Muta™Mouse and Big Blue® mouse or rat:**

The OECD TG 488 assays employ transgenic mice or rats harboring lambda phage DNA carrying reporter gene mutations in all cells (Nohmi et al., 2000, 2017; Thybaud et al., 2003). After chemical treatment, the transgenes are rescued from 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. This assay is advantageous for further evaluation of rodent carcinogens because gene mutations can be detected in any organ or tissue, aiding evaluation of the target organs for carcinogenesis and the ability to distinguish DNA-reactive genotoxic carcinogens from DNA-non-reactive (or non-genotoxic) carcinogens. For substances lacking standard carcinogenicity bioassays, transgenic assays such as with *gpt*, *lacI*, *lacZ* and *cII* that detect point mutations (base substitution or frameshift) and the Spi$^-$ that detect deletion mutations can be integrated into 28-day repeat dose toxicity studies with other genotoxicity assays such as the *in vivo* MN assay, *Pig-a* assay, or comet assay. DNA sequencing of mutants can be useful to examine chemical MOA by comparing the mutation spectrum with those of other known mutagens and to eliminate duplicate mutants generated by clonal expansion of single mutants.

***Pig-a* assay in rats or mice:**

This assay uses the constitutive phosphatidylinositol glycan complementation group A gene (*Pig-a*) 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 glycosylphosphatidyl inositol-anchor proteins in the cell surface and thus the mutant cells fail to express the CD59 antigen and be labeled by antibodies targeting this antigen. The absence of CD59

proteins on the cell surface, which is easily detected by flow cytometry, is a direct reporter of *Pig-a* mutation. The assay is rapid and low cost, using only a small volume of blood, and can be conveniently integrated into rodent 28-day repeat dose toxicity studies along with other genotoxicity assays (Khanals 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), although other cell types are being investigated for suitability, such as T-lymphocytes. An adapted version of the Pig-a assay amenable to high-throughput (HT) scoring is described in section 4.5.6.7.

Germ cell assays

Mouse specific locus test (SLT)

The SLT test for mutagenicity in germ cells is rarely used due to cost and the large number of animals needed (Russell and Shelby, 1985). In a typical SLT, 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 color. 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 SLT and several chemicals (e.g., ethyl nitrosourea) were detected as mutagenic in spermatogonial stem cells (Shelby, 1996). A novel approach such as Trio analysis where direct comparison of DNA sequences is made between parent and offspring (Ton et al., 2018; Masumura, 2016a,b), the expanded simple tandem repeats assay (Yauk, 2004), or the transgenic assays described below, may also detect germ cell mutations.

DRAFT

Rodent dominant lethal assay

The dominant lethal assay investigates whether a chemical induces mutations resulting in embryo or fetal death. The mutations originate primarily from chromosomal aberrations in germ cells (OECD 478). While 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 genetic toxicity studies.

***gpt* delta mouse or rat, Muta™Mouse and Big Blue® mouse or rat**

The OECD TG 488 transgenic assays are also applicable to 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.

4.5.2.4 In vitro chromosomal damage assays

Chromosomal aberration assay

The *in vitro* chromosomal aberration assay (OECD TG 473) assesses chemical-induced chromosomal structural damage in cultured mammalian cells (e.g., Chinese hamster ovary cells, human lymphocytes), but is time-consuming, requires skilled and experienced scorers, and cannot accurately measure aneuploidy (numerical changes in chromosome number). In the early years of conducting this assay, excessive cytotoxicity impacting data interpretation was a major

confounding factor in many laboratories. As a result, updated guidelines established for acceptable cytotoxicity levels (OECD, 2014) have improved the reliability of the test.

Micronucleus (MN) assay

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

Both assays must be conducted under strict conditions limiting cytotoxicity to acceptable levels (defined in the OECD TGs). These *in vitro* tests for chromosomal damage, when conducted with appropriate bioactivation, are, in general, more sensitive (i.e. detect more compounds as active) than the *in vivo* tests for chromosomal damage, leading to suggestions that they produce many "irrelevant positives". The increased sensitivity may involve factors such as enhanced exposure of cells in culture compared with target cells *in vivo* and higher achievable concentrations of the test article in cultures. Positive results in the *in vitro* assay are typically followed by an *in vivo* test for chromosomal damage (typically, an *in vivo* rodent MN assay) to evaluate potential *in vivo* genotoxicity (Kirkland et al., 2007).

4.5.2.5 In vivo chromosomal damage assays

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* chromosome aberration assay, is time-consuming, requires skilled and experienced scorers, and cannot accurately measure aneuploidy, a core genotoxicity endpoint. Due to these limitations, the *in vivo* MN test is more commonly used today, as it can capture numerical and structural changes, is less technically exacting, and lends itself to automation (flow cytometry), which speeds up data acquisition and increases the power of the assay (Dertinger et al., 2011).

Micronucleus (MN) assay

Similar to what has occurred with the *in vitro* chromosomal damage assays over time, the *in vivo* chromosomal aberration assay has gradually been replaced by the *in vivo* MN assay (OECD TG 474). The standard assay evaluates the formation of MN in newly formed bone marrow erythrocytes of mice and rats. A new flow cytometry version of the MN assay employs fluorescent dyes to identify cell surface markers (transferrin receptor) specific to immature and mature erythrocyte populations. This ability to distinguish erythrocytes by maturation stage allows the peripheral blood MN assay to be conducted in rats, where only the very youngest portion of the immature erythrocyte population can be evaluated accurately for MN presence due to efficiency of the rat spleen in sequestering and destroying damaged erythrocytes entering the circulation. Positive results in *in vivo* chromosomal damage assays are correlated 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.

4.5.2.6 In vitro DNA damage/repair assays

In vitro assays have historically assessed DNA damage and repair by measuring unscheduled DNA synthesis (UDS) in cultured mammalian

cells (TG 482); however, based on the observation that certain TGs, including TG 482, are rarely used in various legislative jurisdictions and have been superseded by more reliable tests, TG 482 has been deleted. Whilst information from such assays can still contribute to a weight of evidence assessment of mutagenicity, , testing of chemicals using these assays is not now recommended by OECD (2015). 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 4.5.6.

4.5.127 In vivo DNA damage/repair assays

Comet (single-cell gel electrophoresis) assay (OECD TG 489)

The comet assay 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. The assay detects overt strand breaks or lesions (e.g., double-strand breaks, as well as oxidized bases, alkylations, bulky adducts, cross-links that can be converted to single-strand breaks under alkaline (pH>13) conditions and 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.

The comet assay is increasingly employed as a second *in vivo* assay to accompany the *in vivo* MN assay, since 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., MN) 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 a screen 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 that may lead to mutation or chromosomal damage (i.e. permanent, viable, heritable change). It is generally recognized that cross-links cannot be reliably detected with the standard experimental conditions of the comet assay (OECD TG489) but can be assessed using modified test protocols.

DNA adduct assays

The detection and characterization of DNA adducts can provide mechanistic information on the MOA of genotoxic 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., 2010; Brown, 2012). A broadly applicable and non-specific, but highly sensitive, method is the ^{32}P-postlabelling assay (e.g., Phillips, 1997; Jones, 2012). A number of physical detection methods may be suitable for agents with the requisite properties (e.g. fluorescence or electrochemical detection, coupled with high-performance liquid chromatography [HPLC]). 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 radiolabeled compounds (usually with ^{14}C), accelerator mass spectrometry offers the highest sensitivity in detection, but does not provide structural information.

4.5.3 Interpretation of test results

Genotoxicity can be a hazard endpoint of concern *per se* or a potential key event in the MOA for an adverse outcome such as carcinogenicity or developmental toxicity. Assessment of genotoxicity both qualitatively and quantitatively can therefore be of great value in interpreting the

DRAFT

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. Qualitatively, it can add to WOE for genotoxicity as a key event in an adverse outcome, in different species, tissues, life stages, etc.

4.5.3.1 Identification of relevant studies

Since the assessment of genotoxicity is preferably based on all available data, an appropriate literature search should be performed. WHO (2017) guidance on systematic literature searches[2] can be consulted for general aspects such as selection of database, inclusion and exclusion criteria (e.g., language(s)), documentation of search strategy, and screening of the results.

Generally, information on the chemical(s) of interest is obtained using a database such as ChemIDplus[3] which enables combining the Chemical Abstracts Services (CAS)-number, chemical names, and literature search terms in databases such as PubMed. Structure searches should be performed with care and consider stereochemistry, tautomerism, salt form, and counter ions, if applicable.

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

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*

[2] http://www.who.int/foodsafety/chem/jecfa/Literature_Search.pdf?ua=1

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

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

"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, mutagenicities, 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, and/or aneuploidy. Lacking these data, the chemical is considered data-poor. For data-poor chemicals with known chemical structures, read-across, (Q)SAR, structural alert and/or TTC-based approaches can be used and are discussed in section 4.5.4.

It may be appropriate to further limit the search, such as by language(s) 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 increased transparency. For example, a publication lacking primary 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.:

Table 4.2. Some open-access sources of genetic toxicology data (non-exhaustive list)*

Database	Description
ATSDR	U.S Agency for Toxic Substances and Disease Registry (ATSDR) chemical database with genotoxicity information https://www.atsdr.cdc.gov

Database	Description
CCRIS	Chemical Carcinogenesis Research Information System (CCRIS) database with summary carcinogenicity and genotoxicity results of studies conducted 1985 – 2011 https://toxnet.nlm.nih.gov/newtoxnet/ccris.htm
CTD	Comparative Toxicogenomics Database (CTD) open access database that 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	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 for Ames positive compounds. http://data.jrc.ec.europa.eu/dataset/jrc-eurl-ecvam-genotoxicity-carcinogenicity-ames
EPA 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
GENE-TOX	Externally peer reviewed genetic toxicology data from literature published 1991 – 1998 https://www.nlm.nih.gov/databases/download/genetox.html
IPS INCHEM	International Program on Chemical Safety database of summary documents including genotoxicity. http://www.inchem.org
IRIS	Integrated Risk Information System database from the U.S. EPA with chemical risk assessments, including genotoxicity https://www.epa.gov/iris
ISSCAN	Carcinogen structures and empirical data from Istituto Superiore di Sanità. http://old.iss.it/meca/index.php?lang=1
ISSSTY ISSMIC	*In vitro Salmonella and in vivo* MN results from Istituto Superiore di Sanità. http://old.iss.it/meca/index.php?lang=1
Japanese NIHS Ames list	Japanese National Institute of Health Sciences: Ames mutagenicity data for approximately 12,000 new chemicals, list of strongly positive chemicals available via 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	MAK value documentations for chemical substances at the workplace including data on genotoxicity and carcinogenicity; https://onlinelibrary.wiley.com/doi/book/10.1002/3527600418

DRAFT

Database	Description
NTP – CEBS	Chemical Effects in Biological Systems (CEBS) database of U.S. National Toxicology Program study results including genotoxicity https://tools.niehs.nih.gov/cebs3/ui/
NTP – Tox21 Toolbox	Tox21 database including DrugMatrix toxicogenomics data and companion ToxFX database https://ntp.niehs.nih.gov/results/tox21/tbox/

*Modified from Amberg et al., 2016

4.5.2.2 Presentation and categorization of results

Criteria for the evaluation of genotoxicity test results, similar to those described in several OECD guidelines, should be used to judge a study result as positive, negative, or equivocal:

1. at least one of the test concentrations (or doses) exhibits a statistically significant increase compared with the concurrent negative control
2. the increase is dose-related in at least one experimental condition when evaluated with an appropriate trend test
3. any of the results are outside the distribution of the historical negative control data (e.g., Poisson-based 95% control limits).

The result should be considered clearly positive if all three criteria are fulfilled and clearly negative if none of the criteria are fulfilled, given no major methodological deficiencies exist. The result is equivocal if only one or two criteria are fulfilled. However, while these criteria could generally be applied for unpublished study reports, which may or may not conform to an OECD TG, historical control data are unfortunately rarely reported in published studies. In such cases, the reproducibility and strength of the result might also be considered. The reproducibility should be considered when separate[4] 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" versus "inconclusive" by EFSA (2011) may be informative to assist in an evaluation. The term

[4] The term "independent experiments" should only be used for experiments or studies that were performed in different laboratories by different scientists.

"equivocal" usually refers to a situation where not all the requirements for a clear positive result have been met. Whereas, 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 should produce a clear result.

Specific aspects that should be considered for the evaluation of positive and negative findings have been addressed by the European Chemicals Agency (ECHA, 2017):

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.
- For *in vitro* mammalian cell assays, factors such as the cell line, the maximum concentration tested, the measure of cytotoxicity, and/or the metabolic activation system can influence specificity
- Responses generated only at highly toxic doses or cytotoxic concentrations should be interpreted with caution (i.e. based on criteria defined in OECD TGs)
- The presence or absence of a dose (concentration)-response relationship
- Presence of known genotoxic impurities

Particular considerations when evaluating ***negative*** results include:

- Were the doses or concentrations tested sufficiently high such as to elicit signs of (cyto)toxicity?
- Was the test system adequately sensitive? For example, 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 DMSO, methanol and ethanol inhibit CYP2E1 and thus may interfere with bioactivation.
- Excessive cytotoxicity

Evaluation of *data quality* for hazard/risk assessment includes adequacy, relevance and reliability (Klimisch et al., 1997; OECD, 2005; ECHA, 2011). **Relevance** and **reliability** as they relate specifically to genotoxicity are described further below as their combination helps define the adequacy of the genotoxicity database to support a conclusion on genotoxic potential for hazard/risk assessment purposes. **Adequacy** is discussed in section 4.5.3.4; Weighting and integration of available information (section 4.5.3.3) are pivotal to determining adequacy. When carcinogenicity data are also available, the database may include specific mechanistic or MOA studies, particularly if the substance is carcinogenic (discussed later in 4.5.3.5 and 4.5.3.6).

Relevance:

Relevance of available genotoxicity data should be evaluated based on whether the data inform one of the three genetic endpoints (i.e. induction of gene mutations, structural and numerical chromosomal alterations) versus other genotoxic effects, with the former being most relevant and the latter considered supporting information. Some considerations that could impact relevance of the study results include (EFSA 2011):

- *"Purity of test substance*: Generally, tested 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 guidelines) may not ensure 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).
- *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 toxicokinetic data) on which such an assumption could be based (see ICH, 2011b and Kirkland et al., 2019).

- Problem formulation, i.e. whether this is 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 target tissue exposure, there will be no risk of mutagenicity *in vivo*. This is the case with phenol.
- *Reproducibility of results*: Conflicting results in tests 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.
- *Equivocal results* are generally less relevant than clearly positive results, however, they may suggest genotoxic potential which should be clarified by further testing as this is also recommended by OECD test guidelines. A modification of the experimental conditions may be taken into consideration.

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

If data to assess gene mutations, structural chromosomal aberrations, and aneuploidy are available, it is useful to tabulate the results grouped by endpoint as described in the JMPR Guidance Document for WHO Monographers and Reviewers (https://www.who.int/foodsafety/publications/JMPR-guidance-document/en/) with columns on 'reliability/comments', 'relevance of the test system' and 'relevance of the study result'. *In vivo* tables should include the 'test system' (e.g. bone marrow MN assay; ten 12-week old male B6C3F1 mice per dose), 'route' (e.g. oral gavage, feed, i.p.), '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)), 'reference' as well as the three additional columns mentioned above.

The result should be presented as judged by the genotoxicity expert/reviewer(s), preferably as positive, negative, equivocal or inconclusive. Discordance between judgements of the genotoxicity expert and those of the study authors should be described in the comments.

Generally, the relevance 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*.

Reliability:

Some tools for evaluating **reliability,** while not specific to genotoxicity, include the SYRCLE risk of bias tool for animal studies (Hooijmans et al., 2014), Toxicological data Reliability Assessment Tool (ToxRTool) (Schneider et al., 2009), and Science in Risk Assessment and Policy (SciRAP) (Mollander et al., 2015; Beronius and Ågerstrand, 2017). Klimisch et al. (1997) provide a classification approach such as (1) Reliable without restriction, (2) Reliable with restrictions, (3) Insufficient reliability, or (4) Reliability cannot be evaluated. These tools may be particularly helpful when assessing unpublished studies based on secondary source(s). However, their utility for primary study reports, including peer reviewed literature, should be considered on a case-by-case basis, based on the problem formulation and given their resource-intensive nature.

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

Any limitation that results or contributes to a judgement of limited or insufficient reliability should be described in the 'Reliability/Comments' column. Klimisch reliability scores of 2, 3 or 4 may result, for example, from limitations related to no original data, is a conference abstract without subsequent full publication, inappropriate study design (e.g. inappropriate dose levels, lack of appropriate controls, inappropriate solvent/carrier), insufficient protocol details, inappropriate data analysis, source and purity of chemical not reported, chemical mixture (unless target substance), potential for bias (e.g. samples not reported as analyzed blind); for human studies, uncharacterized or mixed exposure, inappropriate sampling time, etc.

Relevance of the test system

The relevance of the test system (high, limited or low) is based on the genetic endpoint with gene mutations, structural and numerical chromosomal aberrations and results obtained in an *in vivo* comet assay that detects DNA damage generally considered to be of high relevance, unless they occur under conditions described above which may result in a lower classification. DNA adducts may be considered of high (or lower) relevance, depending for example on the methodology used to assess their occurrence and on the type(s) of adducts induced (e.g., bulky adduct). Other endpoints of limited or low(er) relevance may provide useful supporting information. The available studies should be categorized according to the endpoint assessed. Results from oral *in vivo* genotoxicity studies are generally preferred for chemicals in food rather than data obtained through dermal or inhalation exposures. Recommended templates are provided for *in vitro* studies (Table 4-3) and *in vivo* studies (Table 4-4).

Only the relevant and reliable studies should be tabulated, rather than an exhaustive list. Studies considered to have both low relevance of the test system and study result should be omitted. For example, the relevance of the study result is low if either the reliability has a Klimisch score of greater than 2 or the relevance of the test system is low (or both).

DRAFT

Table 4.3. In vitro studies

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 studies

Test System	Route	Doses	Result	Reference	Klimisch reliability/ Comments	Relevance of Test System	Relevance of Study Result
					1	High	High

A general footnote can be included to indicate that studies with low relevance and/or reliability are omitted. After the data are tabulated, the most notable data gaps, whether *in vitro* or *in vivo*, that impact the evaluation should be discussed narratively.

4.5.3 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 relevance, reliability as described above, reproducibility and consistency, significance and mechanism of the genetic alteration, phylogenetic relationship to humans, study type (*in vivo* vs. *in vitro*), and physiologically relevance of the dose and route of administration with respect to human exposures (see below 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 judgment to reach conclusions about the significance of the genotoxicity results. The WOE approach should account for the key genetic endpoints (i.e. gene mutations, structural and numerical chromosomal aberrations) and the appropriateness of *in vivo* follow-up for positive *in vitro* results.

Studies with the following characteristics are generally given more weight in assessing human health risks:

- Highly relevant and reliable studies, as described in Section 4.5.3.2. 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 endpoints (i.e. gene mutations, structural and numerical chromosomal aberrations);
- *In vivo* studies in humans, other mammals, or other species known to respond similarly to humans. Human studies with well- characterized exposures and an absence of co-exposures or other potential confounders;
- Studies conducted using a physiologically-relevant exposure route (i.e. oral, dermal, or inhalation) and under other conditions (e.g. acceptable concentrations/doses, levels of toxicity, and diluents, absence of co-exposures) within generally accepted guidelines as opposed to non-physiological routes;
- Oral route studies are preferred when evaluating chemicals present in the diet;
- Studies using accepted and validated models, 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;
- 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);
- 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, from Eastmond (2017), DNA damage occurring

in the bladder of saccharin-treated rats secondary to urinary crystal formation (NTP, 2016), or 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 alpha-2u-globulin nephropathy (Swenberg, 1993) is weighted less in an evaluation.

In many cases, carcinogens exhibit genotoxicity 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. Contrasting results for the same endpoint in studies using comparable methodology should be evaluated on a case-by-case basis using the weighting considerations outlined above.

There are several key considerations when evaluating a chemical for mutagenicity. For core endpoints, 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 relevance of non-compliant or pre-guideline studies may need expert judgement, particularly when guideline studies exist.

Another scientific consideration is that the results 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 cautiously if not reproduced in other laboratories.

Another consideration is whether a consistent pattern exists. Are the observed results plausible given the known mechanisms of toxicity or action of the agent? While not required, it is anticipated that a substance that is clastogenic *in vitro* will also be clastogenic *in vivo*; and that an agent that is clastogenic in germ cells *in vivo* will also be clastogenic to somatic cells and vice versa (with appropriate toxicokinetic and/or sex

DRAFT

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 due 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.

The WOE evaluation should also note whether evidence exists to support alternative, non-mutagenic MOAs (discussed in further detail below), whether structural relationships to known genotoxic substances exist, to identify data gaps and uncertainties, and should ultimately enable a final conclusion on genotoxicity and more specifically, mutagenicity (described further in section 4.5.3.4).

4.5.3.4 Adequacy of the genotoxicity database

After critical review of relevant and reliable genotoxicity data, JMPR (2015) 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 tioxazafen, pyriofenone (JMPR, 2019) and abamectin (JMPR, 2016).

On the other hand, 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 (2018) was unable to complete the evaluation of the copolymer food additive AMC (anionic methacrylate copolymer). While 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, methacrylic acid and requested further studies to clarify *in vivo* carcinogenic potential.

Chemicals of interest, such as residues or contaminants, lacking data from the minimum range of tests (i.e. an indication of its ability to induce gene mutations, structural chromosomal aberrations, and aneuploidy)

require genotoxicity evaluation using (Q)SAR, read-across and/or TTC-based approaches (see Section 4.5.4).

There is considerable flexibility in the description when positive and/or equivocal test results exist (JMPR, 2015). For example, in an adequate range of *in vitro* and *in vivo* assays, the compound "gave a positive/equivocal response in the *in vitro* [names of endpoint/assay], but it was negative in the *in vivo* [names of endpoint(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 (2016) found no evidence of genotoxicity in numerous *in vivo* assays for acetochlor despite weak mutagenicity *in vitro* with less pure material 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*. As positive results *in vitro* do not confirm direct DNA-reactivity, *in vivo* follow-up is expected. As mentioned in 4.5.2, the comet assay is increasingly employed as a second *in vivo* assay to accompany the *in vivo* MN assay.

Exposure context, such as whether the observed genotoxicity 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, dermal route or by inhalation, when concluding on genotoxic potential.

4.5.5 *Mutagenic mode of action and carcinogenicity*

When carcinogenicity data in humans and/or laboratory animals are available, these data should be interpreted in conjunction with a mutagenicity WOE conclusion. 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 and there is DNA in every cell; therefore, it is reasoned, there can be no DNA damage threshold that has no consequence; 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 and Williams, 2019). For example, studies for chromosomal damage and mutations in mice repeatedly exposed to the mutagen, EMS, demonstrated a clear practical threshold or no-observed genotoxic effect level (Pozniak et al., 2009). Thus, even for DNA-reactive mutagens, non-linear threshold-type dose responses can be seen. For all genotoxicants, there may be a level of exposure below which genotoxicity cannot be distinguished from background levels that are tightly monitored by endogenous systems designed to control cellular perturbations, including DNA damage, caused by exogenous and endogenous stressors.

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 genotoxic 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 defense mechanisms (anti-oxidants or metal homeostasis), high cytotoxicity, metabolic overload and physiological perturbations (e.g. induction of erythropoiesis) (OECD, 2011). Nonetheless, some indirect interactions that may give rise to non-linear dose-responses can occur at very low exposures, such as for arsenite carcinogenicity, where DNA repair inhibition occurs at very low, environmentally relevant concentrations (Hartwig, 2013).

Determining that a carcinogen is genotoxic is not sufficient to conclude that it has a mutagenic MOA for carcinogenicity (Cimino, 2006). A WOE approach that applies various weights to different endpoints/assays is recommended when evaluating whether a human or rodent carcinogen is likely to act via a mutagenic MOA. The level of evidence is specific to the endpoint 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. Professional judgment is

necessary with respect to data quality described in section 4.5.3.3 (i.e. relevance, reliability, adequacy). For example, some factors that provide more weight include:

- The substance is mutagenic in the target organ(s) in which carcinogenesis was observed
- The substance is DNA-reactive, or there is significant conversion to a DNA-reactive intermediate that is confirmed to be associated with carcinogenesis
- There is evidence of substantial covalent binding to DNA, preferably *in vivo* in the target tissue.
- The substance is a multi-route, multi-site, and multi-species carcinogen in animal bioassays, particularly if tumors arise in tissues that do not have high spontaneous incidences nor are hormonally sensitive
- Evidence that the substance acts as an initiator in a suitable assay
- Highly similar structural analogues operate via a mutagenic MOA; WOE is increased if the substance contains structural alerts for DNA mutagenicity and reactivity

Alternative MOAs

In addition to classical tumor promotors which would stimulate cell growth of initiated cells, epigenetic alterations (i.e. DNA/histone methylation) or non- or indirectly genotoxic (i.e. non-DNA-reactive) events are important in both experimental animal and human cancer. Indirect genotoxic MOAs particularly relevant for cancer involve interactions with proteins and enzymes involved in maintaining genomic stability, such as inhibition of DNA repair processes, tumor suppressor functions, cell cycle regulation and apoptosis. Some of these mechanisms are active on the level of tumor initiation and may lead indirectly to an increase in mutation 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, respective interactions have been observed at particularly low concentrations and thus appear to be relevant also under low-exposure

DRAFT

conditions (e.g., Hartwig 2013). Epigenetic alterations refer to heritable changes in gene expression without alterations in DNA sequences. They include alterations in DNA methylation patterns, histone and chromatin modifications, histone positioning and noncoding RNAs. Disruption can lead to altered gene function, such as activation of proto-oncogenes or inactivation of tumor suppressor genes. Thus, they can contribute to cancer initiation and progression (for review see Kanwal, Gupta et al., 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., Chervona et al., 2012, Costa, 2019). From a risk-assessment point of view, these MOAs usually are thought to exhibit a threshold which, in principle, would at low doses, protect from the respective adverse outcome. However, frequently the NOAEL in humans is 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 arsenite (e.g., Langie, Koppen et al. 2015).

DNA-reactive, epigenetic, and/or non-DNA-reactive mechanisms can cooperate in chemical carcinogenesis. Indeed, epigenetic changes often occur as a result of initial mutagenic events (c.f. Nervi et al., 2008).

4.5.3.6 Integration of carcinogenicity and genotoxicity

Similar to standard phrases for genotoxic potential mentioned in section 4.5.3, standard phrases with defined scenarios for chemicals with mutagenicity and carcinogenicity evaluations may include (adapted from JMPR, 2015):

[compound not carcinogenic or genotoxic]
In view of the lack of genotoxicity 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, in the evaluation of chlormequat by JMPR in 2017, it was concluded 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 genotoxic in vivo with positive in vitro genotoxicity]

In view of the lack of genotoxicity *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, in the evaluation of imazethapyr by JMPR in 2016, it was concluded that "Overall, the Meeting considered that the available studies provided no evidence of genotoxic effects *in vivo* and concluded that imazethapyr was unlikely to be genotoxic to humans from exposure through the diet." and that imazethapyr was "Unlikely to pose a carcinogenic risk to humans via exposure from the diet."

or

[carcinogenic compound not mutagenic]

In view of the lack of mutagenicity, the absence of carcinogenicity in [species] and the fact that only [tumors] 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. For example, in the evaluation of ethiprole by JMPR in 2018, it was concluded 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." *[There is considerable flexibility in wording here.]*

or

[carcinogenic compound with positive in vitro mutagenicity]

As [compound] was not mutagenic *in vivo* and there is a clear NOAEL ~~threshold~~ for [tumor type] in [sex] [species], it is concluded that [compound] is unlikely to pose a risk for mutagenic carcinogenicity to humans from the diet. For example, in the evaluation of fenpicoxamid by JMPR in 2018, it was concluded 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." *[There is considerable flexibility in wording here.]*

or

[carcinogenic compound 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 ~~threshold~~ for [tumor type] in [sex] [species], it

is concluded that [compound] should be considered a carcinogen acting by a mutagenic MOA.

or

[compound lacks carcinogenicity data]

Compounds lacking carcinogenicity data, or that have carcinogenicity data with major limitations, with or without adequate genotoxicity data, should be specified as lacking carcinogenicity data to conclude on carcinogenic potential. The major limitations of the existing data should be specified. Derivation of a health-based guidance value may not be appropriate if adequate genotoxicity data are available to support a WOE conclusion that the substance is mutagenic *in vivo*. For example, in the evaluation of natamycin by JMPR in 2017, it was concluded 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." The Meeting did not establish an ADI or an ARfD due to the inadequate database available to the Meeting. Alternatively, if adequate data on genotoxicity are available, it may be possible to use a weight-of-evidence approach to reach a conclusion on risk of carcinogenicity from exposure via the diet, even in the absence of data from carcinogenicity bioassays.

4.5.4 Special Considerations

4.5.4.1 *In silico* approaches for data-poor substances

In the regulatory arena, QSAR methods are used to predict the bacterial mutagenicity. These have been used for drug impurities lacking empirical data according to the ICH M7 (2014, 2017) guidelines (see Amberg et al., 2016, Sutter et al., 2013, Wichard, 2017). QSAR and read-across approaches have been used to assess genotoxicity of pesticide residues, including active substances and their metabolites, for dietary risk assessment (see EFSA 2016, Worth et al., 2010). QSAR models are also applied under the auspices of the EU REACH regulation, most commonly though not exclusively to support WOE approaches for mutagenicity prediction (e.g. Annex VII).

Available *in silico* tools (QSARs and structural alerts) for genotoxicity

In silico approaches pertaining to genetic toxicology typically comprise (Q)SARs, SARs (also referred to as structural alerts) and 'expert systems' comprised of QSARs, SARs or both that are categorized as statistical-(QSAR) or knowledge-based (SAR) or hybrids (Patlewicz et al., 2014).

Relative to other hazard endpoints, structural alerts for genotoxicity are the most established and many software tools exist to identify them. The breadth and scope of alert schemes may differ between different tools with the quantity of alerts within a given tool not necessarily 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/or QSARs are also available for gene mutations in mammalian cells, chromosomal aberrations, MN formation, and DNA binding, all of which contribute to genotoxicity assessment, for example to determine the TTC tier (see 4.5.4.2). *In silico* models and tools, and the data availability for model development for different genotoxicity endpoints have been recently reviewed (Benigni et al., 2019; Hasselgren et al., 2019; Tcheremenskaia et al., 2019). Table 4.6 provides examples of genotoxicity assessment approaches within commercial, open-source, or freely available software.

Table 4.6. Examples of commercial, freely-available, or open-source in silico tools

Type of model	Effects	Availability	Link
Expert system – Knowledge based	Alerts for genotoxicity, 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
Expert system – Hybrid – mix of QSARs and SARs	Ames mutagenicity	TIMES – commercial	http://oasis-lmc.org/products/models/human-health-endpoints/?page=2&

Type of model	Effects	Availability	Link
underpinned by a metabolism simulator	*In vitro* chromosomal aberration *In vivo* micronucleus test *In vivo* liver genotoxicity in vivo Liver TGR mutagenicity in vivo Liver Clastogenicity Comet genotoxicity		
Expert systems – statistical	Various genotoxicity endpoints	Leadscope Model Applier – commercial	http://www.leadscope.com/product_info.php?products_id=67
	Various genotoxicity endpoints	CASE Ultra – commercial	http://www.multicase.com/case-ultra-models
	Various genotoxicity endpoints ("Impurity Profiling Module")	ACD/Percepta – commercial	https://www.acdlabs.com/products/percepta/index.php
	Various genotoxicity endpoints	ChemTunes ToxGPS – commercial	https://www.mn-am.com/products/chemtunestoxgps
	Ames mutagenicity	TOPKAT – commercial	https://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/qsar-admet-and-predictive-toxicology.html
	Ames mutagenicity	LAZAR – freely available	https://lazar.in-silico.de/predict
	Ames mutagenicity	USEPA T.E.S.T. – freely available	https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NRMRL&dirEntryId=232466
	Ames mutagenicity	Sarah Nexus - commercial	https://www.lhasalimited.org/products/sarah-nexus.htm

Type of model	Effects	Availability	Link
	Ames mutagenicity	VEGA platform – freely available	https://www.vegahub.eu/
	Chromosomal aberrations	ADMET Predictor	https://www.simulations-plus.com/software/admetpredictor/toxicity/
Read-across tools – also incorporate weight of evidence QSAR results	Ames mutagenicity	ToxRead-open source	http://www.toxread.eu/
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/
Structural alerts	Carcinogenicity rules-based by ISS (incorporates Ashby Tennant rules), ISS *in vitro* Ames test alerts and *in vivo* mutagenicity (micronucleus); 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 micronucleus by OASIS, Benigni Bossa (ISS) alerts for *in vitro* mutagenicity	OECD QSAR Toolbox – freely available	https://qsartoolbox.org/

DRAFT

Type of model	Effects	Availability	Link
	Ames and *in vivo* mutagenicity (micronucleus)		

Confidence in *in silico* 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 and/or risk assessment
- Classification and labeling (under GHS)
- 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 assessed is within the applicability domain of the model. Generally, the predictivity of (Q)SAR models is closely related to the data available for model development and their quality. A recent project aimed to improve the quality of Ames data as the basis of related (Q)SAR models by extending the datasets with new data and re-evaluating historic Ames test results (Honma et al., 2019).

The performance of different *in silico* approaches for genotoxicity prediction has been reviewed elsewhere (see Hanser et al 2016, Netzeva et al, 2005), including analyses specifically for food ingredients and food .contact materials (for example, Benigni et al., 2019, Worth et al., 2010, Bakhtyari et al, 2013, Van Bossuyt et al., 2018, Vuorinen et al., 2017, Cassano et al., 2014, Greene et al., 2015). General aspects of confidence in and applicability of (Q)SAR models have also been reviewed recently, providing a list of guiding assessment criteria (Cronin et al 2019).

Quantitative consensus models can be used to deal with multiple QSAR predictions by leveraging the strengths, compensating for weaknesses of any individual model, and quantifying uncertainties, such

as for example for *in vitro* estrogenic activity (Mansouri et al., 2016) and acute oral toxicity (Kleinstreuer et al., 2018). A strategy for integrating different QSAR models for screening and predicting Ames mutagenicity in large datasets of plant extracts has recently been proposed (Raitano et al., 2019).

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 judgment to conclude on divergent results (Wichard, 2017; Powley et al., 2015; Greene et al., 2015). 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 Powley et al., 2015, Barber et al., 2015, Amberg et al., 2016, Myatt et al., 2018, Dobo et al., 2012. The expert review in a weight of evidence can include analogue information (read-across, see 4.5.4.3) (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 Hasselgren et al., 2019; Myatt et al., 2018) which is based on a combination of different experimental and *in silico* evidence lines to arrive at an overall conclusion about the genotoxic hazard of a substance. This approach includes Klimisch scores extended to more general Reliability scores also assessing *in silico* results (Table 4.7).

Table 4.7. Reliability of (geno)toxicity assessments based on in silico models and experimental data (modified from Myatt et al., 2018)

Reliability Score	Klimisch Score	Description	Summary
1	1	Data reliable without restriction	Well documented study from published literature Performed according to valid/accepted test guidelines (e.g., OECD) and preferably according to good laboratory practices (GLP)
2	2	Data reliable with restriction	Well documented study/data partially compliant with test guideline and may not have been GLP-compliant

3	-	Expert Review	Read-across Expert review of *in silico* result(s) and/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 endpoint Not sufficiently documented for an expert review
5	4	Data not assignable	Lack of experimental details Referenced from short abstract or secondary literature

Use of in silico methods for genotoxicity assessment

In the context of the present guidance, *in silico* approaches for genotoxicity assessment can be used (see also Figure 4.1, boxes 17, 22):

a) as the basis for application of the TTC approach depending on the presence or absence of genotoxicity structural alerts (or WOE that the substance might be genotoxic) to determine the TTC tier applied (see 4.5.4.2)
b) when insufficient empirical genotoxicity data exist on the compound to reach a conclusion, additional information should be sought from related analogues (i.e. read-across, see 4.5.4.3) and *in silico* approaches (e.g. (Q)SARs) and considered in an overall WOE evaluation of genotoxic potential (see also 4.5.4.3).

When using *in silico* models for genotoxicity assessment, it is recommended to apply two complementary models (i.e. a statistically-based model and an expert-rule-based system), as recommended in ICH M7(R1) (2017) and EFSA (2016b). The two models should not be based on the same training set data, descriptors or learning methods (Barber et al., 2017), in order to allow a WOE of two independent approaches (see also Greene et al., 2017). Practical application of QSAR models to predict genotoxicity is discussed in Amberg et al., 2016, Sutter et al., 2013,

Barber et al., 2015, Wichard, 2017, Mombelli et al., 2016. 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 chemicals.

4.5.4.2 Threshold of Toxicological Concern

While an understanding of the potential for a chemical in the diet to pose a genotoxic hazard is an important element of the overall safety assessment of food, it is also recognized that food can contain many contaminants/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/packaging materials; they can also be formed during food processing and/or cooking. Analytical chemists are now able to routinely detect chemicals at sub-ppb levels and, as analytical tools continue to improve, the detection limits continue to be lowered. At some point, one could consider exposure to a constituent to be sufficiently low that it does not pose a safety concern, and testing is not needed. This is the principle behind the concept of the Threshold of Toxicological Concern (TTC).

TTC is a screening tool that can be used to decide if experimental genotoxicity testing is required for compounds present in the diet at very low levels and that are not added intentionally to food.[5] 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 U.S. FDA's Threshold of Regulation (FDA, 1995) which was developed as a tool to facilitate the safety evaluation of food packaging materials, components of which (might) have the potential to migrate into food at very low levels.

TTC is used widely to assess low level exposures to substances with insufficient toxicity data; it was reviewed most recently by EFSA/WHO

[5] TTC is described in more detail in Chapter 9.1.1 of EHC 240.

DRAFT

(2016). It has been expanded from a single value (the FDA Threshold of Regulation) to encompass a range of exposure limits based on potency bins for chemicals. Substances posing a real or potential genotoxicity hazard are assigned to the bin with the most stringent exposure limit of 0.0025 μg/kg bw/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 TTC. There is work ongoing to further substantiate the TTC exposure limit for compounds considered to pose a possible genotoxicity hazard (Boobis et al., 2017). The recommended review will update 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 (genotoxic) carcinogens. It is also recognized that there are opportunities to refine the 0.0025 μg/kg bw/day exposure limit for the TTC genotoxicity 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 and Renwick, 2013). This assumption has been accepted in guidance for genotoxic 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 US FDA (Cheeseman et al., 1999) have shown that, on average, Ames-positive carcinogens are more potent than Ames-negative carcinogens

Chemicals are assigned to the "genetox tier" based on existing data (e.g. from genotoxicity 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 Toolbox, Derek Nexus, see 4.5.4.1). While this approach is generally considered to be robust, it is also recognized that different programs can result in binning chemicals differently, such that EFSA/WHO (2016) concludes 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. A WOE approach should be taken when binning chemicals into the genetox tier for 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 judgment. Although there remains more work to do on TTC, this is true for all safety assessment approaches. TTC remains an important tool for evaluating low-level exposures in food and can be used as an initial screen to determine whether genotoxicity testing/evaluation is needed.

4.5.13 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 substance ("read-across"). The WOE for evaluating genotoxic potential may come from read-across, structural alerts and/or (Q)SAR models, using expert judgment on all available information, including empirical data, if limited data exist.

Groups of substances with similar human health and/or environmental toxicological properties, typically based on an aspect of chemical similarity, are known as chemical categories. A category of two substances (an untested target substance of interest and a source analogue with data to read-across from) is referred to as an analogue approach. Hanway and Evans (2000) were one of the first to report read-across as part of the regulatory process for new substances in the UK. 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), which was developed in collaboration with 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 OECD 2014; ECHA, 2008; ECETOC, 2012).

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. An example of a top-down approach is the grouping of pesticides based either on phenomenological effects by the European Food Safety Authority (EFSA, 2013) or on common modes of action 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 (PBT); and/or carcinogenicity, mutagenicity, and reproductive toxicity (CMR). In contrast, the bottom up approach tends to encompass scenarios where 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, 2017a). In the United States, application of read-across varies widely between and within Agencies and decision contexts (Patlewicz et al., 2019). For example, applications within 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 (PPRTV) derivation in quantitative risk assessments for chemicals of interest to the USEPA Superfund program (Wang et al., 2012).

A critical aspect in a read-across determination is the identification and evaluation of analogues (i.e. the definition of similarity), which are both related to the chemistry and biological activity. In the genotoxicity field, these aspects are facilitated by the mature understanding of the modes of action and the associated test systems that characterize them. As such, the existence of genotoxicity structural alerts for mutagenicity,

clastogenicity, and DNA reactivity (see section 4.5.4.1) inform initial chemical categories.

There is a wide range of publicly accessible read-across tools (see Table 4.6 for examples and Patlewicz et al., 2017 for a detailed review) and databases with genotoxicity/mutagenicity data (see for example Amberg, 2016, Worth et al., 2010, Benigni et al., 2013, Corvi and 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 endpoint of concern, decision context, and similarity metric chosen. Similarity should be based not only on structural and physicochemical properties, which tend to have been over-emphasized (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 et al., 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 judgment 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 et al., 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 and Stuard, 2014; Schultz et al., 2015, Patlewicz et al., 2015; Schultz et al., 2019). The ECHA Read-Across Assessment Framework (ECHA 2017a), 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 (NAM) such as high

throughput (HT) or high content screening data and linkages to Adverse Outcome Pathways (AOPs) may help reduce uncertainty in read-across evaluations (see OECD, 2016, 2017b, 2018, 2019; Nelms et al., 2018, Zhu et al., 2016). More general approaches to quantifying performance and uncertainty transition towards a QSAR-like approach (Shah et al., 2016, Helman et al., 2018, Patlewicz et al., 2018, Zhu et al., 2016).

Read-across and (Q)SAR approaches are underpinned by the same principles and continuum of relating property/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, while read-across approaches tend to be more case-by-case based on expert review and judgement.

Read-across and (Q)SAR can be used in different contexts depending on the endpoint and its complexity. For genotoxicity, either approach may be feasible to address a specific data gap in lieu of testing. There are also instances where a WOE approach integrating relevant empirical and *in silico* information is appropriate, such as to propose waiving a specific higher order study or to bolster a less ideal study, both common scenarios in the REACH arena.

4.5 Considerations for Specific Compounds

4.5.1 Mixtures

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

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 intake. However, in some cases the safety of natural extracts added to

food should be evaluated based on experimental and/or *in silico* data. Genotoxicity testing in particular is complicated by the dilution of individual components, which may hinder their identification through conventional testing.

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

Genotoxicity testing of mixtures may apply the tiered approach recommended by the European Food Safety Authority (EFSA 2018a). First, 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 chemical characterisation of botanicals (e.g. EFSA 2009), novel foods (e.g. EFSA 2016), herbal medicinal products (e.g. EMA 2011; U.S. FDA 2016) and combined exposure to multiple chemicals (e.g. OECD/WHO, 2011, 2018; EFSA 2019). Analytical methods to identify and control mutagenic impurities and degradation products of pharmaceuticals (e.g. Görög 2018, Teasdale and Elder 2018), while not directly applicable to food, could also be consulted.

For well-characterized mixtures (i.e. 'simple' mixtures in which all components above a certain level are identified and quantified), genotoxic 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 genotoxicity data and if limited, supplemental (Q)SAR models.

If the mixture contains a significant fraction of unidentified substances (i.e. complex mixtures) and/or substances lacking empirical data, the chemically-identified substances are first assessed individually for potential genotoxicity. If none of the identified substances are

genotoxic, the genotoxic potential of the unidentified fraction should be evaluated. If possible, the unidentified fraction should be isolated for testing. Further fractionation of the unidentified material could be considered case-by-case to remove inert, toxicologically irrelevant components (e.g. high molecular weight polymers) in order to minimise the dilution of the components of interest, or to remove highly toxic components (e.g. surface-active substances) which may prevent testing adequately high doses of the mixture due 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 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 OECD for mixtures is 5 mg/ml compared to 2 mg/ml for single substances (For example, see OECD TG 473).

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

Regulatory guidelines for the assessment of the potential genotoxicity of botanical/herbal medicinal products (EMA 2006; U.S. FDA 2016) may also be useful when evaluating complex mixtures used in food.

4.5.2 Flavouring Agents

The Codex 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 (CAC/GL 66-2008). Flavouring agents represent a variety of liquid extracts, essences and flavours that are added to natural food products to impart taste and aroma or enhance them when 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).

Depending on the origin and means of production, flavouring substances identified as a single constituent include those obtained by chemical synthesis or isolated through chemical processes, and natural substances. Alternatively, flavouring preparations 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 molecules. Constituents that occur naturally in a flavouring preparation, due to their presence in the source materials (e.g. intrinsic fruit water) as well as foods/food ingredients used during the manufacturing process (e.g. ethanol, edible oil, acetic acid) can be considered to be part of the flavouring preparation.

A category of complex flavouring agents are smoke flavours 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). Due to the intrinsic chemical complexity of flavouring agents (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, genotoxicity testing, if needed, should be tailored accordingly. Benzo(a)pyrene, a 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, JEFCA (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 their safety data were limited, novel uses of smoke flavourings should be approached with caution (JECFA, 1987).

Currently, the JECFA procedure for the safety evaluation of flavouring agents considers whether the WOE from empirical genotoxicity data and/or structural alerts suggests that the flavouring agent is potentially a mutagenic carcinogen. If either answer is affirmative, then the 'JECFA Procedure' (described in Chapter 9.1.2.1 of EHC 240) cannot be applied.

Whether flavouring agents are composed of **single or multiple constituents,** EFSA (2011) uses a conventional genotoxicity testing approach to evaluate mutagenicity for each component with genotoxicity structural alerts. Consistent with recommendations of Kirkland et al. (2011), EFSA (2011) recommends testing using an *in vitro* bacterial reverse mutation assay and an *in vitro* MN test. If an individual component is positive *in vitro*, additional *in vivo* data are needed to complete the genotoxicity assessment. A positive outcome for a single component can result in the flavouring agent being considered genotoxic. Conversely, when components do not raise genotoxicity concerns, the flavouring agent is considered non-genotoxic. EFSA (2018b) recently reaffirmed this position.

Flavouring agents that are complex mixtures should be tested according to the procedure recommended for extracts from natural sources (see 4.5.5.1).

4.5.5.3 Minor Constituents

Minor components include metabolites of pesticide or veterinary drug active ingredients found as residues in food of plant and animal

origin, impurities of the active ingredient, degradates of pesticides due to non-enzymatic processes, or those found in food following application of pesticide active ingredients.

A step-wise approach to evaluate genotoxicity of these 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 genotoxicity.

The evaluation of genotoxic potential is part of the general toxicological evaluation of such impurities/degradation products as illustrated in Figure 4.2. Sections of the assessment scheme pertaining to genotoxicity are described below, assuming that, for the compound under evaluation, there are no empirical genotoxicity data available:

- ***Step 1***: Identify any available toxicological information for the compound of interest
- ***Step 2***: Evaluate whether the compound of interest is formed in mice, rats, and/or dogs, and hence whether the compound has been tested for (geno)toxicity in tests with the parent compound. As a general rule, the compound is considered tested if urinary levels represent at least 10% of the absorbed dose. Conjugates and downstream metabolites that derive only from the metabolite of interest are also included in the total.
- ***Step 3***: Evaluate the possible role of the metabolite in the (geno)toxicity, 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 degradates, the read-across approach is used to evaluate the genotoxicity information of compounds considered to have sufficient structural similarities to the parent/known compound(s) to permit read-across (see section 4.5.4.3 for details). If read-across is deemed not possible, such as due to the lack of sufficiently similar tested analogs, proceed to step 5.
- ***Step 5***: This step starts with consideration of whether specific

exposure estimates are available[6]. If specific exposure estimates are available, proceed to step 6. If not, list all available relevant information, such as:

- read-across from related substance(s)
- genotoxicity structural alert(s)
- estimate of upper bound of exposure, if available
- other relevant information

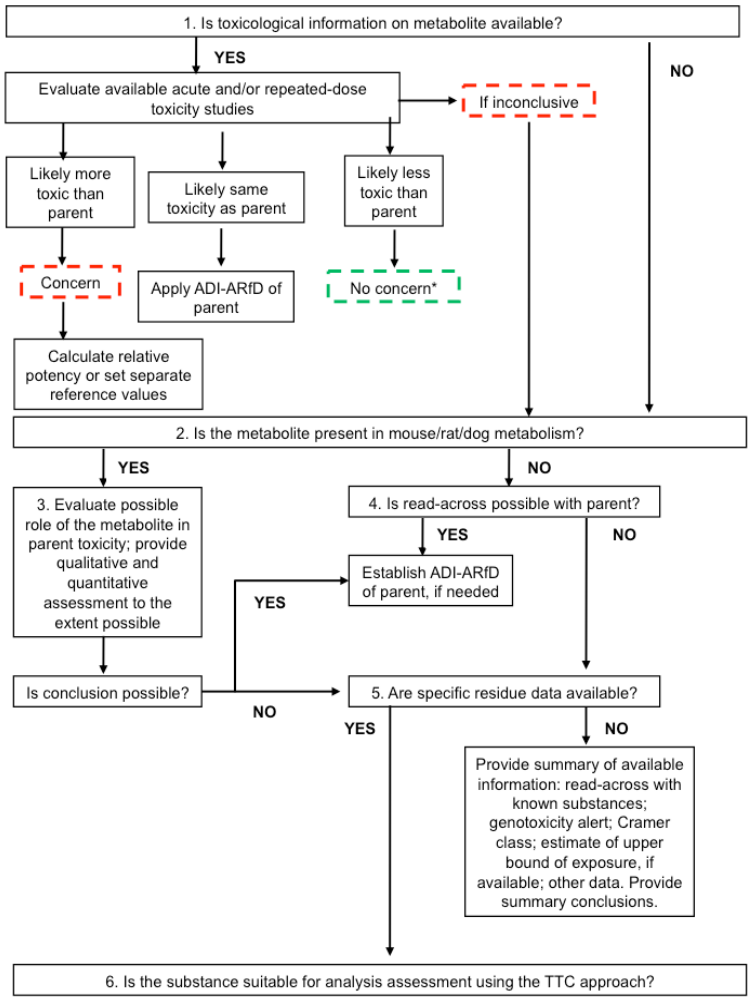
then determine whether the metabolite is of potential (geno)toxic concern, if possible, and/or provide advice for further assessment.

- ***Step 6***: Substances currently not suitable for assessment by the TTC approach (see section 4.5.4.2) 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 genotoxic compounds (0.0025 µg/kg bw) the evaluation can be terminated with low genotoxic concern. Otherwise, proceed to step 8. See section 4.5.4.2 for more details on application of the TTC.
- ***Step 8***: Structural alert models (see section 4.5.4.1) are available that are suitable for this step. If there are no genotoxicity structural alerts, or if the only structural alert is also present in the parent compound and this was negative in an adequate range of genotoxicity tests, the evaluation can be terminated with low genotoxicity concern. Otherwise, proceed to steps 9/10.
- ***Steps 9/10***: Adequate *in vitro* or *in vivo* genotoxicity data are required to assure that DNA-reactive carcinogenicity is unlikely despite the presence of structural alert(s), based on a WOE evaluation (see 4.5.3.3 and 4.5.3.4).

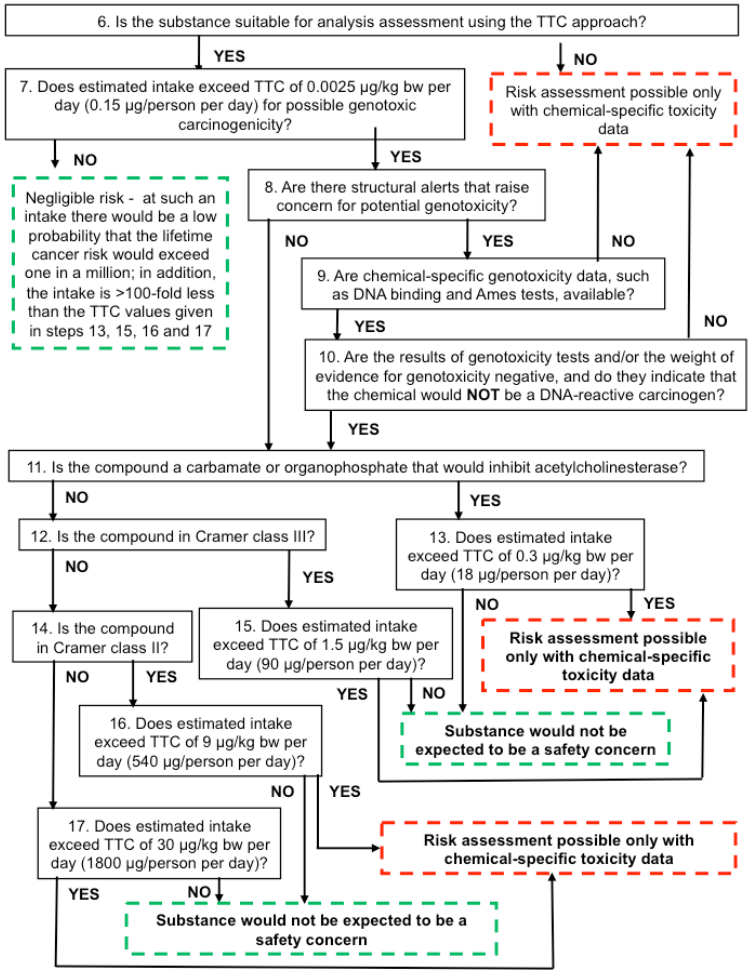
Note that, based on structural considerations, if there are several compounds for which read-across would be possible, testing might be limited to one (or more) representative metabolite.

[6] Dietary exposure assessment is detailed in EHC 240 Chapter 6.

Figure 4.2 Plant and animal metabolite scheme (from JMPR, 2015)



* Note: For compounds already included in residue definition.



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

4.5.14 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, that also includes the food enzyme of interest, has traditionally been used in genotoxicity 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 GI tract and the likelihood of them showing immunological cross-reactivity with known allergenic proteins.

The JECFA (2006) General Specifications and Considerations for Enzyme Preparations Used in Food Processing are based on Pariza and Foster (1983) and the Scientific Committee on Food (SCF, 1991) in Europe. 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 and Foster, 1983; Pariza and Johnson 2001) (see also chapter 9.1.4.2 in EHC 240).

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 genotoxic secondary metabolites. Microbial secondary metabolites are low molecular weight entities that are not essential for the growth of producing cultures. JECFA (2006), based on SCF (1991), recommended the following tests be performed:

- A test for gene mutation in bacteria;
- 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, diluents). The SCF emphasized that these tests were intended to reveal genotoxic effects of unknown compounds synthesized during the fermentation process.

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
- the actual strain used has a well-documented origin,

then it was possible to use the enzyme preparation from such an organism without any genotoxicity 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 (JECFA, 2002) based on a JECFA (1972) conclusion that enzymes from microorganisms traditionally accepted as natural food constituents or normally used in food preparation should themselves be regarded as foods. Up until 2018, JECFA has evaluated over 80 food enzyme preparations from microorganisms such as *Trichoderma reesei*, *Bacillus subtilis*, *B. amyloliquefaciens*, *B. licheniformis*, *Aspergillus niger* and *A. oryzae* but never recorded a positive result in any genotoxicity assay. These data suggest that there are several strains of

microorganisms that could constitute safe strain lineages for food enzyme production and therefore not require genotoxicity testing.

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

4.5.6 Recent Developments and Future Directions

The need to evaluate the genotoxic hazards posed by thousands of chemicals in commerce remains an urgent priority. This need also pertains to the quantitative assessment of the risk associated with realistic environmental exposures. The former necessitates the development and validation of novel, HT tools for 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 et al., 2016, Riebeling et al., 2018). The latter requires the establishment of a computational framework for dose-response analysis that includes Point of Departure (PoD) determinations for the interpretation of genetic toxicity test data in the context of risk assessment (White and Johnson, 2016).

Recently developed HT tools exploit advances in informatics and instrumentation technologies to rapidly assess traditional genotoxicity endpoints (e.g., mutations and chromosome damage), and/or molecular endpoints indicative of DNA damage and/or a DNA damage response. Additionally, (Q)SAR-based models developed by commercial (e.g., Leadscope, MultiCase, Lhasa Limited) or public-sector (e.g., OECD) organizations are increasingly being used for predicting bacterial mutagenicity and chromosomal damage (see Table 4.6 and section 4.5.4.1). HT and *in silico* methods can rapidly screen and prioritize potential genotoxicants, recognizing that their direct utility to establish HBGV (Health-based Guidance Values) (e.g. ADI, ARfD, MOS) is presently limited.

DRAFT

4.5.6.1 Novel in vivo genotoxicity approaches

HT technologies such as flow cytometry and automated microscopy permit rapid detection and quantification of induced gene mutations and chromosomal aberrations *in vivo* (Section 4.5.2). Since many of these assays evaluate genotoxicity biomarkers in peripheral blood, they can be readily integrated into ongoing subchronic bioassays, thus reducing the need for independent genotoxicity tests (Dertinger et al., 2002, Witt et al., 2007). Additionally, some methods are amenable to evaluating genotoxicity biomarkers in humans (Collins et al., 2013; Dertinger et al., 2015, Fenech et al., 2013; Olsen et al., 2017; Witt et al., 2007).

In addition to the HT approaches highlighted earlier (Section 4.5.2), novel *in vivo* approaches (Table 4.8) 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 RaDR Mouse; Hendricks et al., 2003, Sukup-Jackson et al., 2014). No international guidelines yet exist for these approaches, but data could be used in support of guideline test data.

4.5.6.2 Novel in vitro genotoxicity approaches

The need to rapidly screen the enormous number of as yet uncharacterized compounds while simultaneously reducing animal use for (geno)toxicity assessment has led to the development of novel, HT *in vitro* tools. Despite noteworthy advantages related to cost, throughput, and information content of these assays, incorporation of realistic and effective xenobiotic metabolism is a concern. Nevertheless, numerous novel *in vitro* assays are available to rapidly assess induction of DNA damage and repair, while others assess more traditional endpoints, such as the induction of gene mutations, chromosomal damage, or DNA strand breaks (Table 4.9). Since genotoxicity screening for regulatory purposes generally requires assessment of gene mutation and chromosome damage, assays that streamline detection of these endpoints 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*, *λcII*, *gpt*). These assays do not require clonal selection, and can more efficiently measure mutagenicity compared to, for example, traditional *tk* and *hprt* locus mutation assays.

Some of the HT *in vitro* assays summarized in Table 4.8 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., beta-lactamase). Reporter gene activation is visualized via, e.g., automated micro-confocal imaging, fluorescent or luminescent readouts, or flow cytometry. Examples include the ToxTracker® (Hendricks 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 U.S. Tox21 Program (https://ntp.niehs.nih.gov/results/tox21/tbox/) or the U.S. Environmental Protection Agency's ToxCast Program (https://actor.epa.gov/dashboard2/). Importantly, in addition to genotoxic hazard, the simultaneous or sequential examination of multiple endpoints representing several distinct pathways permits delineation of the genotoxic MOA. Related assays, such as the MultiFlow® DNA Damage Assay, assess the presence and/or 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-labeled 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, qPCR and RNAseq approaches have been used as HT approaches for measuring DNA damage signaling. 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 signaling) to identify DNA-damage-inducing (DDI) substances (Li et al., 2015, 2017; Williams et al., 2015; Yauk et al., 2015, 2016; Corton et al., 2018, 2019). Similarly, a quantitative HT 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/or repair as indirect measures of genetic toxicity.

1 **Table 4.8. 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 endpoint, 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 reporter direct repeat 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, Balboa et al., 2014, Hemeryck et al., 2016, Lai et al., 2016; Chang et al., 2018, Takeshita et al., 2019
In vitro Assays that assess the frequency of mutations or DNA damage				
In vitro 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 et al., 2015, 2016; Bemis and Heflich, 2019
In vitro TGR reporter mutagenicity assays	Positive selection assay to detect mutations at a variety of transgenic loci (e.g., *lacI*, *lacZ λcII*, *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	Laborious compared with HT reporter-based assays, transgenes not endogenous loci, no consensus regarding assay protocol, not HT	White et al., 2019

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
		clonal selection, some versions partially validated.		
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 and Pfeifer 2010; Kucab et al, 2010
Cisbio® γH2AX assay	Quantification of H2AX phosphorylation	Positive responses highly predictive of genotoxicity (clastogenicity); homogeneous format with no wash steps required; HTS compatible; suitable for use with adherent or suspension cells	Requires an HTRF® compatible reader and a -60 C freezer	Hsieh et al., 2019
In vitro Reporter assays (indirect measures of genotoxicity)				
ToxTracker® assay	Expression of specific reporter genes up-regulated by DNA damage	Simultaneously monitors genes involved in DNA damage response, oxidative stress, and protein damage response, flow cytometry scoring	Restricted to specifically-constructed cell lines	Ates et al. 2016; Hendriks et al. 2012, 2016
MultiFlow® DNA Damage assay	In vitro high content assays for multiple endpoints	Determines MOA for micronucleus induction, flow cytometry scoring	Method developed for suspension cell lines only	Bryce et al. 2013; Bemis et al. 2016; 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 kinase B, 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 signaling 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

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
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
GreenScreen®, BlueScreen	GADD45a based reporter system, GFP (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	Simpson et al., 2013; Hastwell et al., 2006
High Throughput real-time qPCR	Gene expression assessment of 95 genes involved in genomic stability	Can be used for cell lines, primary cells, 3D 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; Willliams et al., 2015; Yauk et al., 2015, 2016; Corton et al., 2018

To date, none of the HT tools listed in Table 4.8 have OECD Test Guidelines, nor have they been incorporated into widely-accepted genotoxicity assessment platforms, such as those recommended by the ICH (2011), USFDA (2007), or ECHA (2017c). A future role for these tools in regulatory decision-making would be consistent with global trends to modernize the current genotoxicity assessment frameworks, to reduce and replace experimental animals, and to generate genotoxicity MOA information. For example, Dearfield et al. (2017) outlined a paradigm shift whereby a variety of mechanistic endpoints indicative of genomic damage are incorporated into a "next generation testing strategy". Indeed, HT tools are already supporting regulatory evaluations based on traditional *in vitro* assays. For example, the European Commission Scientific Committee on Consumer Safety considers additional *in vitro* tests that include gene expression and recombinant cell

reporter assays (SCCS, 2018). Similarly, Corton et al. (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 genotoxicants (i.e., Disperse Orange and 1,2,4-benzenetriol), resulting in compound classification consistent with more traditional endpoints (e.g., *in vitro* micronucleus formation). Private-sector organizations are now routinely using HT *in vitro* assays to evaluate genotoxicity of products in development (e.g., therapeutic candidates, industrial chemicals) (Dertinger et al., 2018, Sullivan et al., 2016; van Goethem, 2017; Shell, 2017; International Antimony Association, 2018).

The *in vitro* tools and approaches summarized in Table 4.8 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 (3D) cell culture systems have been developed to score endpoints such as chromosome (i.e., MN) and DNA damage (i.e., comet assay). Several novel assays are summarized in Table 4.9.

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 and Luepke 1997, Wolf et al., 2003; Hothorn et al., 2013).

Table 4.9. Novel in vitro genotoxicity assessment systems based on multicellular, 3D constructs

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
3D MN test	MN frequency in reconstructed skin model	Traditional endpoint, 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

Test system	Principle	Advantages	Disadvantages, limitations	Key reference(s)
3D comet assay	DNA damage assay in reconstructed skin model	Traditional endpoint, simple to score, application in reconstructed skin models	Questions remain concerning metabolism	Reisinger et al., 2018; Pfuhler et al., 2014
Hen's Egg MN assay	MN frequency in extraembryonic peripheral blood of fertilized hen's eggs	Traditional endpoint, some metabolic capacity	Non-mammalian test, limited metabolism	Wolf et al., 1997, 2003; Hothorn et al., 2013

Advances in HT detection of DNA damage/repair, chromosome aberrations, and gene mutations may soon be eclipsed by error-corrected, next generation DNA sequencing (NGS) approaches. While 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 et al., 2018), with the precision and accuracy required to assess genetic alterations in only a few DNA molecules within a cell population. Though error-corrected NGS technologies are not yet fully validated or widely applied, the technology is rapidly advancing and may soon be routinely available, particularly since 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 subchronic or translational studies linking observations to humans.

4.5.3 Adverse outcome pathways for genotoxicity

The OECD Adverse Outcome Pathway (AOP) framework organises 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, 2017a). AOP also feed into IATA (Integrated Approaches to Testing and Assessment), a pragmatic approach to hazard

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

characterization that integrates *in silico*, *in vitro*, and *in vivo* assessment tools, including HT *in vitro* tools based on toxicogenomic or recombinant-cell reporter technologies (Sakuratano Sakuratani et al., 2018). The OECD IATA Case Studies Project reviews case studies related to different endpoints, including mutagenicity and genotoxicity,[8] and publishes the learnings and identified guidance needs (OECD, 2016, 2017b, 2018, 2019). The AOP on 'Alkylation of DNA in Male Pre-Meiotic Germ Cells Leading to Heritable Mutations' was the first AOP on genotoxicity published in the OECD AOP series (Yauk et al., 2016). To date, several other AOPs related to genotoxicity are under development in the AOP Wiki[9] (one module of the AOP Knowledge Base) and several ongoing initiatives should contribute to populating the AOP Knowledge Base with more AOPs on genotoxicity in the near future, increasing the development of AOP networks and supporting further tiered-testing and IATA strategies.

4.5.17.4 Quantitative utility for safety assessment

National and international genotoxicity evaluation committees have highlighted a desire to employ quantitative methods for regulatory interpretation of genotoxicity dose-response data (MacGregor et al., 2015a,b; UK COM, 2018). Lacking carcinogenicity data, quantitative analysis of *in vivo* genotoxicity dose-response data can be used for deriving MOEs (White and Johnson, 2016). This is particularly relevant for risk assessment and management of unavoidable food contaminants with positive genotoxicity results and/or genotoxicity structural alerts, and exposures exceeding the genotoxicity TTC of 0.0025 μg/kg bw/day (see 4.5.4.2). Such a quantitative approach requires a paradigm shift from hazard identification of gentoxicants and recognizes that compensatory cellular responses (i.e., DNA damage processing) are quantitatively manifested as mechanistically-plausible dose-response thresholds (Parry et al., 1994; Nohmi, 2008; Carmichael et al., 2009; Johnson et al., 2014;

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

[9] https://aopwiki.org/

Nohmi and Tsuzuki, 2016). With respect to the latter, this is still under debate, and there is currently no international consensus.

Several researchers have employed dose-response PoD (point of departure) values, such as the BMD (Benchmark Dose), Td (Threshold Dose), and NOGEL (No Observed Genotoxicity Effect Level) for quantitative interpretation of *in vitro* and *in vivo* genotoxicity dose-response data. With respect to *in vitro* dose-response data, the BMD approach has been used for MOA determinations, and to rank potency across test articles, cell types, and experimental protocols (Bemis et al., 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. *In vivo* mutagenicity data were used to determine a permissible daily exposure (PDE) to ethylmethanesulfonate, an impurity detected in Viracept, an antiretroviral drug (Gocke and Wall, 2009; Muller and Gocke, 2009). Although the regulatory utility of quantitative interpretation of *in vivo* dose-response data is increasingly recognised, use of genotoxicity-based BMD values to estimate MOEs for genotoxic food contaminants will require consensus regarding, for example, choice of test/endpoint, an appropriate BMR (Benchmark Response) for genotoxicity endpoints, and appropriate safety factors for exposure limit determination (Ritter et al., 2007; Nielsen et al., 2008; IPCS, 2014; Dankovic et al., 2015).

DRAFT

4.5.7 References

Amberg A et al. (2016) Principles and procedures for implementation of ICH M7 recommended (Q)SAR analyses, Reg Toxicol Pharmacol 77:13-24.

Amberg A et al. (2019) Principles and procedures for handling out-of-domain and indeterminate results as part of ICH M7 recommended (Q)SAR analyses, Regul Toxicol Pharmacol 102:53-64.

Ames BN, Lee FD, Durston WE (1973) An improved bacterial test system for the detection and classification of mutagens and carcinogens. Proc Nat Acad Sci 70(3):782-786.

Ates G, Favyts D, Hendriks G, 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.

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: 219-224.

Auerbach C, Robson JM, Carr JG (1947) The chemical production of mutations. Science 105(2723):243-247.

Avlasevich et al. (2011) Flow cytometric analysis of micronuclei in mammalian cell cultures: past, present and future. Mutagenesis 26:147-152.

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.

Barber C, Amberg A, Custer L, et al., (2015). Establishing best practise in the application of expert review of mutagenicity under ICH M7. Regul Toxicol Pharmacol, 73:367-77.

Barber C, Hanser T, Judson P, Williams R (2017) Distinguishing between expert and statistical systems for application under ICH M7. Regul Toxicol Pharmacol, 84:124-130.

Balbo S, Turesky RJ, Villalta PW. (2014) DNA adductomics. Chem Res Toxicol. 27(3):356-66.

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.

Bemis JC, Wills JW, Bryce SM et al. (2016) Comparison of in vitro and in vivo clastogenic potency based on benchmark dose analysis of flow cytometric micronucleus data. Mutagenesis 31:277-285.

Bemis JC, Heflich RH. (2019) In Vitro mammalian cell mutation assays based on the Pig-a gene: A report of the International Workshops on Genetic Toxicology (IWGT) workgroup. Mutat Res. https://doi.org/10.1016/j.mrgentox.2019.03.001

Bemis JC, Bryce SM, Nern M et al. (2016) Best practices for application of attachment cells to in vitro micronucleus assessment by flow cytometry. Mutat Res Genet Toxicol Environ Mutagen. 795:51-9.

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 31:329-331.

Benigni R, Bossa C, Battistelli CL (2013) ISSTOX Chemical Toxicity Databases, http://old.iss.it/meca/index.php?lang=1&anno=2013&tipo=25

Benigni R, Battistelli CL, Bossa C, 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. EFSA supporting publication, EN-1598. 221 pp. doi: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.

Beronius A, Ågerstrand M (2017) Making the most of expert judgment in hazard and risk assessment of chemicals. Toxicol Res (Camb). 6(5):571-577.

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.

Beyersmann D, Hartwig A (2008) Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms. Arch Toxicol 82(8):493-512.

Blackburn K, Stuard SB (2014) A framework to facilitate consistent characterization of read across uncertainty, Regul. Toxicol. Pharmacol. 68: 353-362.

Boobis A, Brown P, Cronin MTD 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 48(8):710-32.

Bossa, C., Benigni, R., Tcheremenskaia, O. Et al. (2018) (Q)SAR Methods for Predicting Genotoxicity and Carcinogenicity: Scientific Rationale and Regulatory Frameworks. Meth Mol Biol 1800, 447-473.

Brown K. Methods for the detection of DNA adducts. In Genetic Toxicology 2012 (pp. 207-230). Springer, New York, NY.

Bryce SM, Avlasevich SL, Bemis JC, 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.

Bryce et al. (2016) Genotoxic Mode of Action Predictions From a Multiplexed Flow Cytometric Assay and a Machine Learning Approach. Environ Molec Mutagen 57:171-189.

Bryce SM, Bernacki DT, Bemis JC, et al. (2017) Interlaboratory Evaluation of a Multiplexed High Information Content *In Vitro* Genotoxicity Assay. Environ Mol Mutagen 58(3):146-161.

Bryce et al. (2018) Investigating the Generalizability of the MultiFlow DNA Damage Assay and Several Companion Machine Learning Models With a Set of 103 Diverse Test Chemicals Tox Sci 162:146-166.

Burden N, Mahony C, Müller BP et al. (2015) Aligning the 3Rs with new paradigms in the safety assessment of chemicals. Toxicol 330:62-66.

CAC/GL 66-2008 - Guidelines for the use of flavourings. http://www.fao.org/fao-who-codexalimentarius/codex-texts/guidelines/en/

Carmichael P, Kirsch-Volders M, Vrijhof H (2009) Thresholds for genotoxins and their application in risk assessment. Mutat Res 678:71.

Cassano, 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, Part C 32, 273-298.

Chang YJ, Cooke MS, Hu CW et al (2018) Novel approach to integrated DNA adductomics for the assessment of in vitro and in vivo environmental exposures. Arch Toxicol. 92(8):2665-2680.

Chapman KE, Thomas AD, Wills JW et al. (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.

Cheeseman MA, Machuga EJ, Bailey AB (1999) A tiered approach to threshold of regulation. Food Chem Toxicol. 37, 387-412.

DRAFT

Cimino MC (2006) Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. Environ Mol Mut 47(5):362-390.

Cisbio link for gH2AX assay: https://www.cisbio.com/usa/drug-discovery/phospho-h2ax-ser139-cellular-assay-kit

Collins A, Koppen G, Valdiglesias V et al. (2014) The comet assay as a tool for human biomonitoring studies: the ComNet project. Review. Mutat Res. 759:27-39.

Committee on Mutagenicity (2018) Statement on the quantitative approaches to the assessment of genotoxicity data. Statement 2018/S1 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/704462/COM_statement_on_quantitative_approaches.pdf

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):772-784.

Corton JC, Witt KL, Yauk CL. 2019. Identification of p53 Activators in a Human Microarray Compendium. Chem Res Toxicol. *32*(9):1748-1759.

Corvi R, Madia F (2018) EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database of Ames Positive Chemicals (2018) European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-eurl-ecvam-genotoxicity-carcinogenicity-ames

Chervona Y, Arita A, Costa M (2012) Carcinogenic metals and the epigenome: understanding the effect of nickel, arsenic, and chromium. Metallomics **4**(7):619-627.

Costa M (2019) Review of arsenic toxicity, speciation and polyadenylation of canonical histones. Toxicol Appl Pharmacol **375**:1-4.

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, Reg Toxicol Pharmacol 106:90-104.

D'Gama AM, Pochareddy S, Li M et al. (2015) Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms. Neuron 88:910-917.

Dankovic DA, Naumann BD, Maier A et al. (2015) The Scientific Basis of Uncertainty Factors Used in Setting Occupational Exposure Limits. J Occup Environ Hyg 12:Suppl 1:S55-68.

Department of National Health and Welfare (1993) The assessment of mutagenicity. Health Protection Branch Mutagenicity Guidelines. Environ Molec Mutagen 21:15-37.

Dearfield KL, Cimino MC, McCarroll NE, et al. (2002) Genotoxicity risk assessment: a proposed classification strategy. Mutat Res/Gen Toxicol Environ Mutagen 521(1-2):121-135.

Dearfield KL, Gollapudi BB, Bemis JC et al. (2017) Next generation testing strategy for assessment of genomic damage: A conceptual framework and considerations. Environ Mol Mut 58(5):264-283.

Dertinger SD, Torous DK, Hall NE et al. (2002) Enumeration of micronucleated CD71-positive human reticulocytes with a single-laser flow cytometer. Mutat Res 515(1-2):3-14.

Dertinger SD 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.

Dertinger SD et al. (2011) Flow cytometric scoring of micronucleated erythrocytes: an efficient platform for assessing in vivo cytogenetic damage. Mutagenesis 26:139-145.

Dertinger SD, Avlasevich SL, Bemis JC et al. (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.

Dertinger SD, Phonethepswath S, Weller P et al. (2011) International Pig-a gene mutation assay trial: evaluation of transferability across 14 laboratories. Environ Mol Mutagen. 52(9):690-8.

Dewangan, J., Pandey, P.K., Divakar, A et al. (2018) Detection of Gene Mutation in Cultured Mammalian Cells. In *Mutagenicity: Assays and Applications* (pp. 49-67).

Dewhurst I, and Renwick AG (2013) Evaluation of the Threshold of Toxicological Concern (TTC) – Challenges and Approaches. Regul Toxicol Pharmacol 65:168–177.

Dobo KL, Greene N, Fred 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: 449-455.

Douglas GR, Jiao J, Gingerich JD et al. (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 USA, **92:**7485-7489.

ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals). 2012. ECETOC Technical Report 116 Category Approaches, Read-Across, (Q)SAR. ECETOC, Brussels, 2012. http://www.ecetoc.org/technical-reports

Eastmond DA, Hartwig A, Anderson D et al. (2009) Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS Harmonized Scheme Mutagenesis 24(4):341–349.

ECHA (European Chemicals Agency) (2008) Guidance on information requirements and chemical safety assessment: Chapter R.6 QSARs and grouping of chemicals.

ECHA (2011) European Chemical Agency. Guidance on information requirements and chemical safety assessment. Chapter R.4: Evaluation of available information. https://echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf/d6395ad2-1596-4708-ba86-0136686d205e (accessed on 30 November 2018) https://echa.europa.eu/-/pathfinder-chapters (to check for the current version)

ECHA (European Chemicals Agency) (2017a) Read-Across Assessment Framework (RAAF). ECHA-17-R-01-EN

ECHA (European Chemicals Agency) (2017b) The use of alternatives to testing on animals for the REACH Regulation. Third report under Article 117(3) of the REACH Regulation. ECHA-17-R-02-EN

ECHA (2017) European Chemical Agency. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7a: Endpoint specific guidance. https://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf/e4a2a18f-a2bd-4a04-ac6d-0ea425b2567f (accessed on 30 November 2018)

https://echa.europa.eu/-/pathfinder-chapters (to check for the current version)

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 Journal 2009; 7(9):1249. [19 pp.]. https://www.efsa.europa.eu/de/efsajournal/pub/1249

EFSA (2011) EFSA Scientific Committee; Scientific Opinion on genotoxicity testing strategies applicable to food and feed safety assessment. EFSA Journal 2011;9(9):2379. [69 pp.] doi:10.2903/j.efsa.2011.2379. Available online: www.efsa.europa.eu/efsajournal

EFSA Panel on Plant Protection Products and their Residues (PPR) (2013) Scientific Opinion on the identification of pesticides to be included in cumulative assessment groups on the basis of their toxicological profile, EFSA Journal 2013;11(7):3293

EFSA (2016a) European Food Safety Authority. NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), Guidance on the preparation and presentation of an application for authorisation of a novel food in the context of Regulation (EU) 2015/2283. EFSA Journal 2016;14(11):4594, 24 pp http://www.efsa.europa.eu/en/efsajournal/pub/4594

EFSA (2016b). EFSA Panel on Plant Protection products and their Residues. Guidance on the establishment of the residue definition for dietary risk assessment. EFSA Journal. 14, e04549-n/a. 10.2903/j.efsa.2016.4549

EFSA (2018) European Food Safety Authority. Scientific Committee. Statement on genotoxicity assessment of chemical mixtures. Adopted 22 November 2018. https://www.efsa.europa.eu/en/consultations/call/180626

EFSA (2019) 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 Journal 2019;17(3):5634, 77 pp. https://doi.org/10.2903/j.efsa.2019.5634

EFSA (European Food Safety Authority)/WHO (World Health Organization). 2016. Review of the Threshold of Toxicological Concern (TTC) approach and development of new TTC decision tree. EFSA supporting publication 2016: EN-1006. 50 pp.

EMA (2006) European Medicines Agency. Committee on Herbal Medicinal Products (HMPC). Guideline on non-clinical documentation for herbal medicinal products in applications for marketing authorization (bibliographical and mixed applications) and in applications for simplified registration (EMEA/HMPC/32116/2005). London, 7 September 2006. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003576.pdf

EMA (2009) European Medicines Agency. Committee on Herbal Medicinal Products (HMPC). Guideline on selection of test materials for genotoxicity testing for traditional herbal medicinal products/herbal medicinal products. (EMEA/HMPC/67644/2009). London, 12 November 2009. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500059583.pdf

EMA (2011) European Medicines Agency. Specifications: test procedures and acceptance criteria for herbal substances, herbal preparations and herbal medicinal products/traditional herbal medicinal products.
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 543:125-136.

Erickson RP (2010) Somatic gene mutation and human disease other than cancer: an update. Mutat Res 705:96-106.

FAO (2001) Smoke flavourings.
http://www.fao.org/fileadmin/user_upload/jecfa_additives/docs/Monograph1/Additive-386.pdf

Felter S, Lane RW, Latulippe ME et al. (2009). Refining the Threshold of Toxicological Concern for risk prioritization of trace chemicals in food. Food Chem. Toxicol. 47: 2236–2245.

Fenech M, Kirsch-Volders M, Rossnerova A 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.

Fischer BM, Neumann D, Piberger AL et al. (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):2745-2761.

Ge J, Chow DN, Fessler JL et al. (2015) Micropatterned comet assay enables high throughput and sensitive DNA damage quantification. Mutagenesis. 30:11–19.

Greene N et al (2015) A practical application of two in silico systems for identification of potentially mutagenic impurities. Regul Toxicol Pharmacol. 72(2):335-49.

Gocke E, Wall M (2009) In vivo genotoxicity of EMS: statistical assessment of the dose response curves. Toxicol Lett 190:298-302.

Gollapudi BB, Lynch AM, Heflich RH, et al. (2015) The in vivo Pig-a assay: A report of the International Workshop On Genotoxicity Testing (IWGT) Workgroup. Mutat Res Genet Toxicol Environ Mutagen 783:23-35.

Görög S (2018) Critical review of reports on impurity and degradation product profiling in the last decade. Trends in Analytical Chemistry, 101:2-16.

Guo X., Zhang S, Dial SL et al. (2014) In vitro investigation of the mutagenic potential of Aloe vera extracts. Toxicol. Res. 3: 487-496.

Guo X, Heflich RH, Dial SL et al. (2018) Quantitative differentiation of whole smoke solution-induced mutagenicity in the mouse lymphoma assay. Environ Mol Mutagen 59:103-113.

Hanser T, Barber C, Marchaland JFF, Werner S (2016) Applicability domain: towards a more formal definition, SAR QSAR Environ. Res., 27: 865-881.

Hanway, R. H. and P.F.Evans. (2000) Read-across of toxicological data in the notification of new chemicals, Toxicol. Lett. 116 (Suppl. 1):61.

Hartwig A (2013) Metal interaction with redox regulation: an integrating concept in metal carcinogenesis? Free Radic Biol Med 55:63-72.

Hasselgren et al (2019) Genetic toxicology in silico protocol. Regul Toxicol Pharmacol, 107:104403.

Hastwell PW, Chai LL, Roberts KJ, 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 607:160-75.

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.

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.

Hendricks CA, Almeida KH, Stitt MS et al. (2003) Spontaneous mitotic homologous recombination at an enhanced yellow fluorescent protein (EYFP) cDNA direct repeat in transgenic mice. Proc Natl Acad Sci USA 100:6325-633

Hendriks G, Atallah M, Morolli B, et al. (2012) The ToxTracker assay: novel GFP reporter systems that provide mechanistic insight into the genotoxic properties of chemicals. Toxicol Sci 125:285–298.

Hendriks G, Derr RS, Misovic B, et al. (2016) The Extended ToxTracker Assay Discriminates Between Induction of DNA Damage, Oxidative Stress, and Protein Misfolding. Toxicol Sci 150:190–203.

HGDB (Human Gene Mutation Database). 2017. http://www.hgmd.cf.ac.uk/ac/index.php

Honma M, Kitazawa A, Cayley A, et al. (2019) Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challege Project. Mutagenesis 34(1): 3-16.

Hothorn LA, Reisinger K, Wolf T, et al. (2013). Statistical analysis of the hen's egg test for micronucleus induction (HET-MN assay). Mutat Res 757: 68-78.

Hooijmans CR, Rovers MM, de Vries RB, et al. (2014) SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol.,14:43.

Hsieh J-H, Smith-Roe SL, Huang R, et al. Identifying Compounds with Genotoxicity Potential Using Tox21 High Throughput Screening Assays. Chem Res Toxicol 32(7):1384-1401.

ICH (International Council for Harmonisation). 2017. Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk Guidance m7 (R1). Available online: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_R1_Addendum_Step_4_31Mar2017.pdf

ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use). (2011) Impurities: Guidelines for Residual Solvents Q3C(R5)

ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use). (2014) Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk (M7)

IPCS (International Program on Chemical Safety) (2014) Guidance Document on Evaluating and Expressing Uncertainly in Hazard Characterization. Harmonization Project Document 11.

JECFA (1972) Toxicological evaluation of some enzymes, modified starches and certain other substances. Fifteenth meeting of the Joint FAO/WHO Expert Committee on Food Additives. http://www.inchem.org/documents/jecfa/jecmono/v001je01.htm

JECFA (1987) http://www.inchem.org/documents/jecfa/jecmono/v22je07.htm

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

JECFA (2006) - General Specifications and Considerations for Enzyme Preparations Used in Food Processing - http://www.fao.org/docrep/009/a0691e/A0691E03.htm. Accessed 2 Aug 2018).

JECFA (2018), Evaluation of certain food additives (Eighty-sixth report of the Joint FAO/WHO Expert Committee on Food Additives) WHO Technical Report Series, No. 1014, 2019

JMPR (2019) Pesticide Residues in Food – Evaluations 2018, Part II Toxicological. https://apps.who.int/iris/bitstream/handle/10665/325787/9789241655330-eng.pdf?ua=1

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

Johnson GE, Soeteman-Hernandez LG, Gollapudi BB et al. (2014) Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. Environ Mol Mutagen 55:609-623.

Jones NJ (2012) ^{32}p-Postlabelling for the sensitive detection of DNA adducts. Methods Mol Biol 817:183-206.

Kanwal R, Gupta K, Gupta S (2015) Cancer epigenetics: an introduction Methods Mol Biol 1238:3-25.

Khanal S, Singh P, Avlasevich SL et al. (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: 30-35.

Kleinstreuer NC, Karmaus A, Mansouri K et al. (2018) Predictive Models for Acute Oral Systemic Toxicity: A Workshop to Bridge the Gap from Research to Regulation. Comput Toxicol 8:21-4.

Kimoto T, Horibata K, Miura D 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.

Kirkland D, Pfuhler S, Tweats D 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. Mutation Res 628:31-55.

Kirkland D, Reeve L, Gatehouse D, Vanparys P (2011) A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. Mutat. Res. 721: 27-73.

Kirkland D, Uno Y, Luijten M, Beevers C, van Benthem J, Burlinson B, Dertinger S, Douglas GR, Hamada S, Horibata K, Lovell DP, Manjanatha M, Martus HJ, Mei N, Morita T, Ohyama W, Williams A. In vivo genotoxicity testing strategies: Report from the 7th International workshop on genotoxicity testing (IWGT). Mutat Res. 2019 Nov;847:403035.

Kirsch-Volders M, Decordier I, Elhajouji A et al. (2011) In vitro genotoxicity testing using the micronucleus assay in cell lines, human lymphocytes and 3D human skin models. Mutagenesis 26(1):177-84.

Kligerman AD, Allen JW, Erexson GL, Morgan DL (1993) IARC Sci Publ 127:217-224.

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.

Kroes R, Renwick AG, Cheeseman M 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.

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: 2429-2443.

Krüger CT, Fischer BM, Armant O et al. (2016) The in vitro PIG-A gene mutation assay: glycosylphosphatidylinositiol (GPI)-related genotype-to-phenotype relationship in TK6 cells. Arch Toxicol 90, 1729-1736.

Kucab JE1, 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.

Kuseva C, Schultz TW, Yordanova D, et al. (2019) The implementation of RAAF in the OECD QSAR Toolbox, Regul Toxicol Pharmacol 105: 51-61.

Lai Y, Yu R, Hartwell HJ et al. (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.

Langie SA, Koppen G, Desaulniers D et al. (2015) "Causes of genome instability: the effect of low dose chemical exposures in modern society." Carcinogenesis 36 Suppl 1:S61-88.

Leonard JA, Nelms M, Craig E 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. Reg Toxicol Pharmacol. 107:104419.

Lester et al. (2018) Structure activity relationship (SAR) toxicological assessments: The role of expert judgment. Regul Toxicol Pharmacol 92: 390–406

Li HH, Chen R, Hyduke DR 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.

Li HH, Hyduke DR, Chen R 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.

Lloyd M. and Kidd D. 2012. The mouse lymphoma assay. In *Genetic Toxicology* (pp. 35-54). Springer, New York, NY.

Luo JL, Yang Q, Tong WM et al. (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 20:320-328.

Lupski JR (2013) Genetics. Genome mosaicism--one human, multiple genomes. Science 341:358-359.

MacGregor JT, Bishop ME, McNamee JP, 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.

MacGregor JT, Frötschl R, White PA 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 - Gen Toxicol Environ Mutagenesis 783:55-65.

MacGregor JT, Frötschl R, White PA 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 - Gen Toxicol Environ Mutagenesis 783:66-78.

Mansouri, K., A. Abdelaziz, A. Rybacka et al. (2016). CERAPP: collaborative estrogen receptor activity prediction project. Environ Health Perspect 124:1023-1033.

Marchetti F, Aardema MJ, Beevers C, 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/Gen Toxicol Environ Mutagen 832:7-18.

Masumura K, Toyoda-Hokaiwado N, Ukai A et al. (2016a) Estimation of the frequency of inherited germline mutations by whole exome sequencing in ethyl nitrosourea-treated and untreated *gpt* delta mice. Genes and Environment. 38. doi: 10.1186/s41021-016-0035-y

Masumura K, Toyoda-Hokaiwado N, Ukai A et al. (2016b) Dose-dependent de novo germline mutations detected by whole-exome sequencing in progeny of ENU-treated male *gpt* delta mice, Mutat Res 810: 30-39.

Meier MJ, O'Brien JM, Beal MA et al. (2017) In Utero Exposure to Benzo[a]Pyrene Increases Mutation Burden in the Soma and Sperm of Adult Mice. Environ Health Perspect 125:82-88.

Mellor CL, Marchese Robinson RL, Benigni R, et al. (2019) Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use, Regul Toxicol Pharmacol 101: 121-134.

Miura D, Dobrovolsky VN, Kasahara Y et al. (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: 614-621.

Miura D, Dobrovolsky VN, Mittelstaedt RA et al. (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: 622-630.

Molander L, Agerstrand M, Beronius A et al. (2015). Science in Risk Assessment and Policy (SciRAP): An Online Resource for Evaluating and Reporting In Vivo (Eco)Toxicity Studies, Hum. Ecol. Risk Assess. 21:753–762.

Mombelli E, Raitano G, Benfenati E (2016) In Silico Prediction of Chemically Induced Mutagenicity: How to Use QSAR Models and Interpret Their Results. Meth Mol Biol, 1425, 87-105.

Muller HJ. (1927) Artificial transmutation of the gene. Science 66(1699):84-87.

Muller L, Gocke E (2009) Considerations regarding a permitted daily exposure calculation for ethyl methanesulfonate. Toxicol Lett 190:330-332.

Myatt GJ et al. (2018) In silico toxicology protocols. Regul Toxicol Pharmacol. 96: 1-17.

Nelms MD, Mellow CL, Enoch SJ et al. (2018). A Mechanistic Framework for Integrating Chemical Structure and High-Throughput Screening Results to Improve Toxicity Predictions. Comput Toxicol. 8:1-12.

Nervi C, Fazi F, Grignani F (2008) Oncoproteins, heterochromatin silencing and microRNAs: a new link for leukemogenesis. Epigenetics. 3(1):1-4.

Netzeva TI, Worth AP, Aldenberg T, 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, ATLA 33:155-173.

Nielsen E, Ostergaard G, Larsen J (2008) Toxicological Risk Assessment of Chemicals: A Practical Guide. Informa Healthcare USA, New York, NY.

Nishihara K, Huang R, Zhao J et al. (2016) Identification of genotoxic compounds using isogenic DNA repair deficient DT40 cell lines on a quantitative high throughput screening platform. Mutagen 31(1):69-81.

Nohmi T, Suzuki T, Masumura K (2000) Recent advances in the protocols of transgenic mouse mutation assays. Mutat Res 455:191-215.

Nohmi H (2008) Possible Mechanisms of Practical Thresholds for Genotoxicity. Genes and Environ. 30:108-113.

Nohmi T, Tsuzuki T (2016) Possible Mechanisms Underlying Genotoxic Thresholds: DNA Repair and Translesion DNA Synthesis. In Nohmi, T. and Fukushima, S. (eds.) Thresholds of Genotoxic Carcinogens. From Mechanisms to Regulation. Academic Press, Tokyo.

Nohmi T, Masumura K & Toyoda- Hokaiwado N (2017) Transgenic rat models for mutagenesis and carcinogenesis. Genes and Environ., **39**: 11.

Nohmi T (2018) Thresholds of Genotoxic and Non-Genotoxic Carcinogens. Toxicol Res 34: 281-290.

OECD (2005) Manual for the investigation of HPV chemicals. Chapter 3.1 Guidance for Determining the Quality of Data for the SIDS Dossier (Reliability, Relevance and Adequacy) (Last updated: December 2005).

OECD (2011) – (To be included).

OECD (2014). Guidance on Grouping of Chemicals, Second Edition. Series on Testing & Assessment No. 194. ENV/JM/MONO(2014)4. OECD Publishing, Paris

OECD (2016), 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, Series on Testing & Assessment No. 250, ENV/JM/MONO(2016)48, OECD, Paris. France, 2016

OECD (2017), Report on Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA), Second Review Cycle (2016), Series on Testing & Assessment No. 270, ENV/JM/MONO(2017)22, OECD, Paris. France, 2017

OECD (2018), Report on Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA), Third Review Cycle (2017), Series on Testing & Assessment No. 289, ENV/JM/MONO(2018)25, OECD, Paris. France, 2018.

OECD (2019), Report on Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA), Fourth Review Cycle (2018), Series on Testing & Assessment No. XXX, ENV/JM/MONO(2019)XX, OECD, Paris. France, 2019, in press.

Ohta T (2006) Mutagenic activity of a mixture of heterocyclic amines at doses below the biological threshold level of each. Genes and Environment. 28: 181-184.

Olsen AK, Dertinger SD, Kruger CT, 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.

Pariza MW, Foster EM (1983) Determining the safety of enzymes used in food processing. J. Food Protect. 46: 453-468.

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: 173-186.

Parry JM, Fielder RJ, McDonald A (1994) Thresholds for aneuploidy-inducing chemicals. Advisory Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment of the UK Department of Health. Mutagenesis 9:503-504.

Patlewicz G, Ball N, Becker RA et al. (2014). Food for thought: Read-across approaches - misconceptions, promises and challenges ahead. Altex 2014, 31(4): 387-396.

Patlewicz G., Ball N., Boogaard PJ et al. (2015) Building scientific confidence in the development and evaluation of read-across. Reg Toxicol Pharmacol 72(1):117-133.

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.

Patlewicz G, Cronin MTD, Helman G et al. (2018) Navigating through the minefield of read-across frameworks: A commentary perspective. Comp Toxicol, 6: 39-54.

Patlewicz G, et al. (2019) Exploring current read-across applications and needs among selected U.S. Federal Agencies. Regulatory Toxicology and Pharmacology 106: 197-209.

Pawar G, Madden JC, Ebbrell DJ, et al. (2019) In Silico Toxicology Data Resources to Support Read-Across and (Q)SAR. Front Pharmacol, accepted

Petkov PI et al (2019) Procedure for toxicological predictions based on mechanistic weight of evidences: Application to Ames mutagenicity, Comp Toxicol, 12. 100009.

Pfuhler S, Fautz R, Ouedraogo G, Latil A, Kenny J, Moore C, Diembeck W, Hewitt NJ, Reisinger K, Barroso J (2014) The Cosmetics Europe strategy for animal-free genotoxicity testing: project status update. Toxicol In Vitro 28(1):18-23.

Phillips DH (1997) Detection of DNA damage by the 32P-postlabelling assay. Mutat Res 378(1-2):1-12.

Phillips, D.H., Farmer, P.B., Beland, F.A., Nath, R.G., Poirier, M.C., Reddy, M.V. and Turteltaub, K.W., 2000. Methods of DNA adduct determination and their application to testing compounds for genotoxicity. Environ Mol Mut, 35(3):222-233.

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 71:295-300.

Pozniak A, Müller L, Salgo M, et l. (2009) Elevated ethyl methanesulfonate (EMS) in nelfinavir mesylate (Viracept®, Roche): overview. AIDS Res Therapy, 6(1):18.

Raitano et al (2019) Integrating in silico models for the prediction of mutagenicity (Ames test) of botanical ingredients of cosmetics, Computational Toxicology, 12:100108.

Russell LB & Shelby MD (1985) Tests for heritable genetic damage and for evidence of gonadal exposure in mammals. Mutat Res, **154**: 69-84.

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**: 25-36.

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:S63-68.

Rappaport SM, Li H, Grigoryan H et al. (2012) Adductomics: characterizing exposures to reactive electrophiles. Toxicol Lett. 213(1):83-90.

Reisinger K, Blatz V, Brinkmann J et al. (2018) Validation of the 3D Skin Comet assay using full thickness skin models: Transferability and reproducibility. Mutat Res. 827:27-41.

Riebeling C, Luch A, Tralau T (2018) Skin toxicology and 3Rs-Current challenges for public health protection. Exp Dermatol 27(5):526-536.

Ritter L, Totman C, Krishnan K et al. (2007) Deriving uncertainty factors for threshold chemical contaminants in drinking water. J Toxicol Environ Health B Crit Rev 10:527-557.

Schneider K, Schwarz M, Burkholder I et al. (2009) "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicol Lett. 189(2):138-144.

Schultz, T.W., P. Amcoff, E. Berggren et al. (2015) A strategy for structuring and reporting a read-across prediction of toxicity, Regul. Toxicol. Pharmacol. 72:586-601.

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.

Scientific Committee on Food (1991) - (http://aei.pitt.edu/40836/1/27th_good.pdf Accessed 2 Aug 2018.

Serafimova R, Fuart-Gatnik M, Worth A (2010) Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity. EUR 24427 EN.

Shah I., Liu J, Judson RS et al. (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.

Shelby MD (1996) Selecting chemicals and assays for assessing mammalian germ cell mutagenicity. Mutat Res, **352**, 159-167.

Simpson K, Bevan, N, Hastwell P, et al. (2013) The BlueScreen-384 assay as an indicator of genotoxic hazard potential in early-stage drug discovery. J Biomol Screen 18(4):441-452.

Smith-Roe SL, Swartz CD, Shepard KG et al. (2018) Black cohosh extracts and powders induce micronuclei, a biomarker of genetic damage, in human cells. Environ Mol Mutagen59(5):416-426.

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(1):28.

Sukup-Jackson MR, Kiraly O, Kay JE 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:e1004299.

Sutter et al. (2013) Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Reg Toxicol Pharmacol. 67: 39–52.

Suzuki Y, Umemura T, Ishii Y et al. (2012) Possible involvement of sulfotransferase 1A1 in estragole-induced DNA modification and carcinogenesis in the livers of female mice. *Mutat Res/Gen Toxicol Environ Mut*, *749*(1-2), pp.23-28.

Swenberg, J.A., 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.

Sykora P, Witt KL, Revanna P et al. (2018) Next generation high throughput DNA damage detection platform for genotoxic compound screening. Sci Rep. 8(1):2771.

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.

Tcheremenskaia O, Battistelli CL, Giuliani A, et al. (2019) In silico approaches for prediction of genotoxic and carcinogenic potential of cosmetic ingredients. Comp Toxicol 11: 91-100.

Teasdale A, Elder DP (2018) Analytical control strategies for mutagenic impurities: Current challenges and future opportunities? *Trends in Analytical Chemistry*, 101, 66-84.

Thybaud V, Dean S, Nohmi T et al. (2003) *in vivo* transgenic mutation assays. Mutat Res, **540**: 141-151.

Ton ND, Nakagawa H, Ha NH et al. (2018) Whole genome sequencing and mutation rate analysis of trios with paternal dioxin exposure. Human Mutat, 39, 1384-1392.

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):91-99.

Tweats DJ, Johnson GE, Scandale I et al. (2016) Genotoxicity of flubendazole and its metabolites in vitro and the impact of a new formulation on in vivo aneugenicity. Mutagenesis 31:309-321.

UK Committee on Mutagenicity (2018). Statement on the quantitative approaches to the assessment of genotoxicity data. Statement 2018/S1

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/704462/COM_statement_on_quantitative_approaches.pdf

Uno et al. 2015a. Recommended protocols for the liver micronucleus test: Report of the IWGT working group. Mutat Res. 783:13-18.

Uno Y., Morita T, Luijten M et al. 2015b. Micronucleus test in rodent tissues other than liver or erythrocytes: Report of the IWGT working group. Mutat Res 783:19-22.

USEPA (2019). OncoLogic https://www.epa.gov/tsca-screening-tools/oncologictm-computer-system-evaluate-carcinogenic-potential-chemicals

US FDA (2016) US Food and Drug Administration. Botanical drug development, Guidance for Industry. December 2016. https://www.fda.gov/downloads/Drugs/Guidances/UCM458484.pdf

US FDA (1995) Food Additives: Threshold of regulation of substances used in food-contact articles: Final Rule. Federal Register 60:36582-36596.

Van Bossuyt et al. (2018) Performance of in silico models for mutagenicity prediction of food contact materials, Toxicol Sci 163(2): 632-638.

Verma, J.R., Rees, B.J., Wilde, E.C. et al. (2017) Evaluation of the automated MicroFlow((R)) and Metafer platforms for high-throughput micronucleus scoring and dose response analysis in human lymphoblastoid TK6 cells. Arch Toxicol 91:2689-2698.

Vuorinen et al (2017) Applicability of in silico genotoxicity models on food and feed Ingredients, Regulatory Toxicology and Pharmacology 90: 277-288.

Wang, N.C., Q. J. Zhao, S.C. Wesselkamper et al. (2012) Application of computational toxicological approaches in human health risk assessment. I. A tiered surrogate approach, Regul. Toxicol. Pharmacol. 63:10-19.

Wetmore BA (2015) Quantitative in vitro-to-in vivo extrapolation in a high-throughput Environment. Toxicology 332:94-101.

White PA, Johnson GE (2016) Genetic toxicology at the crossroads-from qualitative hazard evaluation to quantitative risk assessment. Mutagenesis 31:233-237.

White, P.A M. Luijten, M. Mishima 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, https://doi.org/10.1016/j.mrgentox.2019.04.002

Wichard J (2017) In silico prediction of genotoxicity, Food and Chemical Toxicology 106: 595-599

Williams A, Buick JK, Moffat I et al. (2015) A predictive toxicogenomics signature to classifygenotoxic versus non-genotoxic chemicals in human TK6 cells. Data Brief. 5:77-83.

Wills JW, Johnson GE, Doak SH et al. (2016) Empirical analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings and support MOA determinations. Mutagenesis 31:255-263.

Witt, K. L., Knapton A., Wehr, C, M., Hook, G, J., Mirasalis, J., Shelby, M. D.j and MacGregor, J.T. (2000). Micronucleated erythrocyte frequency in peripheral blood of B6C3F(1) mice from short-term, prechronic, and chronic studies of the NTP carcinogenesis bioassay program. 36(3):163-94.

Witt KL, Cunningham CK, Patterson KB 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.

Witt KL, Livanos E, Kissling GE 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 8;649(1-2):101-13.

Witt KL, Hsieh JH, Smith-Roe SL 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(7):494-507.

Wolf T, Luepke N-P (1997) Formation of micronuclei in incubated hen's eggs as a measure of genotoxicity, Mutat Res 394:163–175.

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 41:561–573.

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. EFSA External Scientific Report. European Commission Joint Research Centre, Institute for Health & Consumer Protection, Ispra, Italy https://doi.org/10.2903/sp.efsa.2010.EN-50

Wu S, Blackburn K, Amburgey J et al. (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. 56:67-81.

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.

Yamamoto KN, Hirota K, Kono K 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.

Yauk CL (2004) Advances in the application of germline tandem repeat instability for in situ monitoring. Mutat Res, **566:**169-182.

Yauk CL, Buick JK, Williams A et al. (2016) 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.

Yauk, C., et al. (2016) Adverse Outcome Pathway on Alkylation of DNA in Male Pre-Meiotic Germ Cells Leading to Heritable Mutations, *OECD Series on Adverse Outcome Pathways*, No. 3, OECD Publishing, Paris, https://doi.org/10.1787/5jlsvvxn1zjc-en.

Youssoufian H, Pyeritz RE (2002) Mechanisms and consequences of somatic mosaicism in humans. Nat Rev Genet 3:748-758.

Yu Y, Wang P, Cui Y, Wang Y (2018) Chemical Analysis of DNA Damage. Anal Chem 90(1):556-576.

Zeiger, E., 2004. History and rationale of genetic toxicity testing: an impersonal, and sometimes personal, view. Environ Mol Mutag 44(5):363-371.

Zhu, H., M. Bouhifd, E. Donley et al. (2016) Supporting read-across using biological data, ALTEX 33:167-182.