ENDORSED FOR PUBLIC CONSULTATION

DRAFT SCIENTIFIC OPINION

Scientific opinion on genotoxicity testing strategies applicable to food and feed safety assessment [1]

EFSA Scientific Committee[2, 3]

European Food Safety Authority (EFSA), Parma, Italy

SUMMARY

At the request of the European Food Safety Authority, the Scientific Committee has reviewed the current state-of-the-science on genotoxicity testing strategies and provided a commentary and recommendations on testing strategies, bearing in mind the needs of EFSA's various Scientific Panels to have appropriate data for risk assessment. It is hoped that this opinion will contribute to greater harmonisation between EFSA Panels on approaches to such testing.

The purpose of genotoxicity testing for risk assessment of substances in food and feed is:

- to identify substances which could cause heritable damage in humans,
- to predict potential genotoxic carcinogens in cases where carcinogenicity data are not available, and
- to contribute to understanding of the mechanism of action of chemical carcinogens.

For an adequate evaluation of the genotoxic potential of a chemical substance, different end-points, i.e. induction of gene mutations, structural and numerical chromosomal alterations, need to be assessed, as each of these events has been implicated in carcinogenesis and heritable diseases. An adequate coverage of all the above mentioned end-points can only be obtained by the use of more than one test system, as no individual test can simultaneously provide information on all these end-points.

In reaching its recommendations for a basic test battery, the Scientific Committee has considered:

- past experience with various tests when combined in a basic battery,
- the availability of guidelines or internationally accepted protocols,
- the performance of *in vitro* and *in vivo* tests in prediction of rodent carcinogenesis,
- correlations between *in vitro* and *in vivo* positive results for genotoxicity,

1 On request from EFSA, Question No EFSA-Q-2009-00782, endorsed for public consultation on 5 April 2011.

2 Scientific Committee members: Boris Antunović, Susan Barlow, Andrew Chesson, , Albert Flynn, Anthony Hardy, Klaus-Dieter Jany, Michael-John Jeger, Ada Knaap, Harry Kuiper, John-Christian Larsen, David Lovell, Birgit Noerrung, Josef Schlatter, Vittorio Silano, Frans Smulders and Philippe Vannier. Correspondence: scientific.committee@efsa.europa.eu.

3 Acknowledgement: The Scientific Committee wishes to thank the members of the Working Group on Genotoxicity testing Strategies: Gabriele Aquilina, Susan Barlow, Mona Lise Binderup, Claudia Bolognesi, Paul Brantom, Raffaella Corvi, Riccardo Crebelli, Eugenia Dogliotti, Metka Filipic, Corrado Galli (member until February 2011), Rainer Guertler, Andrea Hartwig, Peter Kasper, David Lovell, Daniel Marzin, Jan van Benthem for the preparatory work on this scientific opinion; the hearing expert David Kirkland; EFSA's staff member Daniela Maurici for the support provided to this scientific opinion.

Suggested citation: EFSA Scientific Committee; Draft Scientific Opinion on Genotoxicity Testing Strategies applicable in food and feed safety assessement. European Food Safety Authority (2011). Available online: www.efsa.europa.eu

- the minimum number of tests necessary to achieve adequate coverage of the three required endpoints, and
- the need to avoid unnecessary animal tests.

The Scientific Committee recommends a step-wise approach for the generation and evaluation of data on genotoxic potential, comprising:

- a basic battery of *in vitro* tests,
- consideration of whether specific features of the test substance might require substitution of one or more of the recommended *in vitro* tests by other *in vitro* or *in vivo* tests in the basic battery,
- in the event of positive results from the basic battery, review of all the available relevant data on the test substance, and
- where necessary, conduct of an appropriate *in vivo* study (or studies) to assess whether the genotoxic potential observed *in vitro* is expressed *in vivo.*

The Scientific Committee recommends use of the following two *in vitro* tests as the first step in testing:

- a bacterial reverse mutation assay (OECD TG 471), and
- an *in vitro* micronucleus test (OECD TG 487).

This combination of tests fulfils the basic requirements to cover the three genetic endpoints with the minimum number of tests; the bacterial reverse mutation assay covers gene mutations and the *in vitro* micronucleus test covers both structural and numerical chromosome aberrations. The Scientific Committee concluded that these two tests are reliable for detection of most potential genotoxic substances and that the addition of any further *in vitro* mammalian cell tests in the basic battery would significantly reduce specificity with no substantial gain in sensitivity.

The Scientific Committee did consider whether an *in vivo* test should be included in the first step of testing and broadly agreed that it should not be routinely included. However, if there are indications for the substance of interest that specific metabolic pathways would be lacking in the standard *in vitro* systems, or it is known that the *in vitro* test system is inappropriate for that substance or for its mode of action, testing may require either appropriate modification of the *in vitro* tests or use of an *in vivo* test at an early stage of testing. The Scientific Committee also recognised that in some cases it may be advantageous to include *in vivo* assessment of genotoxicity at an early stage, if, for example, such testing can be incorporated within other repeated-dose toxicity studies that will be conducted anyway.

If all *in vitro* endpoints are clearly negative in adequately conducted tests, then it can be concluded with reasonable certainty that the substance has no genotoxic potential.

In the case of inconclusive, contradictory or equivocal results from *in vitro* testing, it may be appropriate to conduct further testing *in vitro*, either by repetition of a test already conducted, perhaps under different conditions, or by conduct of a different *in vitro* test, to try to resolve the situation. In the case of positive results from the basic battery of tests, it may be that further testing *in vitro* is appropriate to optimise any subsequent *in vivo* testing, or to provide additional useful mechanistic information.

Before embarking on any necessary follow-up of positive *in vitro* results by *in vivo* testing, not only the results from the *in vitro* testing should be reviewed, but also other relevant data on the substance, such as information about chemical reactivity of the substance (which might predispose to site of contact effects), bioavailability, metabolism, toxicokinetics, and any target organ specificity. Additional useful information may come from structural alerts and 'read-across' from structurally related substances. It may be possible after this to reach a conclusion to treat the substance as an *in vivo* genotoxin. If, after such a review, a decision is taken that *in vivo* testing is necessary, tests should

be selected on a case-by-case basis using expert judgement, with flexibility in the choice of test, guided by the full data set available for the substance.

In vivo tests should relate to the genotoxic endpoint(s) identified as positive *in vitro* and to appropriate target organs or tissues. Evidence, either from the test itself or from other toxicokinetic or repeated-dose toxicological studies, that the target tissue(s) have been exposed to the test substance and/or its metabolites is essential for interpretation of negative results.

The approach to *in vivo* testing should be step-wise. If the first test is positive, no further test is needed and the substance should be considered as an *in vivo* genotoxin. If the test is negative, it may be possible to conclude that the substance is not an *in vivo* genotoxin. However, in some cases, a second *in vivo* test may be necessary as there are situations where more than one endpoint in the *in vitro* tests is positive and an *in vivo* test on a second endpoint may then be necessary if the first test is negative. It may also be necessary to conduct a further *in vivo* test on an alternative tissue if, for example, it becomes apparent that the substance did not reach the target tissue in the first test. The combination of assessing different endpoints in different tissues in the same animal *in vivo* should be considered.

The Scientific Committee recommends the following as suitable *in vivo* tests:

- an *in vivo* micronucleus test (OECD TG 474),
- an *in vivo* Comet assay (no OECD TG at present; internationally agreed protocols available), and
- a transgenic rodent assay (draft OECD TG available).

The *in vivo* micronucleus test covers the endpoints of structural and numerical chromosomal aberrations and is an appropriate follow-up for *in vitro* clastogens and aneugens. There may be circumstances in which an *in vivo* mammalian bone marrow chromosome aberration test (OECD TG 475) may be an alternative follow-up test.

The *in vivo* Comet assay is considered a useful indicator test in terms of its sensitivity to substances which cause gene mutations and/or structural chromosomal aberrations and can be used with many target tissues. Transgenic rodent assays can detect point mutations and small deletions and are without tissue restrictions.

The Scientific Committee concluded that routine testing for genotoxicity in germ cells is not necessary. A substance that is concluded to be positive in tests in somatic tissues *in vivo* would normally be assumed to reach the germ cells and to be a germ cell mutagen, and therefore potentially hazardous to future generations. In the contrary situation, a substance that is negative in tests in somatic tissues *in vivo* would be assumed to be negative in germ cells, and moreover no germ cell-specific mutagen is known.

Normally, if the results of appropriate and adequately conducted *in vivo* tests are negative, then it can be concluded that the substance is not an *in vivo* genotoxin. If the results of the *in vivo* test(s) are positive, then it can be concluded that the substance is an *in vivo* genotoxin.

The Scientific Committee recommends a documented weight-of-evidence approach to the evaluation and interpretation of genotoxicity data. Such an approach should not only consider the quality and reliability of the data on genotoxicity itself, but also take into account other relevant data that may be available, such as physico-chemical characteristics, structure-activity relationships (including structural alerts for genotoxicity and 'read-across' from structurally related substances), bioavailability, toxicokinetics and metabolism, and the outcomes of any repeated-dose toxicity and carcinogenicity studies.

The Scientific Committee recognises that in the future EFSA will continue to receive datasets that differ from the testing strategy recommended in this opinion. Such datasets should be considered on a

case-by-case basis. Provided that the three critical endpoints (i.e. gene mutation, structural and numerical chromosomal alterations) have been adequately investigated, such datasets may be considered acceptable. The Scientific Committee recognises that in other cases where there is an heterogeneous dataset, EFSA has to rely on a weight-of-evidence approach.

KEY WORDS

Genotoxicity, testing strategies

149 **TABLE OF CONTENTS**

BACKGROUND

During the earlier work of the Scientific Committee on the welfare of experimental animals in 2007, a report was compiled entitled "Overview of the test requirements in the area of food and feed safety". It summarised the testing requirements adopted by the various EFSA Panels that undertake evaluations for chemical authorisation requests. From that overview, it was apparent that, although there are some similarities in the requirements for genotoxicity testing, they do differ between the various Panels and the types of substance being evaluated. There are differences both in the recommended basic battery of tests and in recommendations for any necessary follow-up tests. It was also noted that existing EFSA guidance on strategies for follow-up of *in vitro* positive or equivocal results is often very general.

Optimising strategies for genotoxicity testing, both with respect to a basic battery and follow-up tests, is an area where there is currently considerable activity worldwide. This probably reflects the fact that the science has progressed considerably in recent times. Research and developments in testing in this area are driven not only by the need to ensure that genotoxic substances can be detected in a basic battery of (usually *in vitro*) tests, but also by the need to ensure that such tests do not generate a high number of false positive results, because that has undesirable implications for animal welfare, e.g. by triggering unnecessary *in vivo* studies. Newer assays have also been advocated for use, such as the *in vitro* micronucleus test, the Comet assay, and tests using transgenic animals. For all these reasons guidance from regulatory bodies needs to be regularly reviewed and updated.
Thus, it would be appropriate and timely for the Scientific Committee to review the state-of-the-science in this area, given that genotoxicity testing and testing strategies are a cross-cutting issue for EFSA and its Panels.

It is recognised that it may not be desirable to completely harmonise genotoxicity testing requirements across EFSA Panels. Even if it were considered desirable, it might not be possible because some guidance (e.g. that for animal feed additives) has only recently been incorporated into legislation, while other guidance (e.g. that for plant protection products) is currently being revised at an EU Member State/Commission level. Some Panels are also currently preparing new or revised guidance on testing requirements.

TERMS OF REFERENCE

Following the suggestion of the Scientific Committee for a self-task on the topic of genotoxicity testing strategies, the European Food Safety Authority requests the Scientific Committee to:
Review the current state-of--the-science and provide a commentary and recommendations on genotoxicity testing strategies, which could contribute to greater harmonisation between EFSA Panels on approaches to such testing.
In its work the Scientific Committee is requested to take into consideration:

- that EFSA evaluates different types of substances with differing use/exposure scenarios,
- ongoing activities at national and international level on genotoxicity testing strategies (e.g. by the Japanese and European centres for the validation of alternative methods, the work of the International Working Group on Genotoxicity Testing (Müller et al., 2003; Tweats et al, 2007a,b; Thybaud et al., 2007a), various ILSI-HESI projects (Thybaud et al., 2007b) and collaborative work between ILSI-HESI and Health Canada (ILSI, 2008)).
- recent and foreseeable developments in genotoxicity that may have an impact on options for basic batteries of tests, including issues of reliability and validation of newer tests,
- optimisation of basic batteries of tests with a view to minimising false positive results,
- strategies for follow-up of indications of genotoxicity (positive findings) from a basic test battery that aim to establish whether genotoxic effects are likely *in vivo*, including not only *in vivo* testing but also approaches that make best use of available data (e.g. information on structural alerts, DNA-binding, metabolism, read-across from structurally related substances and mode of action).

ASSESSMENT

1. Introduction

Information on genotoxicity is a key component in risk assessment of chemicals in general, including those used in food and feed, consumer products, human and veterinary medicines, and industry. Genotoxicity testing of substances used or proposed for use in food and feed has been routine for many years. Genotoxicity information is also frequently essential for risk assessment of natural and environmental contaminants in food and feed. Many regulatory agencies and advisory bodies have made recommendations on strategies for genotoxicity testing (see, for example, review by Cimino, 2006). While the strategies for different chemical sectors may differ in points of detail, the majority recommend use of a basic test battery, comprising two or more *in vitro* tests, or *in vitro* tests plus an *in vivo* test, to evaluate genotoxic potential. This is followed up when necessary, in cases where the results of basic testing indicate that a substance is genotoxic *in vitro*, by further studies to assess whether the genotoxic potential is expressed *in vivo*. Follow-up usually comprises one or more *in vivo* tests.

Research in the area of genotoxicity has been prolific, both at the fundamental level and also with respect to comparative analysis of the performance and predictivity of individual tests and combinations of tests for risk assessment. There is an ongoing debate on the need to modify earlier recommended *in vitro* testing batteries (some of which can generate a high number of misleading (false) positives[4]) in order to avoid false positives and the triggering of unnecessary testing in animals, whilst at the same time ensuring detection of genotoxic potential that may have human health implications. Optimisation of testing batteries to minimise false positives may reduce the likelihood of detecting inherent genotoxic activity. Thus in recommending strategies for genotoxicity testing for risk assessment purposes, a balance needs to be struck that ensures with reasonable certainty that genotoxic substances that are likely to be active *in vivo* are detected. New tests have also been developed and their potential for inclusion in genotoxicity testing strategies, both in basic testing and in follow-up of positive results from basic testing, needs to be considered (see for example, Lynch et al., 2011).

In reviewing the state-of-the-science on genotoxicity testing, the Scientific Committee has taken note of other national and international initiatives. In particular, the Scientific Committee has considered not only the extensive literature on genotoxicity testing strategies but also proposals and recommendations from bodies such as the World Health Organization/International Programme on Chemical Safety (WHO/IPCS) (Eastmond et al., 2009), the European Centre for the Validation of Alternative Methods (ECVAM) (Kirkland et al., 2007a, Pfuhler et al., 2009), the International Workshop on Genotoxicity Testing (IWGT) (Kirkland et al., 2007b; Kasper et al., 2007; Burlinson et al., 2007; Tweats et al., 2007a,b; Thybaud et al, 2010), the European Cosmetics Association (COLIPA) (Pfuhler et al., 2010), the Health and Environmental Sciences Institute of the International Life Sciences Institute (ILSI-HESI) (Thybaud et al., 2007a, b; Dearfield et al., 2011), and the guidance documents developed for REACH (ECHA, 2008a,b). Further information is given on these initiatives in Appendix D.

In reaching its recommendations, the Scientific Committee was mindful that the various EFSA Panels consider different types of substances under their respective remits, with differing exposure conditions and varying test requirements. Test requirements may differ not only with respect to the range of toxicity tests recommended or required, but also with respect to the specific tests recommended for genotoxicity testing. In some cases, testing requirements are not set by EFSA, though EFSA may be consulted for its views (e.g. pesticides, for which testing requirements are agreed by the European Commission and the Member States and incorporated into European Union (EU) legislation). In some cases, current testing requirements are set by EFSA and incorporated into EU legislation (e.g. feed additives, GMOs), while in other cases, testing recommendations are made by EFSA and published in

[4] More details on this are given in chapter 3.3.

EFSA guidance documents but not (as yet) incorporated into EU legislation (e.g. food additives, food contact materials, flavouring substances, enzymes). Both guidance and legal testing requirements are updated from time to time in the light of new science and this opens up opportunities for harmonisation, where appropriate. Against this background, the recommendations set out in this opinion are intended to contribute to closer harmonisation of genotoxicity testing for risk assessment across EFSA's Scientific Panels.

2. Aims and rationale of genotoxicity testing

2.1. Potential health effects of genotoxic substances (both cancer and other diseases)

Genetic alterations in somatic and germ cells are associated with serious health effects, which in principle may occur even at low exposure levels. Mutations in somatic cells may cause cancer if mutations occur in proto-oncogenes, tumour suppressor genes and/or DNA damage response genes, and are responsible for a variety of genetic diseases (Erickson, 2010). Accumulation of DNA damage in somatic cells has also been proposed to play a role in degenerative conditions such as accelerated aging, immune dysfunction, cardiovascular and neurodegenerative diseases (Hoeijmakers, 2009; Slatter and Gennery, 2010; De Flora & Izzotti, 2007; Frank, 2010). Mutations in germ cells can lead to spontaneous abortions, infertility or heritable damage to the offspring and possibly to the subsequent generations.

2.2. Scope of genotoxicity testing

In view of the adverse consequences of genetic damage to human health, the assessment of mutagenic potential is a basic component of chemical risk assessment. To this aim, both the results of studies on mutation induction ("mutagenicity") and tests conducted to investigate other effects on genetic material are taken into consideration. Both the terms "mutagenicity" and "genotoxicity" are used in this opinion. Definitions of these terms given below are taken from the REACH "Guidance on information requirements and chemical safety assessment" (ECHA, 2008b).

> ***"Mutagenicity** refers to the induction of permanent transmissible changes in the amount or structure of the genetic material of cells or organisms. These changes may involve a single gene or gene segment, a block of genes or chromosomes. The term clastogenicity is used for agents giving rise to structural chromosome aberrations. A clastogen can cause breaks in chromosomes that result in the loss or rearrangements of chromosome segments. Aneugenicity (aneuploidy induction) refers to the effects of agents that give rise to a change (gain or loss) in chromosome number in cells. An aneugen can cause loss or gain of chromosomes resulting in cells that have not an exact multiple of the haploid number. For example, three number 21 chromosomes or trisomy 21 (characteristic of Down syndrome) is a form of aneuploidy.*
>
> ***Genotoxicity** is a broader term and refers to processes which alter the structure, information content or segregation of DNA and are not necessarily associated with mutagenicity. Thus, tests for genotoxicity include tests which provide an indication of induced damage to DNA (but not direct evidence of mutation) via effects such as unscheduled DNA synthesis (UDS), sister chromatid exchange (SCE), DNA strandbreaks, DNA adduct formation or mitotic recombination, as well as tests for mutagenicity."*
>
> The tests mentioned in the definition of "Genotoxicity" above that do not detect mutagenicity but rather primary DNA damage are commonly termed "indicator" tests. DNA adduct formation, for example, occurs when a substance binds covalently to DNA, initiating DNA repair, which can either return the DNA to its original state or, in the case of mis-repair, result in a mutation.

Genotoxicity testing is performed with the following aims:

- to identify substances which could cause heritable damage in humans,

- to predict potential genotoxic carcinogens in cases where carcinogenicity data are not available, and
- to contribute to understanding of the mechanism of action of chemical carcinogens.

For an adequate evaluation of the genotoxic potential of a chemical substance, different end-points (i.e. induction of gene mutations, structural and numerical chromosomal alterations) have to be assessed, as each of these events has been implicated in carcinogenesis and heritable diseases. An adequate coverage of all the above-mentioned end-points can only be obtained by the use of multiple test systems (i.e. a test battery), as no individual test can simultaneously provide information on all end-points. All the above mentioned endpoints should be examined in hazard identification irrespective of the expected level of human exposure (see Section 4.2.). A battery of *in vitro* tests is generally required to identify genotoxic substances. *In vivo* tests may be used to complement *in vitro* assays in specific cases, e.g. when the available information points to the involvement of complex metabolic activation pathways, which are expected not to be replicated by *in vitro* exogenous metabolic activation systems, or in case of high or "moderate and sustained" human exposure (Eastmond et al., 2009).

Further *in vivo* testing may be required to assess whether the genotoxic effect observed *in vitro* is also expressed *in vivo*. The choice of *in vivo* follow-up tests should be guided by effects observed in the *in vitro* studies (genetic endpoint) as well as by knowledge of bioavailability, reactivity, metabolism and target organ specificity of the substance. Clear evidence of genotoxicity in somatic cells *in vivo* has to be considered an adverse effect *per se*, even if the results of cancer bioassays are negative, since genotoxicity is also implicated in diseases other than cancer. A germ cell mutagen is expected to be also a somatic cell mutagen, while a substance that is a mutagen in somatic cells, provided it has the ability to reach the gonads, should also be considered a potential germ-line mutagen.

3. Review of key issues in genotoxicity testing

3.1. Operational performance of individual assays

3.1.1. General considerations

The methods most frequently used for the assessment of genotoxic potential *in vitro* and *in vivo* are described below. This list is not meant to be comprehensive of all existing methods, but more a consideration of the strengths, limitations and opportunity for further developments of the most widely used genotoxicity assays. Positive results of an *in vitro/in vivo* test indicate that the tested substance is genotoxic under the conditions of the assay performed; negative results indicate that the test substance is not genotoxic under the conditions of the assay performed.

More information about sensitivity and specificity of the different assays can be found in section 3.3. For a complete list of available *in vitro* and *in vivo* test methods see Dearfield et al., 2011.

3.1.2. Most commonly used *in vitro* methods

The most commonly used methods for assessing the genotoxic potential of substances are listed below, together with the relevant OECD Test Guideline (TG) on the basis of their principal genetic end-point:

Studies to investigate gene (point) mutation:

→ Bacterial reverse mutation assay in *Salmonella typhimurium* and *Escherichia coli* (OECD TG 471)

→ *In vitro* gene mutation assay in mammalian cells (OECD TG 476)

Studies to investigate chromosome aberrations:

→ *In vitro* chromosomal aberration assay (OECD 473)

→ *In vitro* micronucleus assay (OECD TG 487)

All the above mentioned *in vitro* tests should be conducted with and without an appropriate metabolic activation system. The most commonly used system is a cofactor-supplemented S9 fraction prepared from the livers of rodents (usually rat) treated with enzyme-inducing agents such as Aroclor 1254 or combination of phenobarbital and β-naphthoflavone. The choice and concentration of a metabolic activation system may depend upon the class of substance being tested. In some cases it may be appropriate to utilise more than one concentration of S9 mix. For azo dyes and diazo compounds, using a reductive metabolic activation system may be more appropriate (Matsushima, 1980; Prival et al., 1984).

Bacterial reverse mutation test (OECD TG 471 – also named Ames test)

The bacterial reverse mutation test is the most widely used assay to detect gene mutations. The test uses amino-acid requiring strains of *Salmonella typhimurium* and *Escherichia coli* to detect mutations, which involve substitution, addition or deletion of one or a few DNA base pairs. It has the ability to differentiate between frame-shift and base-pair substitutions with the use of different bacterial strains.

The principle of this test is that it detects mutations which revert mutations originally present in the test strains and which restore the functional capability of the bacteria to synthesise an essential amino acid. The revertant bacteria are detected by their ability to grow in the absence of the amino acid required by the parent test strain.

The bacterial reverse mutation test is rapid, inexpensive and relatively easy to perform. The limitation is that it uses prokaryotic cells which differ from mammalian cells in factors such as uptake, metabolism, chromosome structure and DNA repair processes. There have been developments to use it in high throughput screening (Claxton et al., 2001; Flückiger-Isler et al., 2004) but the methods have not been developed to a point where they can be routinely used.

Substances which do not directly interact with DNA will not be picked up as mutagenic by this test system. This may be relevant for example for carcinogenic metal compounds, which have been shown to decrease genomic stability by indirect mechanisms, for example by disturbance of the cellular responses to DNA damage, such as DNA repair systems, cell cycle control and apoptosis. Also, standard test procedures may have to be modified if substances are not taken up readily and longer incubation times may be required to ensure the intracellular bioavailability of the test substance, as may be the case for water insoluble metal compounds. Another example is the testing of nanomaterials, which require careful characterisation of the respective material, not only as added but also in cell culture medium, and may require modification of standard protocols.

***In vitro* mammalian cell gene mutation test (OECD TG 476)**

The *in vitro* mammalian cell gene mutation test can detect gene mutations, including base pair substitutions and frame-shift mutations. Suitable cell lines include L5178Y mouse lymphoma cells, the CHO, CHO-AS52 and V79 lines of Chinese hamster cells, and *TK6* human lymphoblastoid cells. In these cell lines the most commonly-used genetic endpoints measure mutation at thymidine kinase (*tk*) and hypoxanthine-guanine phosphoribosyl transferase (*hprt*) loci, and a transgene of xanthine-guanine phosphoribosyl transferase (*xprt*). The *tk*, *hprt*, and *xprt* mutation tests detect different spectra of genetic events. The autosomal location of *tk* and *xprt* may allow the detection of genetic events (e.g. large deletions) not detected at the hemizygous *hprt* locus on X-chromosomes.

Preference is often given to the L5178Y mouse lymphoma assay (MLA *tk* $^{+/-}$). This test can detect not only gene mutations, but also other genetic events leading to the inactivation or loss of heterozygosity (LOH) of the thymidine-kinase gene, such as large deletions or mitotic recombination. While the standard protocol allows discrimination between gross DNA alterations and point mutations on the

474 basis of colony size, the use of additional analytical methods can give information about the specific
475 event that has occurred (Ogawa et al., 2009; Wang et al., 2009).
476
477 The evaluation and interpretation of results from the L5178Y mouse lymphoma assay has changed
478 over the years and protocol updates have been recently recommended (Moore et al., 2007).
479 Cytotoxicity needs to be controlled to avoid false positive results, as with other *in vitro* genotoxicity
480 tests conducted in mammalian cells.
481
482 ***In vitro* mammalian micronucleus test (OECD TG 487)**
483
484 The purpose of the *in vitro* micronucleus test (MNvit) is to identify substances that cause structural
485 and numerical chromosomal damage in cells that have undergone cell division during or after the
486 exposure to the test substance. The assay detects micronuclei[5] in the cytoplasm of interphase cells and
487 typically employs human or rodent cells lines or primary cell cultures.
488
489 The *in vitro* micronucleus test can be conducted in the presence or in the absence of cytochalasin B
490 (cytoB), which is used to block cell division and generate binucleate cells. The advantage of the using
491 cytoB is that it allows clear identification that treated and control cells have divided *in vitro* and
492 provide a simple assessment of cell proliferation. The *in vitro* micronucleus test can be combined with
493 FISH (Fluorescence *in situ* Hybridisation) to provide additional mechanistic information, e.g. on non-
494 disjunction, which is not detected in the standard *in vitro* micronucleus assay.
495
496 The MNvit is rapid and easy to conduct and it is the only *in vitro* test that can efficiently detect both
497 clastogens and aneugens. Cytotoxicity needs to be controlled to avoid false positive results, as with
498 other *in vitro* genotoxicity tests conducted in mammalian cells.
499
500 ***In vitro* mammalian chromosomal aberration test (OECD TG 473)**
501
502 The *in vitro* chromosomal aberration (CA) test detects structural aberrations and may give an
503 indication for numerical chromosome aberrations (polyploidy) in cultured mammalian cells caused by
504 the test substance. However, this test is optimised for the detection of structural aberrations.
505
506 The *in vitro* chromosomal aberration test may employ cultures of established cell lines or primary cell
507 cultures. Cells in metaphase are analysed for the presence of chromosomal aberrations. Additional
508 mechanistic information can be provided using FISH or chromosome painting.
509
510 The test has been widely used for many decades but it is resource intensive, time consuming and it
511 requires good expertise for scoring. Only a limited number of metaphases are analysed for each assay.
512 Cytotoxicity needs to be controlled to avoid false positive results, as with other *in vitro* genotoxicity
513 tests conducted in mammalian cells.
514

515 ### 3.1.3. Most commonly used *in vivo* methods

516 The most commonly used methods to assess the genotoxic potential of substances *in vivo* are listed
517 below, on the basis of their principal genetic end-point:

518 *Studies to investigate gene mutations:*

519 → Gene mutation assays in transgenic models (draft OECD TG)

520 *Studies to investigate chromosome damage:*

521 → Mammalian erythrocyte micronucleus test (OECD TG 474)

[5] Micronuclei in the cytoplasm of interphase cells may originate from acentric chromosome fragments (i.e lacking a centromere) or whole chromosomes that are unable to migrate to the poles during the anaphase stage of cell division

→ Mammalian bone marrow chromosomal aberration test (OECD TG 475)

Studies to investigate primary DNA damage:

→ COMET assay (no OECD TG as yet, internationally agreed protocols available)
→ Mammalian unscheduled DNA synthesis (UDS) assay *in vivo* (OECD TG 486)

In vivo transgenic rodent (TGR) gene mutation assay

The transgenic rodent mutation assay (TGR) is based on transgenic rats and mice that contain multiple copies of chromosomally integrated phage or plasmid shuttle vectors that harbour reporter genes for detection of mutation and/or chromosomal rearrangements (plasmid model and Spi$^-$ assay) induced *in vivo* by test substances (OECD, 2009; OECD, 2010b; Lambert et al., 2008). TGR mutation assays measure mutations induced in genetically *neutral marker genes* (i.e. genes that have no immediate consequence to the animal) recovered from virtually any tissue of the rodent. These neutral transgenes are transmitted by the germ cells, and thus can be detected in all cells including the germ cells. Mutations arising in a rodent are scored by recovering the transgene and analysing the phenotype of the reporter gene in a bacterial host deficient for the reporter gene.

The transgenic mice models respond to mutagens in a similar manner to endogenous genes and are suitable for the detection of point mutations, insertions and small deletions but not large deletions because the *cos*-sites, at the end of the vector, together with a restrictive length of the vector, are essential (for excision and packaging into phage heads). The Spi$^-$ assay and the plasmid model can detect large deletions and thus are able to detect chromosomal rearrangements. The transgenic rodent models could also be used in repeated-dose toxicity studies as the transgenes are neutral genes.

The International Workshop on Genotoxicity Testing (IWGT) has endorsed the inclusion of TGR gene mutation assays for *in vivo* detection of gene mutations, and has recommended a protocol for their implementation (Heddle et al., 2000; Thybaud et al., 2003). An OECD test guideline based on these recommendations has been drafted (OECD, 2010b) and will soon be adopted.

In vivo mammalian erythrocyte micronucleus test (OECD TG 474)

The purpose of the *in vivo* mammalian erythrocyte micronucleus test (MNviv) is to identify substances that cause structural and numerical chromosomal damage in somatic cells *in vivo*. The damage results in the formation of micronuclei, containing chromosome fragments or whole chromosomes in young (polychromatic) erythrocytes sampled in bone marrow and/or reticulocytes of peripheral blood cells of animals, usually rodents. It might not detect organ-specific compounds and unstable compounds or metabolites. If there is evidence that the test substance or the reactive metabolite will not reach the target tissue, it would not be appropriate to use this test.

This assay has a long history of use and it is also potentially applicable in tissues other than the bone marrow or the peripheral blood. The MNviv is still the most widely used *in vivo* genotoxicity test that detects both clastogens and aneugens. High throughput approaches to the peripheral blood have been published (Torous et al., 2000; De Boeck et al., 2005). Possible confounding effects like hypo- and hyperthermia may affect the formation of micronuclei and therefore the scoring. The MNviv can be combined with FISH to provide additional mechanistic information.

In vivo mammalian bone marrow chromosomal aberration test (OECD TG 475)

The mammalian *in vivo* chromosomal aberration test is used for the detection of structural chromosomal aberrations induced by test substances in bone marrow cells of animals, usually rodents.

Bone marrow is the target tissue of this test, therefore if there is evidence that the test substance or the reactive metabolite does not reach the bone marrow, it would not be appropriate to use this test.

As with the *in vitro* chromosomal aberration test, it requires experienced scientists for the scoring of metaphases. It might not detect organ-specific compounds and unstable compounds or metabolites. This assay is potentially applicable also to tissues other than the bone marrow.

***In vivo* Comet assay**

The *in vivo* Comet assay detects DNA single and double strand breaks, alkali-labile lesions, as well as DNA strand breaks arising during the repair of DNA lesions. No OECD Test Guideline yet exists for this assay but internationally agreed protocols are available for performing this test (hptt://cometassay.com).

The *in vivo* Comet assay has the advantage of being rapid and easy to conduct and may be applied to any tissues that can be subcultured. Cell division is not required and a low number of cells is sufficient for the analysis. It is considered an indicator test detecting pre-mutagenic lesions and can be used for mechanistic studies.

The *in vivo* Comet assay has been suggested by several authors (Tice et al., 2000; Hartmann et al., 2003; Burlinson et al., 2007) as a suitable follow-up test to investigate the relevance of positive *in vitro* tests (gene mutagens and clastogens, but not aneugens).

***In vivo* mammalian unscheduled DNA synthesis (UDS) test (OECD TG 486)**

The *in vivo* UDS test allows the investigation of genotoxic effects of substances in the liver. The endpoint measured is indicative of DNA adducts removal by nucleotide excision repair in liver cells and it is measured by determining the uptake of labelled nucleosides in cells that are not undergoing scheduled (S-phase) DNA synthesis.

It has to be considered as an indicator test for DNA damage and not a surrogate test for gene mutations *per se*. The UDS assay has a long history of use but it is useful only for some classes of substances. Tissues other than the liver may in theory be used. However, UDS has a limited use for cells other than liver and its sensitivity has been questioned (Kirkland and Speit, 2008). It is resource intensive and the scoring time consuming. Moreover, radio-labelled substances are required when performing this test.

3.2. Guidance or requirements of EFSA Panels for genotoxicity testing with different types of substances

In general, guidance for genotoxicity testing given by different EFSA Panels has been established at different times, in some cases dating back several years and therefore reflecting, at least in part, differences in the state of the discussion at those time points. It should also be recognised that different types of substances are evaluated within EFSA's remit and that some guidance documents have been incorporated into EU legal requirements for the group of substances under evaluation. More detailed information on the Panels' guidance and requirements can be found in Appendix A.

With respect to ***in vitro* testing**, substances assessed by all Panels (with the exception of enzymes by FEEDAP and CEF Panels, see Appendix A) currently require the assessment of gene mutations in bacteria, gene mutations in mammalian cells as well as chromosome aberration in mammalian cells. Main differences are related to the number of *in vitro* tests required in the core battery to cover these effects and the core test battery itself, for example with respect to the inclusion of the MNvit as an alternative to the chromosomal aberration test. Other differences are related to the follow-up of positive *in vitro* results.

With respect to ***in vivo* testing,** only the guidance document of the FEEDAP Panel (EFSA, 2008) includes an *in vivo* test in a mammalian species in its basic test battery, independent of the outcome of the *in vitro* tests.
The current legislation on plant protection products (EC Directive 91/414) also requires one *in vivo* test as follow-up of *in vitro* results. A new Regulation [(EC) 1107/2009 of the European Parliament and of the Council of 21 October 2009] will come in force on 14 June 2011. However, revised Annexes II and III, including the data requirements, are not published as yet. Prior to agreement on this new Regulation, the PPR Panel was requested by the Commission to issue an opinion on the data requirements for Annex II and III. The Panel suggested in its opinion (EFSA, 2007) that, for genotoxicity, there was no need for follow-up *in vivo* testing after negative *in vitro* results. It is not yet known whether the Panel recommendation will be taken up in the revised annexes expected to be published by the end of 2011.

All other Panels require the *in vivo* follow-up of positive *in vitro* results, mostly following a flexible approach depending on the results from the *in vitro* studies. Four Panels include the *in vivo* transgenic mouse system as one option for *in vivo* testing and one Panel also includes the *in vivo* Comet assay as an option. *In vivo* germ cell testing may be required on a case-by-case basis. In general, the Panels recommend that current OECD guidelines or international accepted recommendations (see 3.1.2) for the respective tests should be followed, but additional tests without adopted guidelines may be acceptable for further clarification.

3.3. Analysis of sensitivity and specificity of *in vitro* and *in vivo* tests with respect to prediction of rodent carcinogenesis

Cancer is a disease of somatic cells which is strongly linked to the occurrence of mutations. Consequently, the performance of genotoxicity tests can be assessed by evaluating their predictivity for cancer. It does not, however, mean that these tests show the same performance for other (genetic) diseases. The evaluation of the performance of genotoxicity tests in relation to their predictivity for carcinogenicity depends strongly on the databases used. The quality of the tests and the conclusions drawn from the tests contribute to the reliability of the predictions. The total number of substances in the database and particularly the number of non-carcinogens is important as well. Most databases have the limitation of a very low number of non-carcinogens. The databases used for the results discussed below contain both genotoxic and non-genotoxic carcinogens and do not distinguish between rodent carcinogens and human carcinogens, thus limiting the predictivity of genotoxicity tests for human cancer risk.

Table 1 describes the terms used. In addition to the definitions in Table 1, sensitivity and negative predictivity also give an indication of the number of "false negative" results (negative results in genotoxicity tests obtained with carcinogens); specificity and positive predictivity also give an indication of "false positive" results (positive results in genotoxicity tests obtained with non-carcinogens). In fact, false positive and false negative results are incorrect classifications. Such results are not false, but are correct results in that specific test. False negative results are better described as "unexpected" or "misleading negative" results obtained with carcinogens and likewise false positive results as "unexpected" or "misleading positive" results with non-carcinogens. However, since in the scientific literature the term "false" is generally used, for convenience in the present report "false" is also used instead of "unexpected" or "misleading".

Table 1: Terms used to describe the performance of the genotoxicity tests

	Carcinogens	**Non-carcinogens**
Genotoxicity positive	A	B
Genotoxicity negative	C	D

675

Sensitivity	% correct identified carcinogens	A/(A+C) * 100
Specificity	% correct identified non-carcinogens	D/(B+D) * 100
Concordance	% correctly identified carcinogens and non-carcinogens	(A + D)/(A+B+C+D) * 100
Positive predictivity	% correctly predicted carcinogens among positive results	A/(A+B) *100
Negative predictivity	% correctly predicted non-carcinogens among negative results	D/(C+D) * 100

676

677 The number of false negative results may be an over-estimation. Cancer can be triggered by genotoxic
678 or non-genotoxic mechanisms. Carcinogens with a non-genotoxic mode of action may score negative
679 and are then easily considered false negatives in genotoxicity tests whereas in fact they are 'correct'
680 negatives in the specific tests. Secondly, for genotoxicity, three genotoxic endpoints (gene mutations,
681 structural and numerical chromosome aberrations) exist. A negative result in a specific genotoxicity
682 test can be the result from a test that does not cover the genotoxic endpoint which makes the substance
683 tested a carcinogen. For instance, a carcinogen which predominantly induces chromosome aberrations
684 will generally score positive in a chromosome aberration test but may (correctly) be negative in gene
685 mutation tests. Kirkland *et al.* (2005), in a review of the substances in their database, showed that the
686 mechanism of action for carcinogenicity of 80% of the false negative substances is known to be non-
687 genotoxic.
688
689 The number of false positives (specificity) is a bigger problem because this may trigger unnecessary *in*
690 *vivo* tests using or could even lead to the abandonment of further development of promising
691 substances.

692 **3.3.1. *In vitro* genotoxicity tests**

693 Many papers have been published on the performance of *in vitro* tests but two of them are
694 particularly relevant (Kirkland et al., 2005; Matthews et al., 2006).These papers examined the most
695 popular *in vitro* genotoxicity tests for their ability to discriminate between carcinogens and non-
696 carcinogens. Many genotoxicity test results on these substances were re-evaluated by experts because
697 the interpretation of data has changed over time. Table 2 shows the performance of the individual *in*
698 *vitro* tests. The concordance (between 60 and 70 %) is similar in the 5 tests evaluated. On the other
699 hand, although the sensitivity of the Ames test and the *hprt* test is the lowest, the specificity in these
700 tests is higher than in other tests. The information on the specificity of the *in vitro* micronucleus test is
701 limited by the small number of tests performed by 2005.
702
703 **Table 2: Performance of the most common short-term *in vitro* genotoxicity tests in detecting**
704 **rodent carcinogens** (data from Kirkland et al., 2005 and Matthews et al., 2006)
705

	Ames[1]	**Ames[2]**	**MLA[1]**	**MLA[2]**	***hprt*[2]**	**CA[1]**	**CA[2]**	**MNvit [1]**	**MNvit [2]**
No. of substances	**717**	**988**	**350**	**460**	**237**	**488**	**673**	**115**	**97**
Sensitivity %	58.8	49.4	73.1	62.8	59.3	65.6	55.3	78.7	87.3
Specificity %	73.9	80.3	39.0	57.8	72.9	44.9	63.3	30.8	23.1
Concordance %	62.5	62.9	62.9	60.7	63.3	59.8	58.7	67.8	70.1
Positive predictivity %	87.4	76.4	73.7	66.1	83.8	75.5	67.1	79.5	75.6
Negative predictivity %	36.8	55.1	38.3	54.2	42.9	33.5	51.1	29.6	76.9

706
707
708

Figure Legend:
1: Kirkland et al., 2005; **2**: Matthews et al., 2006
Ames: Ames test (gene mutation test in bacteria); **MLA**: mouse lymphoma assay (gene mutation test in mammalian cells); ***hprt***: hprt test (gene mutation test in mammalian cells); **CA**: chromosome aberration test; **MNvit**: micronucleus test *in vitro*.

In a workshop organised by ECVAM (DG JRC - Ispra, Italy) the rate of false positive results in genotoxicity tests was addressed (Kirkland et al., 2007a). During the workshop it was investigated (i) whether it is possible to choose existing cell systems which give lower rates of false results, (ii) whether modifications of existing guidelines or cell systems may result in lower false (positive) results, and (iii) what was the performance of new systems showing promise of improved specificity. It was concluded that there was a need for better guidance on the likely mechanisms (high cytotoxicity, high passage number of cell lines, *p53* status, DNA repair status, etc) resulting in positive results not relevant for humans and on how to obtain evidence for those mechanisms. Collaborative research programs have been started to improve the existing genotoxicity tests and to identify and evaluate (new) cell systems with appropriate sensitivity but improved specificity.

3.3.2. Combinations of *in vitro* genotoxicity tests

Since three genotoxic endpoints, i.e. gene mutations, structural and numerical chromosomal aberrations, have to be investigated, it is more meaningful to evaluate the performance of combinations of tests covering these endpoints. The bacterial reverse mutation test is always accepted as part of every strategy because of its specificity for detection of genotoxic carcinogens and is usually the first test to be performed. Most strategies then consist of two further tests performed in mammalian cells: a gene mutation test in mammalian cells and a test measuring chromosomal damage.

Each individual test may result in false negatives and/or false positives. In a combination of tests, the number of false negatives will decrease because a single positive result is considered as evidence that the substance is positive. On the other hand, a substance is only considered negative if all tests performed are assessed negative. The number of false positives, consequently, will increase in combinations of tests.

An evaluation by Kirkland et al. (2005) on combinations of two or three assays (Table 3) showed that in combinations the sensitivity increases whereas the specificity decreases. A combination of three tests, including the mouse lymphoma assay which measures gene mutations and chromosome aberrations in mammalian cells, had a higher sensitivity but the specificity further decreased compared with two tests combination. It would appear that a strategy of three tests is not better than two tests although it is generally felt to be “safer”. In a recent analysis of an existing database of rodent carcinogens and a new database of *in vivo* genotoxins, together covering over 950 substances, Kirkland et al. (2011) confirmed that data from the gene mutation test in bacteria and the *in vitro* micronucleus test allowed the detection of all the relevant *in vivo* carcinogens and *in vivo* genotoxins for which data exist in these databases (Kirkland et al., 2011). Consequently, it would appear that the starting point should be a combination of two *in vitro* tests. Assuming the choice of the Ames test to identify gene mutations, as one of the tests, the only option for two tests which cover the three endpoints is a combination of the Ames test with the *in vitro* micronucleus test. The latter detects both structural and numerical chromosome aberrations. Although the sensitivity is good, combinations with the *in vitro* micronucleus test result in decreases in specificity, due to the low number of non-carcinogens on which this estimate is based. In a retrospective validation study, an expert panel (Corvi et al., 2008) concluded that the *in vitro* micronucleus test can be regarded as sufficiently validated and can be recommended as an alternative to the *in vitro* chromosomal aberration test. The OECD guideline for the *in vitro* micronucleus assay was adopted in July 2010 (OECD, 2010a).

Table 3: Performance of a battery of *in vitro* tests in detecting rodent carcinogens and non-carcinogens (data from Kirkland et al., 2005)

	Ames and MLA[1]	Ames and MN[1]	Ames and CA[1]	MLA and MN[1]	MLA and CA[1]	Ames and MLA and MN[2]	Ames and MLA and CA[2]
No. of substances	**347**	**110**	**480**	**74**	**299**	**74**	**298**
Sensitivity %	81.0	85.9	75.3	87.0	81.3	90.7	84.7
Specificity %	32.4	12.0	34.6	10.0	27.1	5.0	22.9
Concordance %	66.3	69.1	63.8	66.2	63.9	67.6	64.8
Positive predictivity %	73.4	76.8	74.4	72.3	70.2	72.1	69.8
Negative predictivity %	42.5	20.0	35.6	22.2	40.6	16.7	41.5

1: if at least one test out of the two tests performed is positive; **2**: if at least one out the three tests performed is positive; **Ames**: Ames test (*in vitro* gene mutation assay in bacteria); **MLA**: mouse lymphoma assay; **MN**: *in vitro* micronucleus test; **CA:** *in vitro* chromosomal aberration test

3.3.3. *In vivo* genotoxicity tests

The major aim of *in vivo* genotoxicity tests is to investigate whether the positive results of *in vitro* genotoxicity tests can be confirmed *in vivo* and to identify and eliminate from concern the substances which are false positives in the *in vitro* tests. The *in vivo* follow-up test needs to be a logical choice, *i.e.* the test should cover the same genotoxic endpoint as the one which showed positive results *in vitro*. For instance, if a substance appeared as a clastogen under *in vitro* conditions then further testing should be carried out with an *in vivo* test for clastogenicity.

The classical *in vivo* tests may be limited to certain tissue restrictions (bone marrow, peripheral blood cells, hepatocytes). Considering that *in vivo* testing is often a pre-screen for cancer, it is obvious that the value of the *in vivo* tests increases if the target tissue(s) for carcinogenicity are investigated. Therefore, tests without obvious tissue restriction should be recommended as follow-up tests, where possible.

Similar extensive evaluations on the performance of *in vivo* tests are not available as they are for *in vitro* tests. The evaluations on *in vivo* tests are limited by a rather low number of tests and an imbalance in the ratio between the number of (genotoxic and non-genotoxic) carcinogens and non-carcinogens. The database used by Lambert et al. (2005) was built to promote the *in vivo* gene mutation test with transgenic mice and therefore is biased towards substances investigated in the transgenic mouse assay.

Table 4: Performance of the individual *in vivo* tests in detecting rodent carcinogens and non-carcinogens

	CA[1]	MN[1]	Comet[2]	TGR[3]
No. of substances	**82**	**82**	**190**	**105**
Sensitivity %	43.6	36.4	78.1	78
Specificity %	66.7	77.8	80.0	69
Concordance %	51.2	50.0	78.4	77
Positive predictivity %	72.7	76.9	95.4	95
Negative predictivity %	36.7	37.5	40.7	31

1: Kim and Margolin, 1999; **2:** Sasaki et al., 2000; **3**: Lambert et al., 2005
CA: *in vivo* chromosome aberration assay; **MN**: *in vivo* micronucleus test; **Comet**: Comet assay; **TGR** *in vivo* gene mutation assay in transgenic mice.

Table 4 shows the performance of the different *in vivo* assays in the prediction of carcinogenicity. The Comet assay and the gene mutation test with transgenic animals perform relatively well, which is demonstrated by the relatively high sensitivity and specificity. Strikingly, the sensitivities of the chromosomal aberration test and the micronucleus test are low. This is likely to be a consequence of low exposure of hematopoietic cells *in vivo*. Thus, it is necessary that evidence of target cell exposure is obtained in such studies. Since the Comet assay and the gene mutation assay with transgenic animals are tests without tissue restriction, and have a good sensitivity and specificity, these tests may be recommended as *in vivo* follow-up tests. However, as the number of non-carcinogens in the database is low, firm conclusions on the specificity and negative predictivity for these tests are not possible. In any case, it is noted that evidence of genotoxicity *in vivo* should be considered a relevant toxicological end-point *per se*, independently of the predictive value for carcinogenicity (see section 2).

3.3.4. *In vivo* follow-up tests when *in vitro* tests are positive

As mentioned earlier, the *in vivo* follow-up test needs to be a logical choice, i.e. the test should cover the same genotoxic endpoint as the one which showed positive results *in vitro*. Moreover, with the objective of reducing the use of experimental animals, normally only one *in vivo* test should be conducted. A second test is only then necessary if the first *in vivo* test is negative and does not cover all *in vitro* positive genotoxic endpoints. Traditionally, the *in vivo* micronucleus has been the most widely used *in vivo* test. However, this test suffers from a certain tissue-restriction and does not identify all (rodent) carcinogens. More recently, the use of the Comet assay and the *in vivo* gene mutation assay with transgenic mice has increased, mainly because they are able to detect genotoxic damage in (almost) every tissue. Kirkland and Speit (2008) demonstrated that both the Comet assay and the transgenic mouse assay had a high sensitivity to identify carcinogens acting via both clastogenic (Comet assay) and gene mutation (both assays) mechanisms (Table 5). They also reported that, when a positive result was found in these assays, such responses were seen in those tissues where the tumours occur; responses were, however, also found in non-tumour tissues. Kirkland and Speit (2008) suggested that these assays should be given a higher priority in selection of the follow-up *in vivo* test for genotoxic substances that are positive in *in vitro* tests.

Table 5: Influence of gene mutation or clastogen profile *in vitro* on *in vivo* results for carcinogens (Kirkland and Speit, 2008)

***In vitro* results**	**Number of carcinogens**								
	***In vivo* UDS result**			***In vivo* TGR assay result**			***In vivo* Comet assay result**		
	+	**-**	**E**	**+**	**-**	**E**	**+**	**-**	**E**
+ in Ames	7	13	5	13	5	1	21	1	0
+ in MLA	5	14	3	8	4	1	13	1	0
+ in MNvit	1	3	3	5	3	0	7	2	0
+ in CA	5	11	3	8	5	1	17	1	0

Ames: Ames test (*in vitro* gene mutation assay in bacteria); **MLA**: mouse lymphoma assay; **MNvit**: *in vitro* micronucleus test; **CA**: *in vitro* chromosome aberration test; **+**: positive result; **-**: negative result; **E**: equivocal result.

3.3.5. Analysis of genotoxicity data on substances used in food contact materials

An analysis of the correlation between *in vitro* and *in vivo* positives in genotoxicity tests has also been performed using data submitted to the former Scientific Committee on Food (SCF) or to EFSA for approval of chemically defined food contact materials (FCM). It shows that a large number of substances that test positive *in vitro* do not test positive *in vivo*. The results of this analysis are given and discussed in Appendix B.

3.4. Issues in reduction of false positive and false negative results

Certain characteristics of the cell lines commonly used in genotoxicity assays such as the *p53* status, karyotypic instability, DNA repair deficiencies and the need for exogenous metabolism are recognised as possibly contributing factors to the high rate of *in vitro* false positives (Kirkland et al., 2007a). The use of high concentrations of test substance and high levels of cytotoxicity are also considered to be potential sources of false positive results. Considerations on the impact of these factors are presented below.

3.4.1. The example of p53

On the basis of the key role played by the *p53* tumour suppressor gene in the response to DNA damage, a contribution of *p53* to the outcome of genotoxicity tests with mammalian cells is expected. Two major functions of *p53*, i.e. its role in cell death and mutation frequency and type, are expected to impact on the outcome.

The lack of *p53* leads in general to resistance to cytotoxic drugs and to increased spontaneous and induced mutation frequency. The type of *p53* inactivation (deletion versus viral inactivation or targeted mutation) and the class of chemical are key factors in the outcome. For instance, inactivation of *p53* by E6 transfection predominantly induces sensitisation to cytotoxic drugs whereas a knockout loss of function induces drug-resistance (Cimoli et al., 2004). Mutant *p53* may interfere with recombination, apoptosis and other cellular processes, thus leading to high levels of mutations resulting in loss of heterozygosity (LOH). If *p53* function has been abrogated, recombination-mediated mutations occur at a much lower frequency (Morris, 2002).

Fowler and co-workers (2011, manuscript submitted) tested the hypothesis that *p53* deficiency of commonly used rodent cell lines could affect the rate of false positive results in genotoxicity testing. A selection of substances that were accepted as producing false positive results in *in vitro* assays (Kirkland et al., 2008) was tested for micronucleus induction in a set of *p53*-defective rodent cells (V79, CHL, CHO). The results were then compared with those obtained with *p53*-competent human peripheral blood lymphocytes (HuLy), TK6 human lymphoblastoid cells and the human liver cell line, HepG2. The *p53*-defective rodent cell lines were consistently more sensitive to cytotoxicity and micronucleus induction than *p53*-competent cells. The authors concluded that a reduction in the frequency of false positive results can be achieved by using *p53*-competent cells. Although the data are suggestive of an effect of *p53*, it should be taken into account that in this study the *p53*-defective cells are all rodent cells whereas the *p53*-competent cells are of human origin and species-related confounding factors may affect the outcome. Moreover, the type of *p53* inactivation in the defective cell lines used in this study should be carefully considered for its potential effect on the DNA damage response. Further studies with a set of cell lines of the same origin and with well defined p53 mutations are required to address this issue.

Although it is useful to characterize the *p53* status (and possibly DNA repair profile) of the test cell system, it is questionable whether a cell line proficient in *p53* and DNA repair would be the ideal test system for genotoxicity assays because this would impact on the sensitivity of the assay. During *in vitro* immortalisation, cells undergo significant changes and the mutation of *p53* is one of the most frequent events (Lehman et al., 1993). These changes are unavoidable and their understanding is of great value for a sound interpretation of the results.

3.4.2. The metabolic competence of *in vitro* systems

The xenobiotic metabolising system comprises several hundred enzymes and factors such as animal species, tissue and cell type, expression level of activating/inactivating enzymes determine the relative importance of each bioactivation pathway. No cell type *in vivo* reflects the full biotransformation capacity of an organism and the expression of numerous enzymes ceases or is drastically reduced upon cell culturing. Detoxifying systems that assure the reduction of reactive intermediates *in vivo* are usually inefficient in *in vitro* systems. This premise is the basis for the use of exogenous metabolic

systems in genotoxicity assays. However, the almost universal use of a single metabolic activation system (i.e. Aroclor 1254-induced or phenobarbital/β-naphthoflavone-induced rat liver S9) for all *in vitro* genotoxicity assays has also considerable limitations (Kirkland et al., 2007a,c). Different carcinogens are activated by different cytochrome (CYP) and non-cytochrome (non-CYP) enzymes. Phase 2 enzymes are essentially inactive in standard S9 because of a lack of cofactors and this should not be underestimated as several promutagens are activated by phase 2, non-CYP enzymes (e.g. sulphotransferases). The induction by Aroclor-1254 leads to over-representation of the CYP1A and 2B compared to other CYP forms, thus producing a metabolic profile that differs from that of normal liver. Finally, a small portion of the active metabolite may reach the target when it is generated outside the cell environment.

The use of cell lines engineered to express various enzymes is very attractive because the generation of enzymes within the target cells presents an obvious advantage as opposed to external enzyme systems. However, since very specific enzymes are required, depending on the promutagen, a battery of engineered cell lines expressing panels of metabolic enzymes would be required. In addition the activity of the transgenes would need to be checked on a regular basis considering that epigenetic silencing and/or recombinational events might occur.

Alternatively, cell lines are available that maintain some metabolic competency (Kirkland et al., 2007a; Donato et al., 2008). For instance, the HepG2 or Hepa RG cell lines maintain the expression of some metabolic genes of primary human hepatocytes. However, important endpoints such as gene mutations are difficult to study in this cell system. Methods need to be developed in this direction.

Based on current knowledge, metabolic differences between *in vitro* test systems and that of animals used *in vivo* may affect false positive and false negative rates, but their relative contribution is not known. If genetically engineered cell lines are used, it should be mandatory to address the long-term stability of critical properties. The characterisation of the metabolic capability of cellular models used for genotoxicity testing remains a prerequisite for sound interpretation of the results obtained when using the corresponding tests.

3.4.3. Top dose concentration

Current OECD guidelines for *in vitro* genotoxicity testing in mammalian cells require that the top concentration with soluble and non-toxic substances should be 10 mM or 5000 µg/ml (whichever is lower), except where cytotoxicity or precipitation are limiting factors below this level. However, there has recently been considerable debate that testing at high concentrations could be a possible source of false positive results. The requirement of the top concentration of 10 mM or 5000 µg/ml (whichever is lower) was based on a small number of carcinogens that needed high concentrations before giving positive responses in mammalian cell tests *in vitro*, sometimes using inappropriate metabolic conditions. The published data on these chemicals are quite old, which may suggest that they could be detected at lower concentrations under current protocols. It also has to be considered that a simple coincidence of carcinogenicity findings in rodents and genotoxicity at high *in vitro* concentrations that is not relevant *in vivo,* does not mean there is a mechanistic correlation between the *in vitro* genotoxicity and the *in vivo* carcinogenicity. This issue has been addressed for pharmaceutical testing by the International Conference for Harmonisation of the Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and is currently under investigation for industrial chemicals.

3.4.3.1. Considerations from International Conference on Harmonisation (ICH)S2 revision process

In the proposed S2 guideline revision (ICH, 2010), the International Conference on Harmonisation (ICH) is considering that the highest concentration tested in mammalian cell assays should be reduced to 1 mM or 500 µg/ml (whichever is higher). An alternative of 500 µg/ml has been proposed because 1 mM would be too low for adequate assessment of low molecular weight substances. This suggestion to reduce the current upper limit can be justified based on the following considerations:

(1) A review of human exposure levels for pharmaceuticals (Goodman and Gilman, 2002) shows that pharmacologically active concentrations for drugs are typically below 10 µg/ml (or 20 µM for average molecular weight of 500). Although some drugs may have a higher plasma level and others may accumulate in tissues, there are no examples of a drug which exhibits both characteristics. Thus, a top concentration of 1mM would capture low potency drugs and other high dose drugs including cases of extensive tissue accumulation.

(2) The optimal substrate concentrations for many enzymes (K_m), including those for metabolic activation/inactivation, cellular transport/turnover or defence mechanisms are typically below 100 µM. Higher exposure beyond enzyme saturation can result in artefactual effects with no relevance for *in vivo* conditions.

(3) The original 10 mM limit was based on the intention to set an upper limit where none previously existed. It was defined as a limit low enough to avoid artefactual increases in chromosome damage/mutations due to excessive osmolality, and high enough to ensure detection of a number of *in vivo* clastogens. The latter was based on an examination of a data set to examine whether known *in vivo* clastogens would be detectable in the *in vitro* chromosome aberration assay when limiting the maximum concentration to 10 mM (Scott et al., 1991). This data set was re-examined by an ICH S2 Expert working group and it was noted that all *in vivo* positive chemicals were detected in the Ames test or *in vitro* in mammalian cell assays below 1 mM.

3.4.3.2. Subsequent analyses (including non-pharmaceuticals)

Testing to high concentrations and high levels of cytotoxicity is currently required in *in vitro* mammalian cell genotoxicity tests, not only for pharmaceuticals but also for industrial chemicals, and is likely to contribute to the high frequency of false positive results. This topic was discussed during an ECVAM (DG-JRC, Ispra, Italy) workshop on "How to reduce false positive results in *in vitro* mammalian cell genotoxicity tests", which recommended an evaluation of the top concentration in mammalian cell tests required to detect rodent carcinogens (Kirkland et al., 2007a). Moreover, from the 19 chemicals which were identified as giving false positive results (Kirkland et al., 2008), 12 were shown to be positive only when tested above 1 mM. Consequently, an analysis of existing data on *in vitro* mammalian cell tests has been conducted to assess the effect that a reduction of top concentration would have on the outcome of *in vitro* genotoxicity testing (Parry et al., 2010). This analysis included 384 chemicals classified as rodent carcinogens and reported the results of the Ames test as well as the test concentrations which produced positive results in the mouse lymphoma assay, the chromosomal aberration assay and the micronucleus test. In this analysis of published mammalian cell data, 24 rodent carcinogens that were negative in the standard Ames test have been indicated to require testing above 1mM in order to give a positive result in the *in vitro* mammalian cell tests.

A re-evaluation of these chemicals showed that some of them were known to be probable non-genotoxic (non-mutagenic) carcinogens, tumour promoters or negative for genotoxicity *in vivo,* and 9 were retested according to modern MLA and chromosomal aberration protocols (Kirkland and Fowler, 2010). For 5 of those chemicals, no genotoxic response was observed when they were tested according to current cytotoxicity limits, suggesting that they are not genotoxic either in bacteria or in mammalian cells *in vitro*. The other 4 chemicals were confirmed as genotoxic at concentrations below 1mM. Only methylolacrylamide required higher concentrations (2 mM) for detection of a positive response. However, this concentration corresponded to only 202 µg/ml because of its low molecular weight.
Based on this analysis and re-evaluation, it could be concluded that the 10 mM upper limit for non-toxic chemicals in mammalian cell tests is not justified, and can be reduced without loss of the ability to detect genotoxic rodent carcinogens. Thus, a new limit of 1 mM or 500 µg/ml, whichever is the higher, has been proposed by the ICH for the appropriate detection of genotoxic potential. The Scientific Committee notes that, although in general the scientific community agrees that there is no need to test up to 10 mM, the data are not yet sufficient to reach agreement on this new limit.

4. Considerations for basic test batteries

4.1. Core tests versus indicator tests

For initial screening of substances for genotoxic potential, the *in vitro* core test battery should be able to detect the three important genotoxic endpoints, i.e. gene mutations, structural chromosomal aberrations (i.e. clastogenicity) and numerical chromosomal aberrations (aneuploidy), in order to understand the genotoxic mode of action (genotoxic endpoint) of the tested substance. A range of different *in vitro* tests have been described in chapter 3.1.

Indicator tests (e.g. the Comet assay) are also described in chapter 3.1. Such tests detect pre-mutagenic lesions, which may not result in mutations, e.g. repairable DNA damage measured by the Comet assay. In addition, indicator tests do not give information of the mode of genotoxic action and should therefore not be included in the core set for hazard identification. However, indicator tests can be useful as follow-up test for *in vitro* positives and as supplementary tests for mechanistic studies, e.g. for the detection of oxidative DNA damage in the Comet assay using specific enzymes.

4.2. Number of tests in relation to exposure

4.2.1. High exposures

The issue of whether the extent of exposure (e.g. high or lifetime exposures) to substances might influence decisions on the number and type of tests to be included in a basic battery needs to be considered. For example, the WHO mutagenicity testing strategy for chemical risk assessment (Eastmond, et al., 2009) recommends the use of a basic battery of *in vitro* genotoxicity tests covering the endpoints of gene mutation, chromosomal aberration and aneuploidy. It goes on to recommend inclusion of *in vivo* testing as follow-up of negative results only in case of "high" or "moderate and sustained" human exposure, or for substances of high concern. Other guidance, such as that of the UK Scientific Committee on Mutagenicity has recommended three rather than two *in vitro* tests at the first stage of testing and progression to *in vivo* testing, even if *in vitro* tests are negative, in cases where exposures are "high, moderate or prolonged" (COM, 2000). This guidance is at present under revision.

In the Scientific Committee's view, the level or duration of exposure is not the first consideration in devising a basic test battery. If a basic battery of *in vitro* tests can be devised that adequately assesses the potential for genotoxicity of any substance, covering all three critical endpoints (i.e. induction of gene mutations, structural and numerical chromosomal alterations), then the level or extent of exposure is not relevant. However, it is recognised that inclusion of an *in vivo* test may be appropriate for substances designed to be biologically active (e.g. pharmaceuticals), and particularly if carcinogenicity tests are not available. Also in some cases it may be advantageous to include *in vivo* assessment of genotoxicity by incorporating such testing within other repeated-dose toxicity studies that will be conducted anyway (see 6.1). In other cases the necessity for *in vivo* follow-up should be considered case-by-case.

4.2.2. Low exposures

In situations where there is exposure to very low concentrations of substances in food/feed, an alternative approach, the Threshold of Toxicological Concern (TTC) has been proposed. Application of the TTC approach requires knowledge only of the chemical structure of the substance concerned and reliable information on human exposure. It is a screening tool that has been developed in order to assess substances of unknown toxicity that are present at low levels in the diet. It utilises generic human exposure threshold values below which the probability of adverse effects on human health is considered to be very low. The human exposure threshold values have been established based on data from extensive toxicological testing in animals. Human exposure threshold values have been developed for both cancer and non-cancer endpoints, and also for substances both with and without a structural alert for genotoxicity. The approach can also be used for substances for which genotoxicity data are not available.

The TTC approach will be the subject of a separate opinion from this Scientific Committee and it is anticipated that the opinion will be adopted by the end of 2011.

4.3. Are there unique *in vivo* positives?

Some substances, which are negative or equivocal *in vitro,* demonstrate *in vivo* positive results; among these substances, two categories of compounds should be distinguished, as investigated by International Workshop on Genotoxicity Testing (IWGT) (Tweats et al., 2007a, b). As reported by Tweats et al. (2007a), in the first category are substances inducing disturbances in the physiology of the rodents used in the micronucleus assay that can result in increases in micronucleated cells in the bone marrow that are not related to the intrinsic genotoxicity of the substance under test. These disturbances include changes in core body temperature, such as hypothermia, examples of which are chlorpromazine and reserpine (Asanami et al., 1998; Asanami and Shimono, 1997), and hyperthermia, an example of which is oxymorphone (Shuey et al., 2007). Increases in erythropoiesis following prior toxicity to erythroblasts (for example inhibitors of proteins synthesis like cycloheximide) or direct stimulation of cell division (for example erythropoietin) in these cells are also involved in the generation of positive results in the *in vivo* micronucleus assay. Whether these results are relevant for humans under realistic exposure conditions should be considered case-by-case.

As reported by Tweats et al. (2007b), in the second category, are substances that appear to be more readily detected *in vivo* than *in vitro*, or not highlighted *in vitro*. The reasons for this property vary from substance to substance and include metabolic differences, the influence of gut flora, higher exposures *in vivo* compared to *in vitro*, and pharmacological effects such as folate depletion or receptor kinase inhibition. Some examples are given below.

Urethane was classified by the International Agency for Research on Cancer (IARC) as a carcinogen, category 2B. There are sporadic reports of positive results for urethane in a variety of *in vitro* tests for genotoxicity, usually in the presence of rat liver S9 (Tweats et al., 2007b). Using protocols that include recent recommendations for the *in vitro* micronucleus assay, urethane was judged as negative (Lorge et al., 2006) in several cell lines including human lymphocytes (Clare et al., 2006) except in CHL cells (Wakata et al., 2006). Urethane was shown to be a strong genotoxin in the mouse bone-marrow micronucleus assay (Ashby et al., 1990). It produced significant increases in the *lacZ* mutant frequency in the liver and lung in Muta^{TM}Mouse transgenic model (Williams et al., 1998) and in the lambda/cII mutant frequency of BigBlue® *lacI/cII* transgenic mice (Mirsalis et al., 2005). It induced DNA adducts in mouse liver and lung (Fernando et al., 1996). Forkert and Lee (1997) demonstrated that urethane metabolism in lung microsomes is mediated by CYP2E1 and the carboxylesterase isozyme hydrolase A. Using the standard induction procedures, the level of CYP2E1 in rat liver is actually suppressed and this may account for the negative findings with these substances in the Ames test and other *in vitro* tests (Burke et al., 1994). Similarly, the lack of CYP2E1 in S9 from induced rat liver could be the explanation for the absence of *in vitro* mutagenic activity of benzene (Burke et al., 1994).

Procarbazine is another example of false negative results in *in vitro* tests, for example in the Ames test (Gatehouse and Paes, 1983) and in the human lymphocyte micronucleus assay (Vian et al., 1993), due to inappropriate metabolic activation systems, while it is clearly positive *in vivo* in the liver and lung in the Comet assay (Sazaki et al., 1998) and in the mouse bone marrow micronucleus test (Cole et al., 1981).

Tweats et al. (2007b) presented the cases of salicylazosulfapyridine and sulfapyridine. Salicylazosulfapyridine increases the incidence of urinary bladder tumours in rats and of liver tumours in the mice, but it is negative in the Ames test and in tests for chromosomal aberration and sister chromatid exchanges in CHO cells, but positive in the mouse bone marrow micronucleus test. Micronuclei are mainly, but not exclusively, kinetochore-positive, which suggests that an aneugenic mechanism is involved. Tweats et al. (2007b) also presented several other cases of unique *in vivo* positive substances.

Thus, there are small subsets of substances with particular mechanisms of action or specific metabolic routes for which conventional *in vitro* test batteries may miss true *in vivo* genotoxic agents including carcinogens. If there are indications from other data that such mechanisms or routes of metabolism not covered *in vitro* are applicable to the substances under consideration, then the possibility of *in vivo* testing should be considered.

4.4. The three Rs principle

The 3Rs (Russell & Burch, 1959) constitute an ethical framework by which the use of animals in research projects and for safety testing for regulatory purposes can be reviewed to help ensure humane experimentation. The 3Rs are defined as Replacement, Reduction and Refinement of animal testing.

The basis of the European legislation on the welfare of animal used for scientific purposes is the Council Directive 2010/63/EU on the protection of animals used in scientific experiments. This new Directive, which replaces the former Directive 86/609/EEC, seeks to strengthen significantly the protection of animals still needed for research and safety testing. The "Three Rs" principle is firmly anchored in the new legislation, which strongly supports efforts to find alternative methods to testing on animals. Where this is not possible, the number of animals used must be reduced or the testing methods refined so as to cause the least harm to the animals (EFSA, 2009).

The 3Rs principle applies also to genetic toxicology testing where complete or partial replacement can be envisaged using *in vitro* methods and non-testing methods such as *in silico* methods, read across, etc, and reduction and refinement can be applied to the current *in vivo* tests.

Several factors, discussed earlier, have been identified that may be important in the generation of false positive *in vitro* results. While an improvement in terms of increased specificity of the *in vitro* testing battery will likely reduce the number of *in vivo* studies required to follow-up positive outcomes from *in vitro* tests, additional efforts will be needed to ensure a reduction of the total number of animals used. For *in vivo* studies, many opportunities are currently available to reduce the number of animals and these possibilities are summarised in an ECVAM report (Pfuhler et al., 2009). Most of these are scientifically acceptable and some of them already compliant with regulatory guidelines. They include:

- The possibility to use of one sex only is already foreseen in OECD TG 474 (*in vivo*micronucleus test). While the use of both sexes should be considered if any existing data indicate a toxicologically meaningful sex difference in the species used, a survey on common practice in industry has shown that the majority of laboratories perform most of their studies using both sexes (Pfuhler et al., 2009).
- One administration and two sampling times versus two or three administrations and one sampling time for micronucleus, chromosomal aberration and Comet assays.
- The integration of the micronucleus endpoint into repeated-dose toxicity studies (see section 6.1).
- The combination of acute micronucleus and Comet assay studies. The protocol applied is compliant with guidelines, except for sampling times (see section 6.1).
- The omission of a concurrent positive control in routine chromosomal aberration and micronucleus tests has been shown to be scientifically acceptable, although the OECD guidelines still require this. Several possibilities have been proposed, from complete omission of a positive control animal group in a laboratory that has established competence in use of the assay to the use of a control group only periodically or a reduction in the number of animals in the control group.

5. Recommendations for an optimal testing strategy for food/feed substances

The Scientific Committee recommends a stepwise approach for genotoxicity testing of substances used in food and feed: a first step with testing in a “core set” of *in vitro* tests and, where necessary, a second follow-up step which can include both *in vitro* and *in vivo* tests. The basic battery used in the first step of testing includes a combination of mutagenicity tests which can detect gene mutation,

structural and numerical chromosomal aberrations. “Indicator tests” (see section 4.1) are not part of the basic battery.

5.1. Basic battery options

5.1.1. General considerations

In 1991 there were up to 200 different genotoxicity test systems (Waters et al., 1991) and in 2007 there were 16 OECD guidelines for genotoxicity tests. Over time, a number of batteries of short-term tests have been proposed and various strategies for their use proposed (for an early example, see Ashby (1986)). There has been a dichotomy, not necessarily complete, between pragmatic, usually empirical schemes and those with a theoretical underpinning. Some, for instance, Ennever and Rosenkranz (1986) suggested batteries of *in vitro* and *in vivo* tests based upon their empirical performance while the UK Committee on Mutagenicity (COM, 2000; ECHA, 2008b) developed a strategy based upon tiers, with a set of *in vitro* tests providing the first tier and then, if necessary, a move to a second tier based upon *in vivo* somatic tests, followed by *in vivo* germ cell mutation tests with the potential for quantification of the risk.

The Scientific Committee considered five main points as essential for the development of a test strategy:

- Firstly, there should be a step-wise approach with *in vitro* testing preceding *in vivo* testing.
- Secondly, the tests should aim to evaluate the genotoxic potential of the substance assessing induction of gene mutation, structural (clastogenicity) and numerical (aneuploidy) chromosomal alteration.
- Thirdly, the set of tests should be as small as possible.
- Fourthly, when following up positive *in vitro* tests, if it is decided that *in vivo* testing is necessary, a flexible and intelligent approach should be applied and no more studies should be performed than are required for clarification of the relevance of positive *in vitro* results.
- Fifthly, indicator tests, which detect primary DNA damage, should not be part of the basic battery; however, such tests could be useful in the follow-up of *in vitro* positive results.

Before embarking on any testing, it is important for the appropriate conduct of the tests, to consider other relevant knowledge on the substance such as its physico-chemical properties and experimental data on its toxicokinetics. Supporting information may also be available from Structure Activity Relationship (SAR) data, and ‘read-across’ of data between structurally-related substances. This information can also be important for interpretation of genotoxicity testing results and particularly relevant for the choice of any *in vivo* study.

The Scientific Committee considered whether a separate *in vivo* test should be included in the first tier of testing and broadly agreed that it should not. The Scientific Committee noted that a few substances had been identified as negative by *in vitro* testing although positive *in vivo* (see 4.3.); this may be due for example to the lack of specific metabolic factors in the *in vitro* system or to the involvement of specific conditions such as reactions in the gastro-intestinal tract. If there are indications that this may be the case for the substance of interest, it may either require appropriate modification of the *in vitro* test, or an *in vivo* test at an early stage of testing.

5.1.2. *In vitro* studies

Two *in vitro* tests are proposed for the first step of testing:

- the bacterial reverse mutation assay (OECD TG 471) and
- the *in vitro* micronucleus test (OCED TG 487).

This approach fulfils the basic requirement to cover the three genetic endpoints with the minimum number of tests. The data reviewed earlier in this opinion show that these two tests are reliable in detecting potential genotoxic carcinogens and the addition of further mammalian cell *in vitro* tests

reduces specificity with no substantial gain of sensitivity. Nevertheless, in case of equivocal or contradictory *in vitro* results, further *in vitro* testing may be useful to clarify the genotoxic potential *in vitro* (see also section 7).

If all *in vitro* endpoints are clearly negative in adequately conducted tests, then it can be concluded with reasonable certainty that the substance has no genotoxic potential. However, as mentioned above, the Scientific Committee notes that a small number of substances that are negative *in vitro* have positive *in vivo* results, because, for example, the *in vitro* metabolic activation system does not cover the full spectrum of potential genotoxic metabolites generated *in vivo*. The Scientific Committee acknowledges that the proposed step-wise testing strategy may not pick up every single genotoxic substance. This is not different from other currently used testing strategies. However, it is clear from the published literature, that these exceptions will be rare. The Scientific Committee therefore recommends that consideration of whether to proceed to *in vivo* testing in the case of negative *in vitro* results should be considered case-by-case, using a documented weight of evidence approach.

5.1.3. Follow-up of positive results from a basic battery

If positive results are obtained in the basic battery of *in vitro* tests, before embarking on the next step, all relevant data should be reviewed. The next steps may be (a) a conclusion of the assessment without further testing, (b) further *in vitro* testing, and/or (c) *in vivo* testing. One or more positive *in vitro* tests normally require follow up by *in vivo* testing. However, on occasion it may be demonstrated that the positive *in vitro* findings are not relevant for the *in vivo* situation, or a decision is taken to complete the assessment for other reasons.

5.1.4. *In vivo* studies

The Scientific Committee recommends that any *in vivo* tests should be selected on a case-by-case basis with flexibility in the choice of test, guided by the full data set available for the compound.
In vivo studies should relate to the genotoxic endpoint(s) identified *in vitro* and to appropriate target organs or tissues. The approach should be step-wise. If the first study is positive, no further test would be needed and the substance can be considered as an *in vivo* genotoxin. If the test is negative, it may be possible to conclude that the substance is not an *in vivo* genotoxin. However, in other cases, a second *in vivo* test may be necessary on an alternative tissue. There are also situations where more than one *in vitro* test is positive and an *in vivo* test on a second endpoint may be necessary.

The following *in vivo* tests can be considered for follow-up of *in vitro* positives:

- the *in vivo* erythrocyte micronucleus test (OECD TG 474),
- the *in vivo Comet assay* (no OECD TG at present; internationally agreed protocols available (e.g. see: hptt://cometassay.com), and
- the transgenic rodent assay (draft OECD TG - OECD, 2010b).

It is important that there is kinetic evidence that the agent reaches the tissue under investigation, and if the test is negative, it may be necessary to consider other relevant tissues (e.g. site of contact tissues for highly reactive substances which are not systemically available).

5.1.5. Examples of follow-up approaches

In following up *in vitro* positives, the *in vivo* test(s) selected should relate to the genotoxic endpoint(s) identified as positive in the *in vitro* tests. As examples of how decisions on follow-up might be made, some typical scenarios and approaches are described below. However, the Scientific Committee wishes to emphasise that these are only illustrative and that alternative approaches may be appropriate.

In the case of positive results from the basic battery of tests, the three following scenarios typically occur:

(I) bacterial reverse mutation test positive and *in vitro* micronucleus test negative,
(II) bacterial reverse mutation test negative and *in vitro* micronucleus test positive, or
(III) both bacterial reverse mutation test and micronucleus test positive.

Scenario I: Bacterial reverse mutation test positive – *in vitro* micronucleus test negative

Before any decisions are made about the need for *in vivo* testing to follow-up a positive bacterial reverse mutation test, the possibility of a unique positive response, due for example to a specific bacterial metabolism of the test substance, should be considered .

Appropriate *in vivo* tests to follow-up a bacterial reverse mutation test that is not considered to be a bacteria-specific effect would be to conduct a transgenic rodent mutation assay or a rodent Comet assay. Both assays are also suitable for detection of first site of contact effects. Adequate target tissues should be selected based on information about direct reactivity of the substance with DNA (which might predispose to site of contact effects), bioavailability, metabolism, toxicokinetics, and any target organ specificity (if known from repeat-dose toxicity studies).

A combination of the Comet assay with analysis for micronuclei using the same animals could be considered, even in cases in which the *in vitro* micronucleus test is negative, since most substances that are positive in the bacterial reverse mutation test are DNA reactive substances that should be considered as potentially clastogenic too. If an adequately conducted rodent Comet assay (or combined Comet/*in vivo* micronucleus test) is negative it will normally be possible to conclude that the test substance is not mutagenic *in vivo.*

Scenario II: Bacterial reverse mutation test negative – *in vitro* micronucleus test positive

Key points to consider for selection of appropriate *in vivo* follow-up studies under scenario II include clarification of relevant mode of action for micronuclei induction (e.g. discrimination between clastogenic and aneugenic effects with use of centromere/kinetochore stains or FISH technologies), where such information is available, and possible involvement of genotoxic metabolites (e.g. positive test result only in the presence rat liver S9 mix).

IIa. If the available data show an aneugenic effect *in vitro* (i.e. increase in centromere-positive micronuclei) an *in vivo* rodent micronucleus test (in bone marrow or peripheral blood) would typically be considered appropriate to follow-up the *in vitro* finding. If an adequately conducted *in vivo* micronucleus test (with evidence for significant exposure of the target tissue from an absorption, distribution, metabolism and excretion (ADME) study or from changes in the percentage of polychromatic erythrocytes in the blood) is negative, it will normally be possible to conclude that the test substance is not aneugenic *in vivo*.

IIb. If the available data show a clastogenic effect *in vitro* (i.e. increase in centromere-negative micronuclei) in the absence of rat liver S9 mix, an *in vivo* rodent micronucleus test (in bone marrow or peripheral blood) would typically be considered as appropriate and sufficient to follow-up the *in vitro* finding. If an adequately conducted *in vivo* micronucleus test (with evidence for significant exposure of the target tissue from ADME study or from changes in the percentage of polychromatic erythrocytes in the blood) is negative, it will normally be possible to conclude that the test substance is not an *in vivo* clastogen.

IIc. If available data show a clastogenic effect *in vitro* and the effect is seen exclusively (or predominantly) in the presence of rat liver S9 mix, the involvement of liver-specific clastogenic metabolites should be considered. In this situation a single rodent study combining micronucleus analysis (in bone marrow or blood) and a Comet assay in the liver should be considered. If an

adequately conducted combined *in vivo* micronucleus test/Comet assay (with evidence for significant exposure of the target tissues from ADME study or from changes in the percentage of polychromatic erythrocytes in the blood) is negative, it will normally be possible to conclude that the test substance or its metabolites are not clastogenic *in vivo*.

Scenario III: Bacterial reverse mutation test positive – *in vitro* micronucleus test positive

A combined *in vivo* micronucleus test/Comet assay with adequate target tissue selection (see above) is recommended to follow up compounds that are positive in both of the basic *in vitro* tests. If an adequately conducted combined micronucleus test/Comet assay (with evidence for significant exposure of the target tissues from ADME study or from changes in the percentage of polychromatic erythrocytes in the blood) is negative, it will normally be possible to conclude that the test substance is not genotoxic *in vivo*.

5.2. Role of germ cell testing

The Scientific Committee considers that routine testing for genotoxicity in germ cells is not necessary. Systemic exposure to a substance should usually result in t reaching the germ cells if there is systemic diffusion and it has not been demonstrated that the gonadal-blood barrier prevents the substance reaching the germ cells. A positive *in vivo* genotoxin in somatic tissues would, therefore, be assumed to be a germ cell mutagen. The corollary is that a substance that is negative in somatic cells would, providing adequate testing has been done, be considered a negative germ cell mutagen. The lack of genotoxicity *in vivo* in somatic cells gives reassurance on the absence of genotoxicity at the germ cell level too, and moreover no germ cell specific mutagen is known.

On the other hand, for substances which are genotoxic in somatic cells *in vivo,* the potential for germ cell mutagenicity should be considered. It is recognised that standard reproduction studies do not cover all germ cell effects. Thus, the need to perform genotoxicity tests in germ cells should be decided case-by-case. If there is evidence that germ cells are actually exposed to a somatic mutagen or its active metabolite, it is reasonable to assume that the substance may also be a germ cell mutagen and hazardous to future generations without performing specific tests. If for some reason it is considered necessary to conduct testing in germ cells, the methods fall into two classes: (1) tests on germ cells *per se;* (2) tests on the offspring of exposed animals. Only the latter provide information suitable for the quantitative evaluation of transmissible genetic risk.

6. Other issues in testing substances present in food/feed

6.1. Combining genotoxicity testing with repeated-dose toxicity testing and the micronucleus test with the Comet assay

Recently proposed guidance on genotoxicity testing of pharmaceuticals (ICH, 2010) and chemicals (ECHA, 2008b) encourage integration of genotoxicity tests into repeated-dose toxicity (RDT) studies, whenever possible and scientifically justified. An integrated measurement of genotoxicity endpoints offers the possibility for an improved interpretation of genotoxicity findings since such data will be evaluated in conjunction with routine toxicological information obtained in the RTD study, such as haematology, clinical chemistry, histopathology and exposure data. In addition such an approach obviously contributes to the reduction of animal use in genotoxicity testing as it usually would replace a stand-alone *in vivo* genotoxicity study (Pfuhler et al., 2009).

Integration of the micronucleus endpoint into RDT studies is in compliance with the OECD guideline for the *in vivo* micronucleus test (OECD, 1997). Broad experience with the micronucleus test shows the feasibility of integrating both blood and bone marrow micronucleus analysis into RDT studies in rats, the standard rodent species for general toxicity studies.

With other genotoxic endpoints, there is less or no experience as yet. Due to its flexibility, the *in vivo* Comet assay could easily be incorporated into RDT studies, and when conducted with micronucleus

analysis, such a combination could cover systemic genotoxic effects as well as local effects (site of contact tissue and target organ for toxicity) and different genotoxic mechanisms. Results from a recent collaborative trial confirm that the liver Comet assay can be integrated within RDT studies and efficiently complements the micronucleus assay in detecting genotoxins (Rothfuss et al., 2010).

In Appendix C some practical aspects are discussed that need to be considered when combining micronucleus and Comet assays in RDT studies.

6.2. Evaluation of metabolites, degradation and reaction products

The use of plant protection products results in exposure of consumers to a mixture of compounds including the active substance, its plant metabolites, degradates and other transformation products present in food commodities. In addition, the continuous improvement in analytical methods and sensitivity, results in the detection of an increasing number of compounds at low levels and also in the identification of new compounds. Only the active substances are directly investigated through a range of toxicological studies required by the current regulations, while limited information is available for metabolites and degradates and requests for further toxicological studies are restricted as far as possible to minimise the use of animals in toxicological testing.

The EFSA PPR Panel has an ongoing activity to develop an opinion on approaches to evaluate the toxicological relevance of metabolites and degradates of pesticide active substances in dietary risk assessment. Within the frame of a commissioned project to the UK Chemicals Regulation Directorate (CRD), the applicability of the TTC scheme was tested with 100 actives substances randomly selected from a list of 500 compounds evaluated under the EU Directive 91/414/EEC. It showed the TTC approach to be a potentially useful tool as a preliminary step in safety assessment of metabolites and degradation products of pesticides present in food at very low concentrations. A case study was also carried out with 15 active substances and their metabolites, comparing the exposure estimates with the respective TTC value. An outcome of this exercise, confirmed by further case studies carried out by the PPR working group on pesticide metabolites, was that the TTC for genotoxicity is easily exceeded (The Technical Report on this project is available at: www.efsa.europa.eu/en/scdocs/scdoc/44e.htm).

The applicability of analysis of structure-activity relationships (SAR) in the evaluation of genotoxicity alerts in pesticide metabolites was investigated in a project outsourced to the Computational Toxicology Group of the European Commission Joint Research Centre (DG JRC, Ispra). A range of computer-based predictive models (DEREK, CAESAR, LAZAR, TOPKAT, Hazard Expert, ToxBoxes and Toxtree) was tested with three datasets consisting of 185 pesticides, 1290 heterogeneous chemicals (Distributed Structure-Searchable Toxicity- DSST database), and 113 heterogeneous classified mutagens. A wide range of sensitivity and specificity was found with the different tools, with better performance in predicting bacterial mutagenicity. According to the report, pairwise combinations of these tools could increase the overall sensitivity to about 90% (JRC, 2010). However, the Scientific Committee notes that there are differing views of the usefulness of these approaches.

7. Data interpretation

7.1. Consideration of equivocal and inconclusive results

The Scientific Committee considered the issue of how to classify a test result as either positive or negative and what defining a result as equivocal or inconclusive meant. It was recognised that dichotomising results as either positive or negative carries some risk of an incorrect ‘call’. Dichotomising when the substance is a weak genotoxin could also result in contradictory results between repeat experiments. The Scientific Committee recommends that in the event of an equivocal result, repeat experiments should be run. These might, for instance, involve using different dose-ranges. Consideration should be given to the size of any genotoxic effect in an experiment.

Distinguishing between the meaning of an equivocal and an inconclusive result is difficult as the two words are synonyms and often used interchangeably. The term 'equivocal result' usually refers to a situation where not all the requirements for a clear positive result have been met. An example could be where a positive trend was observed, but the dose-response relationship is not statistically significant.

Equivocal can, therefore, be interpreted as possibly relating to the true state of nature as the true result is on the borderline of the decision criteria for positive or negative. In the context of testing, it could imply a weak positive result as opposed to a clear positive or negative. Repeated testing would then result in results falling just one side or the other of the decision criteria.

An inconclusive result could be considered one where no clear result was achieved but this may have been a consequence of some limitation of the test or procedure. In this case, repeating the test under the correct conditions should produce a clear result.

Results classified in this way should be examined with respect to their quality. It was noted that meeting Good Laboratory Practice (GLP) requirements provides confidence in the integrity of the study but does not necessarily guarantee the quality of the results. If necessary, further testing might be suggested taking into account the supplementary information already available.

7.2. Evaluation of the quality and reliability of data

Evaluation of the quality and reliability of the available data on toxicity (including genotoxicity) is crucial in risk assessment. Generally, genotoxicity tests should be performed according to international standards, preferably according to the current OECD test guidelines or the EU Test Methods Regulation (EC) 440/2008 (EU, 2008), and in compliance with the principles of Good Laboratory Practice and Good Cell Culture Practice (GLP, GCCP). Further advice on the performance of tests is available in guidance from the International Workshops on Genotoxicity Testing (IWGT) (Kirkland et al., 2007b; Kasper et al., 2007; Burlinson et al., 2007; Tweats et al., 2007a,b; Thybaud et al, 2010). The highest level of reliability can be attributed to test results obtained from studies performed under such conditions. While for many of the substances which are intentionally added to food or feed such data can be requested or required from the petitioner, the risk assessment of substances like contaminants in food must be performed on whatever data are available. Therefore, it is in all cases important to evaluate the quality and reliability of the available data.

There is no specific guidance for the evaluation of the quality and reliability of genotoxicity data, however, useful guidance on how to evaluate available information gathered in the context of registration, evaluation and authorisation of chemicals (REACH) is provided in a guidance document of the European Chemicals Agency (ECHA, 2008a).

The evaluation of data quality includes assessment of *relevance, reliability* and *adequacy* of the information. These terms were defined in the ECHA guidance document (ECHA, 2008a) based on definitions by Klimisch et al. (1997) as follows:

> "**Relevance** - covering the extent to which data and tests are appropriate for a particular hazard identification or risk characterisation.
>
> **Reliability** - evaluating the inherent quality of a test report or publication relating to preferably standardised methodology and the way the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings. Reliability of data is closely linked to the reliability of the test method used to generate the data.
>
> **Adequacy** - defining the usefulness of data for hazard/risk assessment purposes. Where

there is more than one study for each endpoint, the greatest weight is attached to the studies that are the most relevant and reliable. For each endpoint, robust summaries need to be prepared for the key studies."

The Scientific Committee noted that in order to evaluate the **relevance** of the available genotoxicity data it should be considered whether the data were obtained from studies providing information on one of the three genetic endpoints (i.e. induction of gene mutations, structural and numerical chromosomal alterations) or on other genotoxic effects. Studies covering one of the three genetic endpoints would be most relevant, however, studies on other effects could provide useful supporting information.

Additionally, there are several further issues which could have an impact on the relevance of the study results. Some examples are as follows:

- *Purity of test substance*: Generally, substances tested for genotoxicity should have high purity. However, data obtained with a substance of lower purity might be more relevant if this was the substance to be used in food.
- *Uptake/bioavailability under testing conditions*: In certain cases, the standard testing protocols (e.g. OECD guidelines) may not ensure the bioavailability of test substances. This should be taken into consideration and may apply for example to poorly water- soluble substances or nanomaterials.
- *High cytotoxicity*: A positive result from an *in vitro* test in mammalian cells would be considered of limited or even no relevance if the effect was only observed at highly cytotoxic concentrations.
- *Metabolism*: A negative result obtained with a substance in an *in vitro* assay in which the standard exogenous metabolising system does not adequately reflect metabolism *in vivo* would be considered of low relevance (e.g. azo-compounds).
- *Exposure of target tissue*: A negative result from an *in vivo* study would have limited or even no relevance if there was no indication from the study that the test substance reached the target tissue and if there were no other data, e.g. toxicokinetic data, on which such an assumption could be based.
- *Reproducibility of results*: If conflicting results that were produced with tests that have similar reliability were observed , it should be judged whether this might be attributable to differences in specific test conditions, e.g. concentrations, animal strains, cell lines, exogenous metabolising systems, etc. If no plausible explanation could be found this might limit the relevance of the data and it should be considered whether a further study would be required in order to clarify the issue.
- *Equivocal results* are generally less relevant than clearly positive results, however, they may be considered as an indication for a possible 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.

Reasons why the **reliability** of data could be different may include the use of non-validated test protocols, outdated test guidelines or the failure to characterise the test substance properly with respect to chemical identity and purity. Other reasons could be poor reporting of information on study design and/or results, and poor quality assurance.

If it is considered necessary to make a formal assessment of quality and reliability of the data, then the Scientific Committee recommends that the approach of Klimisch et al. (1997) be used. This approach uses a scoring system to assess the reliability of toxicological data which is cited in the ECHA guidance document (ECHA, 2008a).

7.3. Utility of toxicokinetic data in the interpretation of genotoxicity data

While *in vitro* genotoxicity test data gives information on the intrinsic genotoxic property of the tested substance, for interpretation of the *in vivo* genotoxicity testing results as well as for the strategy of the

follow-up testing, information on the toxicokinetics of the substance (e.g. systemic availability, exposed organs, pathways possibly involved in its metabolism, and elimination pathways) should be scrutinized. In cases of *in vitro* positive results, *in vivo* testing is generally required to confirm *in vitro* results (see 3.3.4). Since i*n vivo* tests take into account absorption, distribution and excretion (this is not the case for *in vitro* tests), they are considered as potentially relevant to human exposure. In addition, metabolism is likely to be more relevant *in vivo* compared with the systems normally used *in vitro*. When the *in vivo* and *in vitro* results are not consistent, then the differences should be clarified on a case-by-case basis. For example, in the *in vivo* micronucleus test, certain substances may not reach the bone marrow due to low bioavailability or specific tissue/organ distribution. In certain cases, for example when it is known that the test substance is metabolised in the liver and the reactive metabolites formed are too short-lived to reach the bone marrow, even demonstration of the bioavailability of the parent substance in the bone marrow does not indicate that bone marrow is an appropriate target. A negative result of the *in vivo* micronucleus assay can be considered as meaningful only if there is definitive evidence from toxicokinetic data that the tested substance as well as the relevant reactive metabolite(s) can reach the bone marrow in significant amounts.

When follow-up testing is required, the selection of an appropriate experimental protocol for the testing *in vivo* should be based on the available information on the toxicokinetics of the agent (Pfuhler et al., 2007; ECHA, 2008b). In cases where toxicokinetic data indicate that the bone marrow is an inappropriate target, then alternative tissues such as liver, intestine, etc, should be considered. When *in vitro* positive results are seen only in the presence of the S9 activation system, the relevance of any reactive metabolites produced *in vitro* to conditions *in vivo* should be considered. *In vitro* metabolic activation with standard induced S9-mix has different activation capacity than human S9, and also lacks phase II detoxification capability. In addition, non-specific activation can occur *in vitro* with high test substrate concentrations (see Kirkland et al., 2007a). In such cases, analysis of the metabolite profile in the incubation mixture used in the genotoxicity test compared with known metabolite profiles in obtained from toxicokinetic studies can help in determining the relevance of test results (Ku et al., 2007; OECD, 2010c). However, there may be cases where the metabolic activation pathway of a pro-mutagenic agent is not efficiently represented in the standard *in vitro* metabolic activation system (rodent liver S9) because of the low expression of specific enzyme activities (e.g. CYP2E1) or the lack of cofactors (e.g. PAPS for sulphate ester formation). Information on the known or expected pathway of metabolic transformation may help identifying such cases and allow optimisation of the experimental conditions of testing (Ku et al., 2007).

Moreover, when *in vivo* testing is performed to follow-up *in vitro* positive results, the biological plausibility and relevance of the results obtained should always be critically considered, because positive results *in vivo* could arise as a consequence of metabolic overload or physiological disturbance, rather than by direct genotoxicity (Tweats et al., 2007a).

7.4. Consideration of other relevant data (SARs)

Non-testing methods refer to a range of predictive approaches, including Structure-Activity Relationships (SARs), Quantitative Structure Activity Relationships (QSARs), chemical grouping and read-across methods, or computer-based *in silico* tools based on the use of one or more of these approaches.

These methods are based on the premise that the properties (including physicochemical properties and biological activities) of a chemical depend on its intrinsic nature and can be directly predicted from its molecular structure or inferred from the properties of similar substances whose properties are known. The first list of structural alerts for mutagenicity was proposed by Ashby (1985), and was subsequently extended by using a combination of data mining and expert knowledge.

A wide range of commercial and free software tools are today available to predict genotoxicity and carcinogenicity, including:

(a) ruled-based systems combining toxicological knowledge and expert judgment (e.g. DEREK - Deductive Estimation of Risk from the Existing Knowledge)
(b) statistically-based systems (e.g. MultiCASE - Multiple Computer Automated Structure Evaluation), and
(c) hybrid models based on the combination of knowledge-based rules and statistically-derived models (e.g. Toxtree) (Serafimova et al., 2010).

More than 100 papers in the scientific literature are devoted to *in silico* prediction of genotoxicity, comparing performances of different (Q)SAR models, including software models; the large majority of them report the results of evaluation studies for prediction of carcinogenicity. The available models perform better for the prediction of bacterial mutagenicity (the accuracy of Ames test mutagenicity prediction is typically 70-75%) than for *in vitro* mutagenicity or cytogenetics in mammalian cells. A factor that contributes to reduced model performance is the nature of the underlying mutagenicity data, such as inconsistent data interpretation or the lack of quality assurance.

Overall, the present evidence does not justify the application of the (Q)SAR approach alone in predicting the genotoxicity of substances. In cases where it may not be possible to request testing (e.g. contaminants in the food chain), the (Q)SAR approach could be useful in aiding the interpretation of data using a weight-of-evidence approach, by including information from all available sources (QSARs, read across and experimental data).

7.5. Evaluating the outcome of genotoxicity and carcinogenicity studies

Rodent carcinogenicity data have been considered as the "gold standard" in the context of the review work conducted on correlations between *in vitro* genotoxicity and carcinogenicity in order to assess if a specific substance can be considered to be an *in vivo* relevant carcinogen. Historically, the genetic toxicology testing battery has been designed to be used as a surrogate for carcinogenicity testing. An important issue that needs to be discussed is whether a negative rodent carcinogenicity study can overrule a positive genotoxicity result. A decision on whether negative carcinogenicity data can overrule positive *in vitro* genotoxicity test results should be taken on a case-by-case basis. It is doubtful, tough, whether this also holds true for *in vivo* genotoxicity test results. Clear evidence of genotoxicity in somatic cells *in vivo* should be considered an adverse effect *per se*, since genotoxicity is also implicated in degenerative diseases other than cancer.

The prediction of carcinogens with a non-genotoxic mode of action is out of the scope of genotoxicity testing, and thus, in principle, only genotoxic carcinogens should be considered as the 'gold standard' for evaluating the predictive value of short-term tests. This approach, however, may be of limited feasibility because genotoxic carcinogens are usually defined on the basis of a positive score in genotoxicity tests, and therefore cannot also be used to evaluate the ability of short-term tests to detect their genotoxic potential. Interspecies differences in cancer susceptibility and rodent specific mechanisms of carcinogenicity should be considered when rodent carcinogens are used as the reference for the prediction of human risk. On the other hand, when human carcinogens (e.g. IARC class I carcinogens) are used as the reference, it has to be taken into account that in this category strong carcinogens (and mutagens), capable of providing direct evidence of carcinogenicity in humans, are likely to be over-represented compared to the universe of human chemical carcinogens.

7.6. Evaluation of pre-existing or non-standard data using weight of evidence

Although this opinion is broadly about genotoxicity testing strategies, an appreciable amount of EFSA's work in this field is in the assessment of data from experiments which have already been carried out and where the option of further testing may not be feasible in the short term. Such a dossier of genotoxicity data may have been collected over many years of experimentation in many laboratories using different assay methods and protocols. The studies may or may not have been carried out to prevailing guidelines at the time or to GLP. Some substances have no 'owners' or

'stewards' and, consequently, there may not be any groups prepared to produce new data using the set of tests currently favoured.

The Scientific Committee recognises that there is no definitive way to assess such dossiers. An example might be a large dossier with a mixture of positive, negative and equivocal results based upon studies using assays which have OECD Guidelines but no longer are considered core methods in testing strategies, studies which have some limitations in their conduct and studies based upon newer methods which have not been fully validated. Dossiers may also contain studies (e.g. from academic laboratories) which have been well-conducted and published after peer review. These studies should be considered on a case by case basis using expert judgement (see section 7.2). In particular, EFSA's CONTAM Panel often has to consider heterogeneous and non-standard data sets. The Scientific Committee recognises that in these cases EFSA has to rely on a weight of evidence approach to assess such data sets. All available mechanistic information should be taken into account and any uncertainties on genotoxic potential, including significant data gaps, should be explained in the opinion.

8. Recent and future developments

8.1. Thresholds for genotoxicity

According to an approach widely accepted until some years ago, all genotoxic substances were assumed to act through a non-threshold mechanism. This approach was based both on precautionary considerations and on a mechanistic model that considered the theoretical possibility that a single molecule could cause a DNA lesion, which might eventually be converted into a mutation.

There is today a consensus on the existence of a threshold for genotoxic agents that interact with molecular targets different from DNA (e.g. DNA polymerases, topoisomerases, spindle proteins). The interaction of reactive chemicals with spindle fibres or the interference with spindle checkpoint proteins is a potential cause of aneuploidy. It is accepted that spindle function is inhibited by an interaction with multiple binding sites, resulting in a dose–response curve with a threshold (Parry et al., 1994). A threshold mechanism of action has, therefore, been proposed for this class of substances (Elhajouji et al., 1995, 1997).

Topoisomerase I and II are enzymes that control changes in DNA structure by catalyzing the breaking and rejoining of the phosphodiester backbone of DNA strands during the normal cell cycle. Topoisomerase inhibitors block the ligation step necessary for the rejoining, generating single and double strand breaks that harm the integrity of the genome. It is accepted that genotoxic effects arising via such mechanisms show a threshold (ECETOC, 1997).

There is now experimental evidence that mutagens whose mode of action is based on the induction of reactive oxygen species (ROS) could act through a threshold mechanism. ROS are a normal component of the cellular environment, therefore the mutagenic potential of an oxidant depends on its capability to overcome the physiological cellular defences against oxidative damage. DNA-oxidizing agents belonging to different chemical classes have been recently reported to induce *in vitro* genotoxicity with a thresholded non-linear dose-response relationship (Platel et al., 2009).

The non-threshold model has also been questioned for DNA-reactive chemicals, at first on a theoretical basis, taking into account the presence of cellular defence mechanism (scavenging, detoxification, DNA repair etc.) that can protect DNA at low exposure levels. In the last few years several laboratory studies have confirmed that also in the case of some DNA-reactive agents a threshold is experimentally observable, while other substances display a linear dose-response relationship (see, for example, review by Jenkins et al., 2010).

While the existence of a threshold is now accepted for non-DNA reactive agents (e.g. spindle inhibitors), mutagens whose molecular target is DNA may also display non-linear experimental dose-response relationships, depending on the mode of action. A first distinction should be made between agents that indirectly target DNA, such as oxidants, and chemicals that directly interact with DNA, forming adducts. The issue is not the discussion of the theoretical basis of the alternative models but the definition of criteria to decide when a threshold or a non-threshold model is more appropriate and the consequences of this decision for the evaluation of genotoxic risk (EFSA, 2005).

The question is more controversial in the case of alkylating agents. It has been demonstrated that the *in vitro* genotoxicity of some alkylmethane sulphonates (EMS and MMS) shows a non-linear dose-response, containing a range of non-mutagenic low concentrations, and that a no-observed-effect level (NOEL) for genotoxicity can be set. In contrast, alkylnitrosoureas (ENU and MNU) concomitantly tested, appeared to induce genotoxic effects with a linear dose-response relationship (Doak et al., 2007). Similar results have been recently reported *in vivo* (Gocke and Müller, 2009). This difference could be because of different preferred targets for the two classes of alkylators, as alkylnitrosoureas are relatively more capable of alkylating oxygen atoms, producing more of the mispairing base O^6-alkylguanine, and also the poorly repaired O^2-alkylthymine and O^4-alkylthymine. At low dosages of MMS and EMS, the little amount of O^6-alkylguanine could be efficiently repaired by methylguanine DNA methyltransferase (MGMT), while MNU and ENU could rapidly saturate MGMT, causing linearly increasing mutation levels (Doak et al., 2007).

Several factors that modulate the interaction of alkylators and DNA are still under experimental investigation. In particular, little is yet known about the interspecies and inter-individual variability in metabolism and DNA damage response relevant to alkylating agents. For example it is known that DNA repair glycosylases show high inter-individual variability (Paz-Elizur et al., 2007) and significant inter-individual differences in the expression of MGMT were reported in the human population, both in lymphocytes (reviewed in Kaina et al., 2007) and in lung tissues (reviewed in Povey et al., 2007). Therefore the possibility of adopting a threshold model for alkylating chemicals should be considered with some caution and evaluated on a case-by-case basis.

The dose-response relationship is also affected by the metabolism of the chemical, as exemplified by the case of paracetamol, a drug also found as a food contaminant. N-acetyl-p-benzoquinone imine, produced by the oxidative metabolism of paracetamol, can form adducts on DNA, but only after depletion of cellular glutathione. This depletion occurs *in vivo* only at exposure levels inducing pronounced liver toxicity and above, the therapeutic dosage (Bergman et al., 1996). Another example of the role of metabolism is that of Chromium (VI), whose carcinogenic potential is due to a recognised genotoxic mechanism. Chromium (VI) is efficiently reduced in body fluids to Chromium (III), which does not easily cross cell membranes. Therefore, the genotoxic and carcinogenic potential of Chromium (VI) depends on the reductive metabolism being overwhelmed. Based on this, a thresholded mechanism for the carcinogenesis of Chromium (VI) has been proposed (De Flora, 2000).

No experimental evidence of thresholds has yet been found for many DNA-reactive agents. In these cases, a precautionary approach suggests the adoption of a linear dose-response model. The practical consequence of this approach is that no exposure level to these agents would be considered without risk. The strict application of this principle can be problematic in some specific situations. For example, in the case of some DNA-reactive chemicals occurring in food, a certain degree of exposure is unavoidable.

Standard genotoxicity testing is currently based on acute treatments, while human exposure, in particular to food-related chemicals, is generally chronic. The duration and degree of repeated exposure may have a strong influence on the saturation of the defence pathways and on the induction of enzymes associated with the response to the chemical, with important effects on the dose-response relationship, as recently demonstrated in *in vitro* experiments (Platel et al., 2009). While an experimental effort aimed to clarify this issue is desirable, the possibility that a longer duration of

exposure may lower the real threshold for humans should be taken into account as a further uncertainty.

Finally, it should also be considered that many chemical mutagens, including food contaminants, do not act via a single mode of action but through different concomitant mechanisms, with or without a threshold. The drug doxorubicin is a known example of this kind of complex action, but also chemicals relevant to food safety may display multiple mechanisms of genotoxic activity. For example, some metal ions can cause oxidative stress, interact with proteins involved with genome stability and form adducts on DNA (McCarroll et al., 2010). Similarly, topoisomerase inhibition, a potential cause of DNA breakage, was reported also to affect DNA repair; a reduced activity of the incision step of nucleotide excision repair was observed in human fibroblasts treated with different topoisomerase I and II inhibitors (Thielmann et al. 1993). In such cases, a simplistic model based on a single prevalent mode of action could underestimate the actual risk for human health.

8.2. Promising new test methods

8.2.1. Genotoxicity assays based on induction of DNA Damage Response (DDR)/stress pathways gene transcription

In the last few years several attempts have been made to develop and validate the induction of stress pathways/proteins as end-points in genotoxicity assays by using high throughput screening approaches. The choice of the pathways was mostly based on microarray experiments with genotoxic substances. The GreenScreen HC assay, that uses *p53*-competent TK6 lymphoblastoid cell line genetically modified to incorporate a fusion cassette containing the GADD45alfa promoter (and other regulatory elements) and the *GFP* gene as reporter (Hastwell et al., 2006), has been widely characterised and its high specificity confirmed in independent studies (reviewed in Birrel et al, 2010). HTS luciferase reporter assays based on four different stress pathways (RAD51C, Cystatin A, *p53* and *Nrf2*) in the HepG2 cell line have also been developed and shown to be useful for pre-screening in early phases of drug development (Westerink et al., 2010).

A recent study has addressed the question of whether the use of these new assays may reduce false positive results (Birrell et al., 2010). The same list of chemicals used by Fowler et al. (2011, publication submitted) was tested in the GreenScreen HC assay. Of the 17 chemicals tested 76% (13/17) were negative. Of the remaining four, *p*-nitrophenol was only positive at the top dose, 2, 4-dichlorophenol is an *in vivo* genotoxin and two chemicals (i.e. *tert*-butylhydroquinone and curcumin) are antioxidant substances that can act as pro-oxidants in the hyperoxic conditions of cell culture. The results suggest that the generation of false positives is minimized by the GreenScreen HC assay. In the same study substances that should be detected as positive in *in vitro* mammalian cell genotoxicity tests were tested and 18/20 (90%) were reproducibly positive. Substances that should give negative results in *in vitro* genotoxicity tests were also reproducibly negative (22/23, 96%). Although the number of chemicals tested is limited, these data overall suggest a good sensitivity and specificity of this assay. However, the mechanistic basis of these transcriptional assays does not guarantee that DDR/stress pathways gene activation will necessarily involve DNA damage. For example, GADD45alfa activation can be achieved by histone deacetylase inhibitors (e.g. Trichostatin A), various non-steroidal anti-inflammatory drugs such as aspirin, and specific iron chelators such as desferrioxamine. Moreover, changes in osmoregulation and any alteration of the redox cell status will end-up with transcriptional changes of these genes too (reviewed in Siakafas and Richardson, 2009).

Destici et al., (2009) have also observed that DNA damaging agents can synchronise the circadian clock of cells in culture and, as a consequence, the expression of circadian clock genes that include some DDR genes (e.g. *p53*, *p21*) thus blurring the profile of transcriptional response to DNA damage. The alternative approach of running these assays on cells in which intracellular clocks are synchronised prior to exposure should be evaluated also for its potential impact on the sensitivity of these assays.

Thus, on the basis of the currently available information, these assays show promise mostly as a pre-screening step to gain insights into the mechanisms of action of substances and guide the testing strategy. However, they are not ready to be used as potential new genotoxicity tests without further studies.

8.2.2. A new *in vivo* test for gene mutation: the *Pig-a* mutation assay

The majority of current regulatory test batteries do not include an *in vivo* test for gene mutation because an *in vivo* gene mutation test that is sufficiently sensitive and practical to be used for regulatory safety assessments is currently not available. The *Pig-a* gene mutation assay addresses this need, at least to a certain extent. The *Pig-a* gene, located on the X-chromosome, codes for a catalytic subunit of the N-acetylglucosamine transferase complex that is involved in an early step of glycosylphosphatidyl inositol (GPI) anchor synthesis (Takahashi et al., 1993). Although this assay can be carried out on a number of species and cell types at present only blood cells have been successfully used. Most published studies have used rat red blood cells and reticulocytes (reviewed in Dobrovolsky et al., 2010). The test protocol requires small blood volumes (µlitres) if a flow cytometric assay is carried out and this makes integration with, for instance, repeat-dose toxicology tests highly feasible. However, the sensitivity of this assay for detecting known mutagens and carcinogens has not yet been well defined and standard protocols for analysis and data interpretation have not been established.

8.2.3. Cell Transformation Assays

CTAs have been in use for 40 years and are currently used by academia, and by the chemical, agro-chemical, cosmetic, tobacco and pharmaceutical industries. CTAs are conducted to screen for potential carcinogenicity, as well as to investigate mechanisms of carcinogenicity. Currently, CTAs are also used for clarification of *in vitro* positive results from genotoxicity assays as part of a weight of evidence assessment. Data generated by the CTA can be useful where genotoxicity data for a certain substance class have limited predictive capacity or for investigation of substances with structural alerts for carcinogenicity and to demonstrate differences or similarities across a chemical category. CTAs are also used to identify tumour promoters.

In vitro cell transformation assays (CTAs) have been shown to involve a multistage process that closely models key stages of *in vivo* carcinogenesis (LeBoeuf et al., 1999). They are thus used to detect phenotypic changes that are associated with malignant transformation *in vivo*. These morphological changes are a result of the transformation of cultured cells, which involves changes in cell behaviour and proliferation control (e.g. altered cell morphology, changed colony growth patterns and anchorage-independent growth). Moreover, when injected in suitable hosts these cells give rise to tumours.

In order to systematically assess the performance of the CTAs, the OECD published in 2007 a detailed paper on "Cell transformation assays for the detection of chemical carcinogens" aiming at reviewing all available data on the three main protocols for CTA (based on Syrian hamster embryonic primary cells [SHE], BALB/c 3T3 and C3H10T1/2 rodent cell lines) (OECD, 2007). This review concluded that the performance of the SHE and BALB/c 3T3 assays were sufficiently adequate and should be developed into OECD test guidelines. A pre-validation study including two SHE protocols (at pH 6.7 and pH 7.0) and the BALB/c 3T3 protocol was organised by ECVAM to address issues of standardisation of protocols, transferability and reproducibility. The data demonstrated that SHE standardised protocols are available and the assay systems themselves are transferable between laboratories, and are reproducible within and between laboratories. For the BALB/c 3T3 method an improved protocol has been developed, however further testing of this protocol was recommended to confirm its robustness (Vanparys et al., 2010). An ECVAM recommendation on cell transformation assays is currently in preparation.

8.2.4. Toxicogenomics

Toxicogenomics is based on the use of global gene expression data to identify expression changes associated with a toxicological outcome including carcinogenicity and genotoxicity. In the context of genotoxicity testing, its primary use is envisaged to be in providing information on mode of action and such information can be useful supporting evidence. However, it does not replace the need for genotoxicity testing.

The application of toxicogenomics to predict mode of action has been recently reviewed in depth (Ellinger-Ziegelbauer et al., 2009; Waters et al., 2010). Although the published *in vitro* and *in vivo* data set show appreciable variability, common features emerge with respect to molecular pathways. For instance, the DNA damage-responsive *p53* pathway is extensively activated both by DNA reactive genotoxins *in vitro* and genotoxic carcinogens *in vivo*. Conversely, *in vitro* DNA non-reactive genotoxins and *in vivo* non-genotoxic carcinogens mostly induce an oxidative stress response, signalling and cell cycle progression genes. These data represent a first proof of concept that the gene expression profiles reflect the underlying mechanism of action quite well. However, additional studies should be performed to enlarge the number of chemicals tested, to fill the gaps in dose-response and time-course relationship and in the case of *in vivo* toxicogenomics to analyse different routes of exposure and organ systems (most studies so far have used rat liver) and other species.

8.3. Epigenetics

Epigenetics is the occurrence of changes in phenotype as a result of changes in gene expression which persist through cell division into the daughter cell and which are not a consequence of a change in DNA sequence. One postulated mechanism is changes in the methylation of the cytosine base at CpG sites which may be maintained through gametogenesis and which results in gene silencing in the subsequent generation. Such changes have also been postulated to be inherited from generation to generation. Epigenetics may, therefore, provide heritable changes but unlike changes to the base pair sequence these are not permanent with the effect apparently diminishing over subsequent generations. This could be an explanation for observations of male-mediated abnormalities. This mechanism, therefore, shares some but not all the properties of genetic changes in terms of inter-generational events. There have been suggestions, however, that epigenetic changes could lead to irreversible changes in DNA sequence through, for instance, changing the mobility and insertion characteristics of transposable elements which can result in genetic rearrangements and mutational events. At present, there is not a strong evidence base for such a mechanism leading to permanent inherited changes but the research in this area should be monitored.

8.4. Use of Margin of Exposure (MOE) approach for *in vivo* genotoxicity

The 'no safe dose' concept of genotoxic carcinogens led to the risk management concept of ALARA (As Low As Reasonably Achievable) or ALARP (As Low As Reasonably Practical/Possible). However, the ubiquitous nature of genotoxic compounds in the environment from both natural and human-derived sources requires a method to evaluate the possible implication of unavoidable exposures to them.

The concept of the Margin of Exposure (MOE) was developed to try to address this issue and provide a comparison between the observed data and the environmental level of interest. The aim is to help decide on acceptable or tolerable levels of exposure taking into account the risk management options available.

The MOE is defined as the dimensionless ratio of a chosen Point of Departure (POD) or Reference Dose (RD) such as the NOAEL (no-observable-adverse-effect level) or a dose that produces a specified effect, e.g. the benchmark dose (BMD), on a dose–response curve to an estimate of the expected human exposure or dose (MOE=POD/Exposure). Both EFSA (2005) and the Joint FAO/WHO Expert Committee on Food Additives (JECFA) (FAO/WHO, 2006) have proposed the use of the $BMDL_{10}$ (the lower confidence limit on a benchmark dose giving a 10% response) as the POD

for the calculation of MOEs for genotoxic carcinogens. The smaller the dose from exposure, the larger is the margin of exposure. An MOE can be calculated for any specified response but the MOE is not a quantitative measure of risk.

The MOE approach has been considered by various international groups and advisory bodies as a tool for prioritorizing and for risk assessment. One proposal is that the maximum upper limit for the margin of exposure for carcinogenicity might be 10,000 (Gaylor, 1999; Gold et al., 2003). EFSA (2005) said that an MOE greater than 10,000 relative to the carcinogenic BMDL10 would be of "low concern" for genotoxic carcinogens.

As an MOE can, in theory, be calculated for any specific quantitative response, genotoxicity data could be used for the calculation of MOEs for genotoxicity endpoints. However, it has been customary to consider the use of genotoxicity testing as a hazard identification phase with the object of categorizing a chemical as either genotoxic or non-genotoxic. Quantitative assessment of the *in vivo* dose-response relationship or of measures of potency is not routinely used in assessments. There has, though, been increasing interest in the development of methods to try to identify thresholds for genotoxic substances (see section 8.1) and to characterise the dose-response relationships at low doses. Such approaches might be compatible with the development of MOE approaches for genotoxicity data especially if the collection of genotoxicity data became integrated into the standard toxicity tests.

8.5. Work ongoing in other groups

The Scientific Committee has considered recent and likely future developments in the area of genotoxicity testing from work being undertaken by other national and international groups of experts. These activities are summarised in Appendix D.

CONCLUSIONS AND RECOMMENDATIONS

The Scientific Committee has reviewed the state-of-the-science on genotoxicity testing strategies, bearing in mind the needs of EFSA's various scientific panels to have appropriate data for risk assessment. The Scientific Committee has considered relevant publications, including those from a number of international groups of experts, which focus on optimisation of basic test batteries and follow-up of indications of genotoxicity observed in basic test batteries.

The purpose of genotoxicity testing for risk assessment of substances in food and feed is:

- to identify substances which could cause heritable damage in humans,
- to predict potential genotoxic carcinogens in cases where carcinogenicity data are not available, and
- to contribute to understanding of the mechanism of action of chemical carcinogens.

For an adequate evaluation of the genotoxic potential of a chemical substance, different end-points, i.e. induction of gene mutations, structural and numerical chromosomal alterations, need to be assessed, as each of these events has been implicated in carcinogenesis and heritable diseases. An adequate coverage of all the above mentioned end-points can only be obtained by the use of more than one test system, as no individual test can simultaneously provide information on all these end-points.

In reaching its recommendations for a basic test battery, the Scientific Committee has considered:

- past experience with various tests when combined in a basic battery
- the availability of guidelines or internationally accepted protocols
- the performance of *in vitro* and *in vivo* tests in prediction of rodent carcinogenesis,
- correlations between *in vitro* and *in vivo* positive results for genotoxicity,
- the minimum number of tests necessary to achieve adequate coverage of the three required endpoints, and
- the need to avoid unnecessary animal tests.

The Scientific Committee recommends a step-wise approach for the generation and evaluation of data on genotoxic potential, comprising:

- a basic battery of *in vitro* tests,
- consideration of whether specific features of the test substance might require substitution of one or more of the recommended *in vitro* tests by other *in vitro* or *in vivo* tests in the basic battery,
- in the event of positive results from the basic battery, review of all the available relevant data on the test substance, and
- where necessary, conduct of an appropriate *in vivo* study (or studies) to assess whether the genotoxic potential observed *in vitro* is expressed *in vivo.*

Recommendations for the basic test battery

The Scientific Committee recommends use of the following two *in vitro* tests as the first step in testing:

- a bacterial reverse mutation assay (OECD TG 471), and
- an *in vitro* micronucleus test (OECD TG 487).

This combination of tests fulfils the basic requirements to cover the three genetic endpoints with the minimum number of tests; the bacterial reverse mutation assay covers gene mutations and the *in vitro* micronucleus test covers both structural and numerical chromosome aberrations. The Scientific Committee concluded that these two tests are reliable for detection of most potential genotoxic substances and that the addition of any further *in vitro* mammalian cell tests in the basic battery would significantly reduce specificity with no substantial gain in sensitivity.

Concerning the magnitude of the concentrations of test substance used in *in vitro* tests on mammalian cells, the Scientific Committee is aware that many consider that, for the majority of cases the top concentration of 10 mM recommended in current OECD guidelines is too high. However, there is a need to evaluate further data and to reach international consensus on this issue. Until this issue is resolved, the Scientific Committee recommends that EFSA Panels should use a weight-of-evidence approach to reach a decision on whether a substance that is positive only at a high concentration is indeed a relevant positive.

The Scientific Committee did consider whether the extent of human exposure (e.g. high or lifetime exposures) to substances should influence the number and type of tests to be included in a basic battery. It was concluded that, provided the basic battery of *in vitro* tests adequately assesses the potential for genotoxicity of a substance covering all three critical endpoints, then the level or duration of human exposure is not by itself the sole consideration.

The Scientific Committee also considered whether an *in vivo* test should be included in the first step of testing and broadly agreed that it should not be routinely included. However, if there are indications for the substance of interest that specific metabolic pathways would be lacking in the standard *in vitro* systems, or it is known that the *in vitro* test system is inappropriate for that substance or for its mode of action, testing may require either appropriate modification of the *in vitro* tests or use of an *in vivo* test at an early stage of testing. The Scientific Committee also recognised that in some cases it may be advantageous to include *in vivo* assessment of genotoxicity at an early stage, if, for example, such testing can be incorporated within other repeated-dose toxicity studies that will be conducted anyway.

In the case of positive results from the basic battery of tests, it may be that further testing *in vitro* is appropriate to optimise any subsequent *in vivo* testing, or to provide additional useful mechanistic information.

In cases where all *in vitro* endpoints are clearly negative in adequately conducted tests, it can be concluded with reasonable certainty that the substance is not a genotoxic hazard.

In the case of inconclusive, contradictory or equivocal results from *in vitro* testing, it may be appropriate to conduct further testing *in vitro*, either by repetition of a test already conducted, perhaps under different conditions, or by conduct of a different *in vitro* test, to try to resolve the situation.

Recommendations on follow-up of results from the basic battery

Before embarking on any necessary follow-up of positive *in vitro* results by *in vivo* testing, not only the results from the *in vitro* testing should be reviewed, but also other relevant data on the substance, such as information about chemical reactivity of the substance (which might predispose to site of contact effects), bioavailability, metabolism, toxicokinetics, and any target organ specificity. Additional useful information may come from structural alerts and 'read-across' from structurally related substances. It may be possible after this to reach a conclusion to treat the substance as an *in vivo* genotoxin. If, after such a review, a decision is taken that *in vivo* testing is necessary, tests should be selected on a case-by-case basis using expert judgement, with flexibility in the choice of test, guided by the full data set available for the substance.

In vivo tests should relate to the genotoxic endpoint(s) identified as positive *in vitro* and to appropriate target organs or tissues. Evidence, either from the test itself or from other toxicokinetic or repeated-

dose toxicological studies, that the target tissue(s) have been exposed to the test substance and/or its metabolites is essential for interpretation of negative results.

The approach to *in vivo* testing should be step-wise. If the first test is positive, no further test is needed and the substance should be considered as an *in vivo* genotoxin. If the test is negative, it may be possible to conclude that the substance is not an *in vivo* genotoxin. However, in some cases, a second *in vivo* test may be necessary as there are situations where more than one endpoint in the *in vitro* tests is positive and an *in vivo* test on a second endpoint may then be necessary if the first test is negative. It may also be necessary to conduct a further *in vivo* test on an alternative tissue if, for example, it becomes apparent that the substance did not reach the target tissue in the first test. The combination of assessing different endpoints in different tissues in the same animal *in vivo* should be considered.

The Scientific Committee recommends the following as suitable *in vivo* tests:

- an *in vivo* micronucleus test (OECD TG 474),
- an *in vivo* Comet assay (no OECD TG at present; internationally agreed protocols available, e.g. see hptt://cometassay.com), and
- a transgenic rodent assay (draft OECD TG; OECD, 2010b).

The *in vivo* micronucleus test covers the endpoints of structural and numerical chromosomal aberrations and is an appropriate follow up for *in vitro* clastogens and aneugens. The current OECD Test Guideline only considers peripheral blood and bone marrow as target tissues. There may be circumstances in which an *in vivo* mammalian bone marrow chromosome aberration test (OECD TG 475) may be a alternative follow up test.

The *in vivo* Comet assay is considered a useful indicator test in terms of its sensitivity to substances which cause gene mutations and/or structural chromosomal aberrations *in vitro*. It can be performed with many tissues. Transgenic rodent assays can detect point mutations and small deletions and are without tissue restrictions. More detailed advice on strategies for *in vivo* follow up is given in the main body of the opinion.

The Scientific Committee concluded that routine testing for genotoxicity in germ cells is not necessary. A substance that is concluded to be positive in tests in somatic tissues *in vivo* would normally be assumed to reach the germ cells and to be a germ cell mutagen, and therefore potentially hazardous to future generations. In the contrary situation, a substance that is negative in tests in somatic tissues *in vivo* would be assumed to be negative in germ cells, and moreover no germ cell-specific mutagen is known.

Normally, if the results of appropriate and adequately conducted *in vivo* tests are negative, then it can be concluded that the substance is not an *in vivo* genotoxin. If the results of the *in vivo* test(s) are positive, then it can be concluded that the substance is an *in vivo* genotoxin.

Other considerations

The Scientific Committee considered whether genotoxicity data would always be necessary for substances in food and feed for which human exposures are very low and whether, instead, the TTC approach might be helpful in assessing the likelihood of carcinogenic or transmissible genotoxic effects. Low-exposure substances within the EFSA remit include contaminants, and impurities, metabolites and degradation products of deliberately added substances, for which genotoxicity data may be unavailable. The Scientific Committee anticipates that it will adopt an opinion on the use of the TTC approach by the end of 2011.

Interpretation of data

The Scientific Committee recommends a documented weight-of-evidence approach to the evaluation and interpretation of genotoxicity data. Such an approach should not only consider the quality and reliability of the data on genotoxicity itself, but also take into account other relevant data that may be available, such as physico-chemical characteristics, structure-activity relationships (including structural alerts for genotoxicity and 'read-across' from structurally related substances), ADME, and the outcomes of any repeated-dose toxicity and carcinogenicity studies. The use of all the available relevant data is critical to reaching a sound conclusion on genotoxic potential as well as assisting in the design of genotoxicity studies and decision-making on the strategy for follow-up of positive or equivocal results from testing in a basic battery.

The Scientific Committee recognises that EFSA will continue to receive datasets that differ from the testing strategy recommended in this opinion. Such datasets should be considered on a case-by-case basis. Provided that the three critical endpoints (i.e. gene mutation, structural and numerical chromosomal aberration) have been adequately investigated, such datasets may be considered acceptable. The Scientific Committee recognises that in other cases where there is a heterogeneous dataset, EFSA has to rely on a weight-of-evidence approach.

Ongoing developments

The Scientific Committee is aware of a number of ongoing developments in genotoxicity test methods and in testing strategies that are being undertaken by other national and international groups of experts. The Scientific Committee recommends that these developments be followed and, if appropriate, the recommendations in this opinion be reviewed.

References (to be completed)

Asanami S, Shimono K, 1997. Hypothermia induces micronuclei in mouse bone marrow cells. Mutat. Res. 393, 91-98.

Asanami S, Shimono K, Kaneda S, 1998. Transient hypothermia induces micronuclei in mice. Mutat. Res. 413, 7-14.

Ashby J, 1985. Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. Environ. Mutagenesis 7, 919–921.

Ashby J, 1986. Carcinogen/mutagen screening strategies. Mutagenesis 1, 309-317.

Ashby J, Tinwell H, Callander RD, 1990. Activity of urethane and N,N-dimethylurethane in the mouse bone-marrow micronucleus assay: equivalence of oral and intraperitoneal routes of exposure. Mutat. Res. 245, 227-230.

Bergman K, Müller L, Teigen SW, 1996. Series: current issues in mutagenesis and carcinogenesis, No. 65. The genotoxicity and carcinogenicity of paracetamol: a regulatory (re)view. Mutat. Res. 349, 263-288.

Birrell L, Cahill P, Hughes C, Tate M, Walmsley RM, 2010. GADD45a-GFP GreenScreen HC assay results for the ECVAM recommended lists of genotoxic and non-genotoxic chemicals for assessment of new genotoxicity tests. Mutat. Res. 695, 87-95.

Burke DA, Wedd DJ, Herriott D, Bayliss MK, Spalding DJ, Wilcox P, 1994. Evaluation of pyrazole and ethanol induced S9 fraction in bacterial mutagenicity testing, Mutagenesis 9, 23–29.

Burlinson B, Tice RR, Speit G, Agurell E, Brendler-Schwaab SY, Collins AR, Escobar P, Honma M, Kumaravel TS, Nakajima M, Sasaki YF, Thybaud V, Uno Y, Vasquez M, Hartmann A, 2007. Fourth International Workgroup on Genotoxicity testing: results of the *in vivo* Comet assay workgroup. *In vivo* Comet Assay Workgroup, part of the Fourth International Workgroup on Genotoxicity Testing. Mutat. Res.627, 31-35.

Cimino MC, 2006. Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. Environ. Mol. Mutagen. 47, 362-390.

Cimoli G, Malacarne D, Ponassi R, Valenti M, Alberti S, Parodi S, 2004. Meta-analysis of the role of p53 status in isogenic systems tested for sensitivity to cytotoxic antineoplastic drugs. Biochim. Biophys. Acta. 1705, 103-120.

Clare MG, Lorenzon G, Akhurst LC, Marzin D, van Delft J, Montero R, Botta A, Bertens A, Cinelli S, Thybaud V, Lorge E, 2006. SFTG international collaborative study on in vitro micronucleus test II. Using human lymphocytes. Mutat. Res. 607, 37-60.

Claxton LD, Houk VS, Warren S, 2001. Methods for the spiral Salmonella mutagenicity assay including specialized applications. Mutat. Res. 488, 241-257.

Cole RJ, Taylor N, Cole J, Arlett CF, 1981. Short-term tests for transplacentally active carcinogens. I. Micronucleus formation in fetal and maternal mouse erythroblasts. Mutat. Res. 80, 141-157.

COM, 2000. Guidance on a Strategy for Testing of Chemicals for Mutagenicity. Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment (COM). Department of Health, London.

Cimoli G, Malacarne D, Ponassi R, Valenti M, Alberti S, Parodi S, 2004. Meta-analysis of the role of p53 status in isogenic systems tested for sensitivity to cytotoxic antineoplastic drugs. Biochim. Biophys. Acta, 1705: 103-120.

2112 Corvi R, Albertini S, Hartung T, Hoffmann S, Maurici D, Pfuhler S, van Benthem J, Vanparys P,
2113 2008. ECVAM retrospective validation of the in vitro micronucleus test (MNT). Mutagenesis,
2114 271, 271-283.

2115 Dearfield KL, Thybaud V, Cimino MC, Custer L, Czich A, Harvey JS, Hester S, Kim JH, Kirkland D,
2116 Levy DD, Lorge E, Moore MM, Ouédraogo-Arras G, Schuler M, Suter W, Sweder K, Tarlo K, van
2117 Benthem J, van Goethem F, Witt KL, 2011. Follow-up actions from positive results of in vitro
2118 genetic toxicity testing. Environ. Mol. Mutagen. 52: 177-204.

2119 De Boeck M, van der Leede BJ, Van Goethem F, De Smedt A, Steemans M, Lampo A, Vanparys P,
2120 2005. Flow cytometric analysis of micronucleated reticulocytes: Time- and dose-dependent
2121 response of known mutagens in mice, using multiple blood sampling. Environ. Mol. Mutagen. 46,
2122 30-42.

2123 De Flora S, Izzotti A, 2007. Mutagenesis and cardiovascular diseases: molecular mechanisms, risk
2124 factors, and protective factors. Mutat. Res, 621, 5–17.

2125 De Flora S., 2000. Threshold mechanisms and site specificity in chromium(VI) carcinogenesis.
2126 Carcinogenesis 21, 533-541.

2127 Destici E, Oklejewicz M, Nijman R, Tamanini F, van der Horst GT, 2009. Impact of the circadian
2128 clock on in vitro genotoxic risk assessment assays. Mutat. Res. 680, 87-94.

2129 Dobrovolsky VN, Miura D, Heflich RH, Dertinger SD, 2010. The in vivo Pig-a gene mutation assay, a
2130 potential tool for regulatory safety assessment. Env. Mol. Mutagen. 51, 825-835.

2131 Doak SH, Jenkins GJ, Johnson GE, Quick E, Parry EM, Parry JM, 2007. Mechanistic influences for
2132 mutation induction curves after exposure to DNA-reactive carcinogens. Cancer Res. 67, 3904-
2133 3911.

2134 Donato MT, Lahoz A, Castell JV, Gómez-Lechón MJ, 2008. Cell lines: a tool for in vitro drug
2135 metabolism studies. Curr. Drug Metab. 9, 1-11.

2136 Eastmond DA, Hartwig A, Anderson D, Anwar W, Cimino MC, Dobrev I, Douglas GR, Nohmi T,
2137 Phillips DH, Vickers C, 2009. Mutagenicity testing for chemical risk assessment: update of the
2138 WHO/IPCS Harmonized Scheme. Mutagenesis 24, 341-349.

2139 ECETOC, 1997. Aneuploidy. ECETOC Monograph No.27, August 1997. European Centre for
2140 Ecotixicology and Toxicology of Chemicals, Brussels.

2141 ECHA, 2008a. Guidance on information requirements and chemical safety assessment Chapter R.4:
2142 Evaluation of available information. Available at:
2143 http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r4_en.pdf?vers
2144 =20_08_08. ECHA, 2008b. Guidance for the Implementation of REACH. Guidance on information
2145 requirements and chemical safety assessment. Chapter R.7.a: Endpoint specific guidance. Section
2146 R.7.7 Mutagenicity and carcinogenicity, p377. European Chemicals Agency, Helsinki. Available
2147 at:
2148 http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7a_en.pdf?ver
2149 s=02_02_10. EFSA, 2005. Opinion of the Scientific Committee on a request from EFSA related to
2150 A Harmonised Approach for Risk Assessment of Substances Which are both Genotoxic and
2151 Carcinogenic. The EFSA Journal 282, 1-31. Available at:
2152 http://www.efsa.europa.eu/en/efsajournal/pub/282.htm.

2153 EFSA, 2007. Opinion of the Scientific Panel PPR related to the revision of Annexes II and III to
2154 Council Directive 91/414/EEC concerning the placing of plant protection products on the market -
2155 Toxicological and metabolism studies. The EFSA Journal 449, 1 – 60. Available at:
2156 http://www.efsa.europa.eu/en/efsajournal/doc/449.pdf.

2157 EFSA, 2008. Technical Guidance for establishing the safety of additives for the consumer prepared
2158 by the Panel on Additives and Products or Substances used in Animal Feed. Available at:
2159 http://www.efsa.europa.eu/en/efsajournal/doc/801.pdf.

EFSA, 2009. Scientific Opinion of the Scientific Committee on Existing approaches incorporating replacement, reduction and refinement of animal testing: applicability in food and feed risk assessment. The EFSA Journal 1052, 1077. Available at: http://www.efsa.europa.eu/en/efsajournal/doc/1052.pdf.

Elhajouji A, Van Hummelen P, Kirsch-Volders M, 1995. Indications for a threshold of chemically-induced aneuploidy in vitro in human lymphocytes. Environ. Mol. Mutagen. 26, 292-304.

Elhajouji A, Tibaldi F, Kirsch-Volders M, 1997. Indication for thresholds of chromosome non-disjunction versus chromosome lagging induced by spindle inhibitors in vitro in human lymphocytes. Mutagenesis 12, 133-140.

Ellinger-Ziegelbauer H Aubrecht J, Kleinjans JC, Ahr HJ, 2009. Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. Toxicol. Letters 186, 36-44.

Ennever FK, Rosenkranz HS, 1986. Short-term test results for NTP noncarcinogens: an alternate, more predictive battery. Environ Mutagen. 8, 849-865.

EU, 2008. Council Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Official Journal of the European Union, L142, 1-739. 31.05.2008.

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

FAO/WHO, 2006. Evaluation of Certain Food Contaminants. Sixty-fourth Report of the Joint FAO/WHO Expert Committee on Food Additives. World Health Organization, Geneva. Available at: http://whqlibdoc.who.int/trs/WHO_TRS_930_eng.pdf.

Fernando RC, Nair J, Barbin A, Miller JA, Bartsch JA, 1996. Detection of 1,N6-ethenodeoxyadenosine and 3,N4-ethenodeoxycytidine by immunoaffinity/32P-postlabelling in liver and lung DNA of mice treated with ethyl carbamate (urethane) or it's metabolites, Carcinogenesis 17, 1711–1718.

Flückiger-Isler S, Baumeister M, Braun K, Gervais V, Hasler-Nguyen N, Reimann R, Van Gompel J, Wunderlich HG, Engelhardt G, 2004. Assessment of the performance of the Ames II assay: a collaborative study with 19 coded compounds. Mutat. Res. 558, 181-197.

Forkert PG, Lee RP, 1997. Metabolism of ethyl carbamate by pulmonary cytochrome P450 and carboxylesterase isozymes: involvement of CYP2E1 and hydrolase A, Toxicol. Appl. Pharmacol. 146, 245–254.

Fowler P, Smith K, Young J, Jeffrey L, Kirkland D, Pfuhler S, Carmichael P, 2011. Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. I. Choice of cell type. Publication submitted.

Frank SA, 2010. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. Proc. Natl. Acad. Sci. U S A. 107 Suppl 1, 1725-1730.

Gatehouse DG, Paes DJ, 1983. A demonstration of the in vitro bacterial mutagenicity of procarbazine, using the microtitre fluctuation test and large concentrations of S9 fraction. Carcinogenesis 4, 347-352.

Gaylor DW, Kodell RL, Chen JJ, Krewski D, 1999. A unified approach to risk assessment for cancer and noncancer endpoints based on benchmark doses and uncertainty/safety factors. Regul. Toxicol. Pharmacol. 29 (2, Pt 1), 151-157.

Gocke E, Müller L, 2009. *In vivo* studies in the mouse to define a threshold for the genotoxicity of EMS and ENU. Mutat. Res. 678, 101-107.

Gold LS, Gaylor DW, Slone TH, 2003. Comparison of cancer risk estimates based on a variety of risk assessment methodologies. Regul. Toxicol. Pharmacol. 37, 45-53.

Goodman and Gilman, 2002. The Pharmacological Basis of Therapeutics, 10th Edition, New York, McGraw-Hill.

Hartmann A, Agurell E, Beevers C, Brendler-Schwaab S, Burlinson B, Clay P, Collins A, Smith A, Speit G, Thybaud V, Tice RR; 4th International Comet Assay Workshop. 2003. Recommendations for conducting the in vivo alkaline Comet assay. 4th International Comet Assay Workshop. Mutagenesis. 2003 Jan;18(1):45-51.

Hastwell PW, Chai LL, Roberts KJ, Webster TW, Harvey JS, Rees RW, Walmsley RM., 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.

Heddle JA, Dean S, Nohmi T, Boerrigter M, Casciano D, Douglas GR, Glickman BW, Gorelick NJ, Mirsalis JC, Martus HJ, Skopek TR, Thybaud V, Tindall KR, Yajima N, 2000. In vivo transgenic mutation assays. Environ. Mol. Mutagen. 35, 253-259.

Hoeijmakers JH, 2009. DNA damage, aging, and cancer. New Engl. J. Med. 361, 1475-1485.

ICH, 2010. ICHS2(R1): Genotoxicity: Testing and Data Interpretation for Pharmaceuticals Intended for Human Use. Step 2 Document, 2008. Available at: http://www.ich.org/products/guidelines/safety/article/safety-guidelines.html.
http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S2_R1/Step2/S2_R1__Guideline.pdf

ILSI, 2008. HESI and Health Canada collaborate on toxicity testing. Ilnternatinal Life Sciences Institute, Washington DC. ILSI News, 26, 2-3.

Jenkins GJ, Zair Z, Johnson GE, Doak SH, 2010. Genotixic thresholds, DNA repair, and susceptibility in human populations. Toxicology 278, 305-310.

JRC 2010. Scientific Report submitted to EFSA. Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Available at: www.efsa.europa.eu/en/scdocs/scdoc/50e.htm.

Kaina B, Christmann M, Naumann S, Roos WP, 2007. MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. DNA Repair 6, 1079–1099.

Kasper P, Uno Y, Mauthe R, Asano N, Douglas G, Matthews E, Moore M, Mueller L, Nakajima M, Singer T, Speit G; IWGT Workgroup, 2007. Follow-up testing of rodent carcinogens not positive in the standard genotoxicity testing battery: IWGT workgroup report. Mutat. Res. 627, 106-116.

Kim BS, Margolin BH, 1999. Prediction of rodent carcinogenicity utilizing a battery of in vitro and in vivo genotoxicity tests. Environ Mol Mutagen. 34, 297-304.

Kirkland D, Aardema M, Henderson L , Müller L, 2005. Evaluation of the ability of a battery of 3 *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity, Mutat. Res. 584, 1-256.

Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F, Elhajouji A, Glatt H, Hastwell P, Hayashi M, Kasper P, Kirchner S, Lynch A, Marzin D, Maurici D, Meunier JR, Müller L, Nohynek G, Parry J, Parry E, Thybaud V, Tice R, van Benthem J, Vanparys P, White P, 2007a. How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop. Mutat. Res. 628, 31-55.

2251 Kirkland DJ, Hayashi M, Jacobson-Kram D, Kasper P, MacGregor JT, Müller L, Uno Y, 2007b. The
2252 International Workshops on Genotoxicity Testing (IWGT): history and achievements. Mutat. Res.
2253 627, 1-4.

2254 Kirkland D, Aardema M, Banduhn N, Carmichael P, Fautz R, Meunier J-R, Pfuhler S, 2007c. *In vitro*
2255 approaches to develop weight of evidence (WoE) and mode of action (MoA) discussions with
2256 positive *in vitro* genotoxicity results. Mutagenesis 22, 161-175.

2257 Kirkland D, Kasper P, Müller L, Corvi R, Speit G, 2008. Recommended lists of genotoxic and non-
2258 genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: A
2259 follow-up to an ECVAM workshop, Mutat. Res. 653, 99-108.

2260 Kirkland D, Speit G, 2008. Evaluation of the ability of a battery of three in vitro genotoxicity tests to
2261 discriminate rodent carcinogens and non-carcinogens III. Appropriate follow-up testing *in vivo*.
2262 Mutat Res.654, 114-1132.

2263 Kirkland D, Fowler P, 2010. Further analysis of Ames-negative rodent carcinogens that are only
2264 genotoxic in mammalian cells in vitro at concentrations exceeding 1 mM, including retesting of
2265 compounds of concern. Mutagenesis, 25, 539-553.

2266 Kirkland D, Reeve L, Gatehouse D, Vanparys P, 2011. A core in vitro genotoxicity battery comprising
2267 the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in
2268 vivo genotoxins. Mutat. Res. 721, 27-73.

2269 Klimisch HJ, Andreae M, Tillmann U, 1997. A systematic approach for evaluating the quality of
2270 experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. 25, 1-5.

2271 Ku WW, Bigger A, Brambilla G, Glatt H, Gocke E, Guzzie PJ, Hakura A, Honma M, Martus HJ,
2272 Obach RS, Roberts S; Strategy Expert Group, IWGT, 2007. Strategy for genotoxicity testing-
2273 Metabolic considerations. Mutat. Res. 627, 59-77.

2274 Lambert IB, Singer TM, Boucher SE, Douglas GR, 2005. Detailed review of transgenic rodent
2275 mutation assays. Mutat. Res. 590, 1-280.

2276 Lambert IB, Singer TM, Boucher SE, Douglas GR, 2008. Detailed review paper on transgenic rodent
2277 mutation assays. OECD Environmental, Health and Safety Publications, available at:
2278 http://www.oecd.org/dataoecd/53/35/40835004.pdf

2279 LeBoeuf RA, Kerckaert KA, Aardema MJ, Isfort RJ, 1999. Use of Syrian hamster embryo and
2280 BALB/c 3T3 cell transformation for assessing the carcinogenic potential of chemicals. IARC Sci.
2281 Publ. 146, 409-425.

2282 Lehman TA, Modali R, Boukamp P, Stanek J, Bennett WP, Welsh JA, Metcalf RA, Stampfer MR,
2283 Fusenig N, Rogan EM, et al. 1993. P53 mutations in human immortalized epithelial cell lines.
2284 Carcinogenesis 14, 833-839.

2285 Lorge E, Thybaud V, Aardema MJ, Oliver J, Wakata A, Lorenzon G, Marzin D, 2006. SFTG
2286 international collaborative study on in vitro micronucleus test I. General conditions and overall
2287 conclusions of the study. Mutat. Res. 607, 13-36

2288 Lynch AM, Sasaki JC, Elespuru R, Jacobson-Kram D, Thybaud V, De Boeck M, Aardema MJ,
2289 Aubrecht J, Benz RD, Dertinger SD, Douglas GR, White PA, Escobar PA, Fornace A Jr, Honma
2290 M, Naven RT, Rusling JF, Schiestl RH, Walmsley RM, Yamamura E, van Benthem J, Kim JH,
2291 2011. New and emerging technologies for genetic toxicity testing. Environ. Mol. Mutagen. 52,
2292 205-223.

2293 Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF, 2006. An analysis of genetic toxicity,
2294 reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens
2295 using surrogate endpoints. Regul. Toxicol. Pharmacol. 44, 83-96

2296 McCarroll N, Keshava N, Chen J, Akerman G, Kligerman A, Rinde E, 2010. An evaluation of the
2297 mode of action framework for mutagenic carcinogens case study II: chromium (VI). Environ. Mol.
2298 Mutagen. 51, 89-111.

2299 Mirsalis JC, Shimon JA, Johnson A, Fairchild D, Kanazawa N, Nguyen T, de Boer J, Glickman B,
2300 Winegar RA, 2005. Evaluation of mutant frequencies of chemically induced tumors and normal
2301 tissues in lambda/cII transgenic mice. Environ. Mol. Mutagen. 45, 17-35.

2302 Moore MM, Honma M, Clements J, Bolcsfoldi G, Burlinson B, Cifone M, Clarke J, Clay P,
2303 Doppalapudi R, Fellows M, Gollapudi B, Hou S, Jenkinson P, Muster W, Pant K, Kidd DA, Lorge
2304 E, Lloyd M, Myhr B, O'Donovan M, Riach C, Stankowski LF Jr, Thakur AK, Van Goethem F;
2305 Mouse Lymphoma Assay Workgroup, IWGT, 2007. Mouse lymphoma thymidine kinase gene
2306 mutation assay: meeting of the International Workshop on Genotoxicity Testing, San Francisco,
2307 2005, recommendations for 24-h treatment. Mutat. Res. 627, 36-40.

2308 Morris SM, 2002. A role for p53 in the frequency and mechanism of mutation. Mutat Res. 511, 45-62.

2309 Müller L, Blakey D, Dearfield KL, Galloway S, Guzzie P, Hayashi M, Kasper P, Kirkland D,
2310 MacGregor JT, Parry JM, Schechtman L, Smith A, Tanaka N, Tweats D, Yamasaki H: IWGT
2311 Expert Group, 2003. Strategy for genotoxicity testing and stratification of genotoxicity test results--
2312 report on initial activities of the IWGT Expert Group. Mutat. Res. 540, 177-181.

2313 OECD, 1997. OECD Guidelines for the Testing of Chemicals – Test Guideline 474, Mammalian
2314 Erythrocyte Micronucleus Test (adopted 21st July,1997). Organisation for Economic Cooperation
2315 and Development, Paris. Available at:
2316 http://lysander.sourceoecd.org/vl=1154212/cl=15/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/v1
2317 n4/contp1-1.htm.

2318 OECD, 2007. Detailed review paper on cell transformation assays for detection of chemical
2319 carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and
2320 Assessment, No. 31. Available at: http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-
2321 mono(2007)18.

2322 OECD, 2009. Detailed review paper on transgenic rodent mutation assays, series on testing and
2323 assessment number 103, OECD, Paris, July 23, 2009. Available at: http://
2324 www.olis.oecd.org/oils/2009doc.nsf/linkto/ENV-JM-MONO(2009)7.

2325 OECD, 2010a. OECD Guideline for the Testing of Chemicals: In Vitro Mammalian Cell
2326 Micronucleus Test. Available at:
2327 http://lysander.sourceoecd.org/vl=1154212/cl=15/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/v1
2328 n4/contp1-1.htm.

2329 OECD, 2010b. OECD Guideline for the Testing of Chemicals: Draft Proposal for a New Guideline
2330 XXX: Transgenic Rodent Gene Mutation Assays. Available at:
2331 http://www.oecd.org/dataoecd/15/27/46777174.pdf.

2332 OECD 2010c. OECD Guideline for the Testing Of Chemicals: Toxicokinetics (417) Adopted 22 July
2333 2010.Available at: http://www.oecd-ilibrary.org/environment/test-no-417-
2334 toxicokinetics_9789264070882-en.

2335 Ogawa I, Furukawa S, Abe M, Tanaka Y, Hayashi S, Usuda K, 2009. Multi-endpoint genotoxic assay
2336 using L5178Y ($Tk^{+/-}$ -3.7.2c) cells. J. Toxicol. Sci. 34, 547-553.

2337 Parry JM, Fielder RJ, McDonald A, 1994. Thresholds for aneuploidy-inducing chemicals. Advisory
2338 Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment of the
2339 UK Department of Health. Mutagenesis 9, 503-504.

2340 Parry JM, Parry E, Phrakonkham P, Corvi R, 2010. Analysis of published data for top concentration
2341 considerations in mammalian cell genotoxicity testing. Mutagenesis, 25, 531-538.

2342 Paz-Elizur T, Elinger D, Leitner-Dagan Y, Blumenstein S, Krupsky M, Berrebi A, Schechtman E,
2343 Livneh Z, 2007. Development of an enzymatic DNA repair assay for molecular epidemiology
2344 studies: Distribution of OGG activity in healthy individuals DNA Repair 6, 45–60.

2345 Pfuhler S, Albertini S, Fautz R, Herbold B, Madle S, Utesch D, Poth A, 2007. Genetic toxicity
2346 assessment: employing the best science for human safety evaluation part IV: Recommendation of a

working group of the Gesellschaft fuer Umwelt-Mutationsforschung (GUM) for a simple and straightforward approach to genotoxicity testing. Toxicol. Sci. 97, 237-240.

Pfuhler S, Kirkland D, Kasper P, Hayashi M, Vanparys P, Carmichael P, Dertinger S, Eastmond D, Elhajouji A, Krul C, Rothfuss A, Schoening G, Smith A, Speit G, Thomas C, van Benthem J, Corvi R, 2009. Reduction of Use of Animals in Regulatory Genotoxicity Testing: Identification and Implementation Opportunities - Report from an ECVAM Workshop. Mutat. Res. 680, 31- 42.

Pfuhler S, Kirst A, Aardema M, Banduhn N, Goebel C, Araki D, Costabel-Farkas M, Dufour E, Fautz R, Harvey J, Hewitt NJ, Hibatallah J, Carmichael P, Macfarlane M, Reisinger K, Rowland J, Schellauf F, Schepky A, Scheel J, 2010. A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: genotoxicity. A COLIPA analysis. Regul. Toxicol. Pharmacol. 57, 315-324. Erratum in: Regul. Toxicol. Pharmacol. 58, 544.

Platel A, Nesslany F, Gervais V, Marzin D, 2009. Study of oxidative DNA damage in TK6 human lymphoblastoid cells by use of the in vitro micronucleus test: Determination of No-Observed-Effect Levels. Mutat Res. 678, 30-37.

Povey AC, Margison GP, Santibanez-Koref MF, 2007. Lung cancer risk and variation in MGMT activity and sequence. DNA Repair 6, 1134–1144.

Rothfuss A, O'Donovan M, De Boeck M, Brault D, Czich A, Custer L, Hamada S, Plappert-Helbig U, Hayashi M, Howe J, Kraynak A, van der Leede B, Nakajima M, Priestley C, Thybaud V, Saigo K, Sawant S, Shi J, Storer R, Struwe M, Vock E, Galloway S, 2010. Collaborative study on 15 compounds in the rat liver Comet Assay integrated into 2- and 4-week repeat-dose studies. Mutat. Res. 702, 40-69.

Russell WMS, Burch RL, 1959. The Principles of Humane Experimental Technique. Methuen, London, 1959. Reprinted by UFAW, 1992: 8 Hamilton Close, South Mimms, Potters Bar, Herts EN6 3QD England. ISBN 0 900767 78 2.

Sasaki YF, Saga A, Akasaka M, Ishibashi S, Yoshida K, Su YQ, Matsusaka N, Tsuda S, 1998. Organ-specific genotoxicity of the potent rodent colon carcinogen 1,2-dimethylhydrazine and three hydrazine derivatives: difference between intraperitoneal and oral administration. Mutat. Res. 415, 1-12.

Sasaki YF, Sekihashi K, Izumiyama F, Nishidate E, Saga A, Ishida K, Tsuda S, 2000. The comet assay with multiple mouse organs: comparison of comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database. Crit. Rev. Toxicol. 30, 629-799.

Scott D, Galloway SM, Marshall RR, Ishidate M, Brusick Jr D, Ashby J, Myhr BC, 1991. Genotoxicity under extreme culture conditions. A report from ICPEMC Task Group 9. Mutat. Res. 257, 147–204.

Serafimova R, Fuart Gatnik M & Worth A, 2010. Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity. EUR 24427 EN. JRC Report. Available at: http://ecb.jrc.ec.europa.eu/qsar/publications/.

Siakafas AR and Richardson DR, 2009, Growth arrest and DNA damage-45alpha (GADD45α) Int J Biochem. Cell Biol. 41, 986-989

Shuey DL, Gudi R, Krsmanovic L, Gerson RJ, 2007. Evidence that oxymorphone-induced increases in micronuclei occur secondary to hyperthermia. Toxicol. Sci. 95, 369-375

Slatter MA and Gennery AR, 2010. Primary immunodeficiencies associated with DNA-repair disorders. Expert Rev. Mol. Med. 12:e9. Published on-line 18 Mar 2010.

Takahashi M, Takeda J, Hirose S, Hyman R, Inoue N, Miyata T, Ueda E, Kitani T, Medof ME, Kinoshita T, 1993. Deficient biosynthesis of N-acetylglucosaminyl-phosphatidylinositol, the first intermediate of glycosyl phosphatidylinositol anchor biosynthesis, in cell lines established from patients with paroxysmal nocturnal hemoglobinuria. J. Exp. Med. 177, 517-521.

Thielmann HW, Popanda O, Gersbach H, Gilberg F, 1993. Various inhibitors of DNA topoisomerases diminish repair-specific DNA incision in UV-irradiated human fibroblasts. Carcinogenesis 14, 2341-2451.

Thybaud V, Dean S, Nohmi T, de Boer J, Douglas GR, Glickman BW, Gorelick NJ, Heddle JA, Heflich RH, Lambert I, Martus HJ, Mirsalis JC, Suzuki T, Yajima N, 2003. In vivo transgenic mutation assays. Mutat. Res. 540, 141-151.

Thybaud V, Aardema M, Clements J, Dearfield K, Galloway S, Hayashi M, Jacobson-Kram D, Kirkland D, MacGregor JT, Marzin D, Ohyama W, Schuler M, Suzuki H, Zeiger E, 2007a. Expert Working Group on Hazard Identification and Risk Assessment in Relation to In Vitro Testing. Strategy for genotoxicity testing: hazard identification and risk assessment in relation to in vitro testing. Mutat. Res. 627, 41-58.

Thybaud V, Aardema M, Casciano D, Dellarco V, Embry MR, Gollapudi BB, Hayashi M, Holsapple MP, Jacobson-Kram D, Kasper P, MacGregor JT, Rees R, 2007b. Relevance and follow-up of positive results in in vitro genetic toxicity assays: an ILSI-HESI initiative. Mutat. Res. 633, 67-79.

Thybaud V, Macgregor JT, Müller L, Crebelli R, Dearfield K, Douglas G, Farmer PB, Gocke E, Hayashi M, Lovell DP, Lutz WK, Marzin D, Moore M, Nohmi T, Phillips DH, Van Benthem J, 2010. Strategies in case of positive *in vivo* results in genotoxicity testing. Mutat. Res. 2010. [Epub ahead of print]. doi:10.1016/j.mrgentox.2010.09.002.

Tice RR, Agurell E, Anderson D, Burlinson B, Hartmann A, Kobayashi H, Miyamae Y, Rojas E, Ryu JC, Sasaki YF, 2000. Single cell gel/comet assay: guidelines for in vitro and in vivo genetic toxicology testing. Environ. Mol. Mutagen. 35, 206-221.

Torous DK, Dertinger SD, Hall NE, Tometsko CR, 2000. Enumeration of micronucleated reticulocytes in rat peripheral blood: a flow cytometric study. Mutat. Res. 465, 91-99.

Tweats DJ, Blakey D, Heflich RH, Jacobs A, Jacobsen SD, Morita T, Nohmi T, O'Donovan MR, Sasaki YF, Sofuni T, Tice R; IWGT Working Group, 2007a. Report of the IWGT working group on strategies and interpretation of regulatory *in vivo* tests I. Increases in micronucleated bone marrow cells in rodents that do not indicate genotoxic hazards. Mutat. Res. 627, 78-91.

Tweats DJ, Blakey D, Heflich RH, Jacobs A, Jacobsen SD, Morita T, Nohmi T, O'Donovan MR, Sasaki YF, Sofuni T, Tice R; IWGT Working Group, 2007b. Report of the IWGT working group on strategy/interpretation for regulatory *in vivo* tests II. Identification of *in vivo*-only positive compounds in the bone marrow micronucleus test. Mutat. Res. 627, 92-105.

Vanparys P, Corvi R, Aardema M, Gribaldo L, Hayashi M, Hoffmann S, Schechtman L, 2010. ECVAM Prevalidation of Three Cell Transformation Assays. ALTEX 27, 267-270.

Vian L, Bichet N, Gouy D, 1993. The in vitro micronucleus test on isolated human lymphocytes. Mutat. Res. 291, 93-102.

Wang J, Sawyer JR, Chen L, Chen T, Honma M, Mei N, Moore MM, 2009. The mouse lymphoma assay detects recombination, deletion, and aneuploidy. Toxicol. Sci. 109, 96-105.

Waters MD, Stack HF, Jackson MA, 1999. Short-term tests for defining mutagenic carcinogens. In: McGregor DB, Rice JM, Venitt S (eds). The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. IARC Sci. Publ. No. 146, pp.499-536, International Agency for Research on Cancer, Lyon.

Waters MD, Jackson M, Lea I, 2010. Characterizing and predicting carcinogenicity and mode of action using conventional and toxicogenomics methods. Mutat. Res, 705: 184-200.

Wakata A, Matsuoka A, Yamakage K, Yoshida J, Kubo K, Kobayashi K, Senjyu N, Itoh S, Miyajima H, Hamada S, Nishida S, Araki H, Yamamura E, Matsui A, Thybaud V, Lorenzon G, Marzin D,

Lorge E, 2006. SFTG international collaborative study on in vitro micronucleus test IV. Using CHL cells. Mutat. Res. 607, 88-124.

Westerink WM, Stevenson JC, Horbach GJ, Schoonen WG., 2010. The development of RAD51C, Cystatin A, p53 and Nrf2 luciferase-reporter assays in metabolically competent HepG2 cells for the assessment of mechanism-based genotoxicity and of oxidative stress in the early research phase of drug development. Mutat. Res. 696, 21-40.

Williams CV , Fletcher K, Tinwell H, Ashby J, 1998. Weak mutagenicity of ethyl carbamate to Lac-Z transgenic mice, Mutagenesis 13, 133–137.

APPENDICES

A. APPENDIX: GUIDANCE OR REQUIREMENTS OF EFSA PANELS FOR GENOTOXICITY TESTING

- **Panel on Food Additives and Nutrient Sources added to food (ANS):** For food additives, three tests are recommended for the assessment of genotoxicity: a bacterial mutagenicity test (Ames test), a mammalian gene mutation assay (preference for mouse lymphoma *tk*) and an *in vitro* chromosomal aberration assay. *In vivo* testing is required in case of positive *in vitro* results. The test should also cover the endpoint of aneuploidy (preferably by the *in vivo* MN test). In case of positive *in vivo* results in somatic cells, the need for *in vivo* studies at the germ cell level should be considered on a case-by-case basis. Explicitly, some test systems already accepted by other Panels but not by ANS, are mentioned in the guidance document as "future developments". This reflects the fact that the current guidance was drawn up some time ago. It mentions, for example, the *in vitro* MN test as an alternative to the *in vitro* chromosomal aberration assay. Also, the mouse lymphoma *tk* assay is not accepted as a surrogate for both gene and chromosomal mutation test. Furthermore, test procedures like the Comet assay or tissue specific mutations in transgenic animals are considered to provide useful information in the future, but since they are still under validation, it was recommended to use such tests with caution. The guidance document, originally adopted by the Scientific Committee on Food in 2000 (SCF, 2001) is at present under revision. The finalisation is expected later in 2011.

- **Panel on Food Contact Materials, Flavourings, Enzymes and Processing Aids (CEF):** This Panel evaluates food contact materials, food enzymes, flavourings in or on food, and smoke flavourings and currently has slightly different recommendations for each category.

    - *Food contact materials:* Testing requirements for food contact materials (mainly plastics) are three *in vitro* mutagenicity tests (bacterial mutagenicity test, mammalian mutagenicity test, mammalian chromosomal aberration test). In case of positive or equivocal results, further mutagenicity tests, including *in vivo* assays, may be required, decided case-by-case (EFSA, 2008b). A revision of the guidance document on test requirements is ongoing and it is planned to be adopted in summer 2011.
    - *Food enzymes:* Genotoxicity assessment requires two *in vitro* tests, one bacterial mutagenicity test or, if not applicable, a mammalian mutagenicity test (preferably mouse lymphoma *tk* with colony sizing). Further, an *in vitro* test for the detection of clastogenicity is required (chromosomal aberrations, micronuclei or mouse lymphoma *tk*). Follow-up of positive *in vitro* results is flexible (expert judgement on a case-by-case basis) and may include *in vivo* rodent bone marrow MN or mouse peripheral blood MN, rodent bone marrow clastogenicity, Comet assay, gene mutations in transgenic rodents, or rat liver UDS (EFSA, 2009).
    - *Flavourings to be used in or on foods:* Until recently, there have been no requirements for genotoxicity testing of flavourings, but in the future evaluations will be conducted, based on the EFSA "Guidance on data requirements for the risk assessment of flavourings to be used in or on foods" (EFSA, 2010). Three tests will be required, a bacterial mutagenicity test (Ames test), a mammalian gene mutation assay (preference for mouse lymphoma *tk*) and an *in vitro* chromosomal aberration or MN assay. This test battery does not consider the mouse lymphoma assay as an acceptable surrogate for tests for chromosomal damage. Follow-up of positive *in vitro* results should be selected from among cytogenetic tests in rodent erythropoeitic cells, Comet assay, gene mutations in transgenic rodents, or rat liver UDS. In general, the approach is flexible and testing can be omitted if previously evaluated and structurally related flavourings gave negative results. Similarly, follow up of positive *in vitro* results can be omitted if it can be adequately demonstrated that they are not relevant for the *in vivo* situation.

- *Smoke flavourings:* The same three test systems as for flavourings in general are proposed; however, there is no specific guidance on *in vivo* follow-up testing in case of positive *in vitro* results (EFSA, 2005).

- **Panel on Contaminants in the food chain CONTAM:** No specific guidance is given, since expert judgement is made on the basis of all available information.

- **Panel on Feed Additives and Products or Substances Used in Animal Feed (FEEDAP):** The guidance document specifies two *in vitro* tests, namely gene mutations, either in bacteria or in mammalian cells (preferably the mouse lymphoma tk assay), and chromosomal aberrations in mammalian cells. However, in addition the core test battery includes one *in vivo* test in a mammalian species, independent of the outcome of the *in vitro* tests. If one test gives a positive result, one further *in vivo* test is required (EFSA, 2008a). This guidance is also incorporated into EU legislation (EU, 2008).

- **Plant Protection Products and their Residues (PPR):** This Panel deals with plant protection products and their residues. Testing requirements are established by the European Commission and Member States and are included in the relevant EU Regulation (EC Directive 91/414, Annex II and Annex III)). Three *in vitro* tests are required (bacterial assay for gene mutation, combined tests for structural and numerical chromosome aberrations, and a test for gene mutations in mammalian cells); and at least one *in vivo* study must be done with demonstration of exposure (e.g. cell toxicity and/or toxicokinetic data).

  A new Regulation (Regulation (EC) 1107/2009 of the European Parliament and of the Council of 21 October 2009) will come in force on 14 June 2011. Revised Annexes II and III, including the data requirements, are not published until now. The PPR Panel was requested by the Commission to issue an opinion on the Commission Working Document on the data requirements for the new Regulation. The Panel suggested in its opinion (*The EFSA Journal* (2007) 449, 1 – 60) that there was no need for follow-up *in vivo* after negative *in vitro* results in the future. Follow-up of equivocal or positive *in vitro* results should be considered on a case-by-case basis, taking into account all relevant information and testing the same endpoint as in the positive *in vitro* test. In addition to guideline *in vivo* tests, the Comet assay could be applied in specific target tissues. Substances identified as *in vivo* somatic cell mutagens should be considered a germ cell mutagens as well, but in some cases the specific evaluation of mutations in germ cells may be appropriate. Within this test battery, the rodent dominant lethal assay is deleted in the future regulation. Altogether, the Panel recommends a rather flexible approach, especially for any *in vivo* testing.

References

EFSA, 2005. Guidance from the Scientific Panel on Food Additives Flavourings, Processing aids and Materials in Contact with Food Guidance on submission of a dossier on a Smoke Flavouring Primary Product for evaluation by EFSA. Available at: http://www.efsa.europa.eu/en/efsajournal/pub/492.htm.

EFSA, 2008a. Guidance for the preparation of dossiers for technical additives. EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP). The EFSA Journal (2008) 774, 1-21. Available at: http://www.efsa.europa.eu/en/efsajournal/pub/774.htm.

EFSA, 2008b. Guidance document on the submission of a dossier on a substance to be used in Food Contact Materials for evaluation by EFSA by the Panel on additives, flavourings, processing aids and materials in contact with food (AFC). The EFSA Journal, 31 July 2008. Available at: http://www.efsa.europa.eu/en/efsajournal/pub/21r.htm.

EFSA, 2009. Guidance of the Scientific Panel of Food Contact Material, Enzymes, Flavourings and Processing Aids (CEF) on the Submission of a Dossier on Food Enzymes for Safety Evaluation by the Scientific Panel of Food Contact Material, Enzymes, Flavourings and Processing Aids. Available at: http://www.efsa.europa.eu/en/efsajournal/pub/1305.htm.

EFSA, 2010. Guidance on data requirements for the risk assessment of flavourings to be used in or on foods. EFSA Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids The EFSA Journal 8(6), 1623, 1-53. Available at:http://www.efsa.europa.eu/en/scdocs/doc/1623.pdf.

EU, 2008. Commission Regulation (EC) No 429/2008 of 25 April 2008 on detailed rules for the implementation of Regulation (EC) No 1831/2003 of the European Parliament and of the Council as regards the preparation and the presentation of applications and the assessment and the authorisation of feed additives. Official Journal of the European Union L133, 1-65. Available at: http://ec.europa.eu/food/food/animalnutrition/feedadditives/guidelines_en.htm.

SCF, 2001. Guidance on submissions for food additive evaluations by the Scientific Committee on Food (opinion expressed on 11 July 2001). Available at: http://ec.europa.eu/food/fs/sc/scf/out98_en.pdf .

B. APPENDIX : ANALYSIS OF FOOD CONTACT MATERIALS DATABASE

Analysis of concordance between *in vitro* positives and *in vivo* positives for genotoxicity in data submitted to SCF/EFSA

An analysis of the concordance between *in vitro* and *in vivo* positives was performed inspecting the data submitted to the former Scientific Committee on Food (SCF) or to EFSA for approval of chemically defined food contact materials (FCM). This database consists of a homogeneous data set, with results from three *in vitro* mutagenicity tests (bacterial reverse mutation test, mouse lymphoma and/or HPRT gene mutation assay and structural chromosomal aberrations test) on all substances. A unique feature of this database is that all entries (i.e. positive and negative) are based on a careful analysis of raw data from GLP-compliant studies: thus the information provided can be considered highly reliable, even though based on a relatively small number of substances

Criteria for inclusion of substances in the analysis were as follows:

- Food contact materials (FCM) evaluated by the EFSA (period 2003-April 2010) or by the Scientific Committee on Food (period 1992-2002), for which a Summary Data Sheet was available for inspection of experimental data;
- Chemically defined substances or mixtures (undefined mixtures or high MW polymeric additives were excluded);
- Tested in three adequate *in vitro* genotoxicity studies, i.e. i) bacterial reverse mutation test (Ames test), ii) mammalian cell gene mutation test (either MLA, mouse lymphoma assay, or *hprt* assay), iii) chromosomal aberration assay (CA);

In total, **204** substances met the inclusion criteria.

Overview of results

The analysis of genotoxicity test results on FCM highlighted a relatively high overall incidence of *in vitro* positives:

147/204 (72.05%) negative in all three tests

57/204 (27.94%) positive in one or more tests

Equivocal results were considered as negative and weakly positive results considered as positive.

The highest proportion of positive results was detected by the *in vitro* chromosomal aberration assay (24%), followed by the mouse lymphoma assay (8.9%). The bacterial reverse mutation and *hprt* assays detected positive results in a small number of tests (2.4% and 2.8%, respectively).

Overall, there was low concordance between tests among the positive results, with the majority being "CA positives only":

	Ames positive *hprt* positive	**Ames positive *hprt* negative**	**Ames negative *hprt* positive**	**Ames negative *hprt* negative**		**Ames positive MLA positive**	**Ames positive MLA negative**	**Ames negative MLA positive**	**Ames negative MLA negative**
CA positive	1/30	0/30	1/30	**25/30**		0/27	0/27	8/27	**15/27**
CA negative	1/30	2/30	0/30	0/30		0	1/27	3/27	0/27

Ames: Ames test (*in vitro* gene mutation assay in bacteria); **MLA**: mouse lymphoma assay; **CA**: *in vitro* chromosome aberration test.

Follow-up *in vivo*

Fifty-one of the 57 *in vitro* positives were tested in one or more *in vivo* assays. Only three substances produced positive results *in vivo*; the remaining substances were completely negative.

The results of the follow-up *in vivo* on the 49 substances positive in the *in vitro* chromosomal aberration assay are shown below:

Result	MNviv in bone marrow	CA in bone marrow
Positive	2*	2*
Negative	42**	6**

** one substance positive in both MN and CA; ** two substances tested in both MNviv and CA*
MNviv: *in vitro* micronucleus test; **CA**: *in vitro* chromosome aberration test.

The results show that most *in vitro* clastogens were negative in cytogenetic tests *in vivo.* A word of caution is needed in the interpretation of these findings, as in principle the outcome of the follow-up *in vivo* testing of FCM could have been biased by the lack of submission by petitioners of dossiers on substances testing positive *in vivo*. However, it is noted that in the majority of cases, *in vivo* studies were performed in a second stage, after the initial submission of dossiers to EFSA (or to the Scientific Committee on Food - SCF). Thus, the outcome of *in vivo* assays could not be anticipated.

Further analysis of *in vitro* positives

Chromosomal aberration assays (CA)

Data on the 46 substances positive in the chromosomal aberration test *in vitro*, and negative in the cytogenetic tests in rodent bone marrow, were further inspected to identify the possible role of exogenous metabolism and/or high doses and excessive toxicity in generating false positive results.
Detailed information on the qualitative and/or quantitative effect of exogenous metabolism on test results was available for 35 substances, which were distributed as follows:

Positive with and without S9	13/35 (37%)
Positive only or predominantly with S9	11/35 (31%)
Positive only without S9	11/35 (31%)

The results indicate that the majority of substances (24/35, 69%) are directly clastogenic *in vitro*: of these, less than half were metabolically inactivated *in vitro*. Conversely, a similar number of substances were metabolically activated by liver S9 into (more) genotoxic derivatives. Overall, even though it has to be considered that most phase II detoxifying enzymes are not active in liver S9, these data may indicate that metabolic detoxification is not a major determinant of the inactivity of these *in vitro* clastogens in cytogenetic tests in rodents.

Positive results in chromosomal aberration assays occurred to a comparable extent in V79 cells (16 positives), CHO/CHL (15 positives) and human lymphocytes (12 positives).

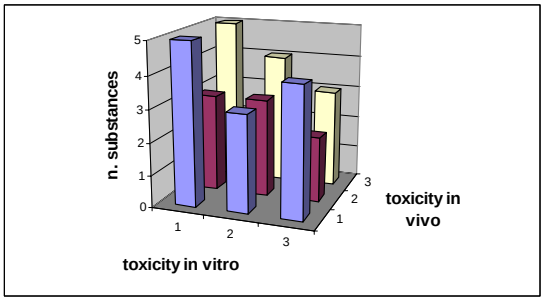
In order to verify whether these irrelevant *in vitro* positives were "high dose" positive only, data on the lowest effective concentration (LEC) *in vitro* were retrieved: LEC values ranged from 7.5 to 5000 µg/ml (median 425 µg/m), with the LEC of 20 out of 49 values lower than 1 mg/ml.

Third, data were further analysed to assess whether toxicity, or the lack of toxicity, could be implicated in the differential response of FCM clastogens in *in vitro* tests and in rodents. Based on the toxicity elicited at the lowest positive dose (*in vitro*) and highest tested dose (*in vivo*), each substance was allocated in one of the following semi-quantitative categories:

	In vitro	***In vivo***
Group 1 (distinct toxicity)	~ 40 % inhibition of mitotic index or survival	Significant decrease of PCE/NCE*
Group 2 (mild toxicity)	Less than 40 % inhibition of MI** or lethality	Clinical signs, with no effect on PCE/NCE
Group 3 (absence of toxicity)	No effect of treatment or no data	No effect observed or no data

*PCE: polychromatic erythrocytes; NCE: normochromatic erythrocytes. **MI: mitotic index

An outline of the distribution of scores is given in the following picture:



Clearly at least some of *in vitro* irrelevant positives did elicit a significant toxicity when tested *in vivo*, nevertheless producing negative results. Notably, some compounds displayed a relatively more pronounced toxicity in tests *in vivo* rather than *in vitro*. Thus, different effectiveness of the *in vitro* treatment, as indicated by more severe toxicity, cannot alone explain the divergent response produced by these FCM in cytogenetic tests *in vitro* and *in vivo*.

Gene mutation assays in mammalian cells

tk (tymidine kinase) locus

Data on ten substances positive in the Mouse Lymphoma Assay (MLA) but negative in either cytogenetic tests in rodent bone marrow (MN or CA) or the UDS assay were further analysed in order to find an explanation for discrepancies between *in vitro* and *in vivo* results (positive versus negative). The following factors were considered: i) influence of metabolic activation, ii) concentration/dose tested, iii) toxicity.

The influence of metabolic activation is shown below:

Positive with and without S9-mix:	2/10 (20%)
Positive only or predominantly with S9-mix:	7/10 (70%)
Positive only without S9-mix:	1/10 (10%)

Most of the substances are positive after metabolic activation (9/10). One explanation for the negative outcome i*n vivo* might be the reactive metabolite(s) formed *in vitro* is not formed at or does not reach the target organ (bone marrow) *in vivo.* The difference could also be due to differences in metabolic activation e.g. deactivation *in vivo* of reactive metabolites.

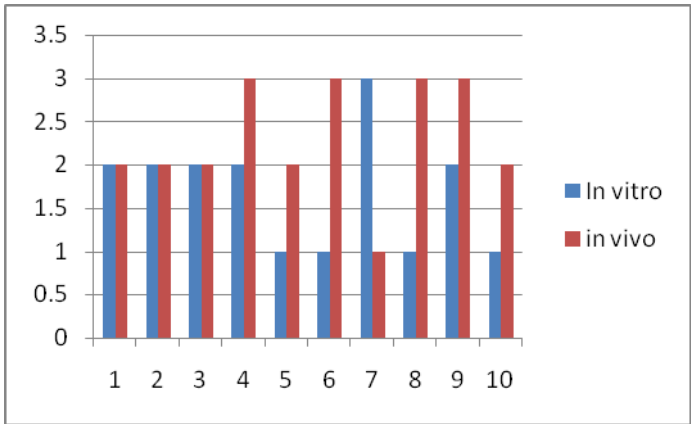
In order to assess whether the "irrelevant" *in vitro* positives were indeed "high concentration positives", the highest tested concentration (HTC) in vitro was determined. About half of the *tk* positive substances (5 out of 11) were tested up to relatively high concentrations *in vitro* (1000 – 5000 µg/ml), while due to toxicity constraints the remaining were tested at much lower concentrations (<100 µg/ml). Overall, most of the positive results are not associated with exceedingly high concentrations *in vitro*, and cannot be considered "high concentration positives".

Finally, the level of toxicity elicited by the substances tested *in vitro* and *in vivo* were recorded and compared to check whether these could provide an explanation for the different results between the in *vitro* and *in vivo* studies.

As for *in vitro* clastogens, substances positive in the MLA were allocated to one of three toxicity levels based on the signs of toxicity observed *in vitro* and *in vivo.* It is noted that different parameters were used to measure toxicity in *in vitro* studies (e.g. Relative Total Growth, relative survival, in some instances only a qualitative indication, e.g. "moderately toxic" was available). Consequently only a semi-qualitative ranking of toxicities is possible.

	In vitro	*In vivo*
Group 1 (distinct toxicity)	10% >RTG < 20%	Toxicity to bone marrow
Group 2 (mild toxicity)	Some toxicity observed but RTG > 20%	Clinical signs with no effect on toxicity
Group 3 (absence of toxicity)	No effect of treatment or no data	No effect observed or no data

The toxicity score for each substance *in vitro* and *in vivo* is shown in the following figure:



X axis: substance number; Y axis: toxicity score (group 1, 2 or 3)

As shown in the figure, distinct toxicity was observed for 4 substances *in vitro*, and all of these were less toxic *in vivo*, indicating detoxification *in vivo*. Only one substance had a severe toxic effect on the bone marrow, showing that the substance did reach the target organ. Five substances had a mild toxic effect *in vivo* (general toxic effect), and for 4 substances no toxic effect was observed. Although very few data were available for this analysis, it seems unlikely that the difference in response *in vitro* and *in vivo* was due to the higher toxicity of the test substance *in vitro*; rather, it seems more likely that this discrepancy may be due to differences in metabolism and/or in the bioavailability of the test substances to the target organ. The latter may be a critical factor as, although most of the substances were tested up to the maximum recommended dose, only one out of the 10 tested substances produced a direct evidence of reaching the bone marrow in a biologically relevant concentration.

hprt (hypoxanthine-guanine phosphoribosyl transferase) locus

Of the 3 substances which tested positive in the *hprt* assay, one was positive both with and without S9-mix, the substance was negative in the bacterial reverse mutation test and positive in the CA test with S9-mix. No *in vivo* tests were performed. However, the substance is classified as a possible carcinogen. Such compounds should not migrate at a detectable level into food.

The other substance was positive with S9-mix only and positive in the bacterial reverse mutation test with and without S9-mix. The substance was negative in the *in vitro* CA test and in an *in vivo* MN assay. No *in vivo* test for gene mutations was performed.

The third substance was positive both with and without S9 mix and positive in the bacterial reverse mutation test with and without S9 mix. The substance was negative in an *in vitro* CA assay and two *in vivo* studies (MN and CA). No data were submitted on gene mutations *in vivo*.

Because these substances were not tested for the same genetic endpoint *in vitro* and *in vivo* no further analysis was performed.

Overall conclusion

In conclusion, the routine application of a battery of three *in vitro* genotoxicity tests (bacterial reverse mutation test, mammalian cell gene mutation and chromosomal aberrations assays) produced a

relatively high incidence (28%) of positive results among food contact materials that were not confirmed *in vivo*. This high incidence may reflect both the proneness to positive results in the assays used (e.g. because of impaired DNA damage response and cell cycle control), as well as a high proportion of chemically reactive substances in this class of compounds, due to their technological function (e.g. as reactive monomers).

Whatever the reason for the high frequency of *in vitro* positives, the findings of the *in vitro* assays were not confirmed by the follow-up *in vivo* assays, in which only a small number of substances were positive (3 out of 49, or 6%). These results may indicate that over 90% of *in vitro* positives were "irrelevant positives" detected by mammalian cell assays (especially chromosomal aberrations assay) under *in vitro* conditions.

As to the reasons for the large discrepancy between *in vitro* and *in vivo* test results, none of the factors frequently invoked to explain the high frequency of *in vitro* positives, namely high dosing, excessive toxicity or artificial metabolic conditions, seems sufficient alone to account for these results. This may indicate on the one hand, as mentioned above, an inherent characteristic of *in vitro* mammalian cell systems to produce "irrelevant positive" results, as well as the relative insensitivity of the *in vivo* assays (cytogenetic tests in rodent erythropoietic cells and rat liver UDS) routinely applied in the past for the follow-up of *in vitro* positives. One reason for the apparent insensitivity of the MN assay *in vivo* could be that the tested substances did not reach the bone marrow. This underlines the importance of insuring that in future studies (i.e toxicity to bone marrow or verification by chemical analysis that the substance or its reactive metabolite can be detected in bone marrow or blood).

C. APPENDIX: SOME PRACTICAL CONSIDERATIONS IN COMBINING GENOTOXICITY TESTING WITH REPEATED-DOSE TOXICITY TESTS

In cases where genotoxicity studies are combined with repeated-dose toxicity (RTD) studies, standard protocols may need modification.

Timing of dosing

Since the DNA damage detected by the Comet assay is transient in nature and can be quickly removed (e.g., by DNA repair) timing of the last treatment before tissue sampling is critical. Therefore, the standard dosing regime of an RDT or micronucleus study needs to be modified by including an additional dose, generally 3-6 hours before sampling. This represents a deviation from the OECD TG 474 for the MN assay (OECD, 1997), where a sampling time of 24 and 48 hours after the last treatment is recommended. However, available data with this approach demonstrate (and theoretical considerations suggest) that dosing 3-6 hours pre-sacrifice has no impact on micronucleus analysis.

Requirements for the top dose

One concern when integrating genotoxicity testing into RDT studies is a possible reduced sensitivity as the top dose would be typically lower under these conditions than in acute studies. For this reason, the ICH S2R1 guideline (ICH, 2008) defines criteria to determine whether the top dose in multiple administration studies is appropriate for genotoxicity evaluation, in particular when the study is used to follow-up positive *in vitro* findings or when the initial (tier 1) battery is done without an *in vitro* mammalian cell test. Any one of the criteria listed below is sufficient under these conditions to demonstrate that the top dose in a toxicology study is appropriate for micronucleus analysis and for other genotoxicity evaluation:

- Maximum feasible dose (MFD) based on physico-chemical properties of the drug in the vehicle, provided the MFD in that vehicle is similar to that achievable with acute administration;
- Limit dose of 1000 mg/kg per day for studies of 14 days or longer, if this is tolerated;
- Maximal possible exposure demonstrated either by reaching a plateau/saturation in systemic exposure or by substance accumulation;
- Top dose is ≥ 50% of the top dose that would be used for acute administration, if such acute data are available.

Selection of a top dose based only on an exposure margin, i.e. the multiple over clinical exposure is not considered sufficient justification according to the ICH S2R1 guideline.

Influence of repeated bleeding

Repeated bleeding of animals, either for obtaining toxicokinetic and/or routine toxicology parameters or for multiple time points in the peripheral blood micronucleus assay have been suggested as a potential cause of increasing background MN frequencies, due to stimulation of erythropoiesis. However, results from recent studies addressing this issue do not indicate that repeated bleeding is a critical confounding factor for micronucleus induction in rats as long as bleeding is kept to reasonable volumes (Rothfuss et al., 2010a). Nevertheless, in order to minimize disturbances to erythropoiesis, it is advisable to limit the number of bleeds and to use the smallest possible volumes of blood.

Impact of toxicity

When blood or bone marrow micronucleus measurement is done in a multiweek RDT study (e.g., 28 days) marked haematotoxicity can affect the ability to detect MN, i.e., a dose that induces detectable increases in MN after acute treatment might be too toxic to analyse after multiple treatments. Therefore, it is advisable to include an early blood sampling at 3-4 days in cases of test substances that

are severely toxic for blood or bone marrow. This includes substances that induce aneuploidy, such as potent spindle poisons. As it is possible to freeze fixed blood samples for extended storage, sampling at an early time-point could be performed routinely and evaluation of the samples could be done if found necessary based on the final outcome of the RDT study.

Cytotoxicity may also have a confounding effect in the Comet assay. While available data suggests that false-positive results due to cytotoxicity can occur in the Comet assay *in vitro* no such evidence has so far been published for the Comet assay *in vivo*. Nevertheless, since tissue-toxicity is more likely to be induced during an RDT study than an acute-dose study this issue might be more critical when performing the Comet assay integrated into an RDT study (Vasquez, 2010). A review of published (and some unpublished) *in vivo* Comet assay data with concurrent cytotoxicity analysis data of mainly non-genotoxic carcinogens recently performed by an IWGT Working Group did not provide evidence that cytotoxicity, by itself, generates increases in DNA migration resulting in false-positive findings (Rothfuss et al., 2010b). Anyway, it is imperative to describe in study reports any confounding factors that may have an influence on the induction of comets, such as cytotoxicity or cell division.

Positive control

In order to avoid the inclusion of separate positive control animal groups for both endpoints (and different tissues) it is highly recommended to use a single positive control substance appropriate for all conditions. It might be sufficient to treat control group animals using an acute-dose protocol. With sufficient experience within a laboratory, the use of concurrent positive controls for the well-established MN endpoint may not be needed with every study but only periodically, e.g. every few months (Pfuhler et al., 2009). A practical approach in which two positive control substances within the same animal group is used has been proposed (Vasquez, 2010). For induction of micronuclei, cyclophosphamide is administered by i.p. injection 24 h (micronucleus analysis in bone marrow) and/or 48 h (micronucleus analysis in blood) prior to sampling. A single oral administration of ethyl methanesulphonate 3-4 hours prior to harvest is used for the same group of animals that was injected with cyclophosphamide for inducing positive comet effects in any tissue sampled.

References

ICH, 2010. ICHS2(R1): Genotoxicity: Testing and Data Interpretation for Pharmaceuticals Intended for Human Use. Step 2 Document, 2008. Available at: http://www.ich.org/products/guidelines/safety/article/safety-guidelines.html. Also available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S2_R1/Step2/S2_R1__Guideline.pdf.

OECD, 1997. OECD Guidelines for the Testing of Chemicals – Test Guideline 474, Mammalian Erythrocyte Micronucleus Test (adopted 21st July,1997). Organisation for Economic Cooperation and Development, Paris. Available at: http://lysander.sourceoecd.org/vl=1154212/cl=15/nw=1/rpsv/cw/vhosts/oecdjournals/1607310x/v1n4/contp1-1.htm.

Pfuhler S, Kirkland D, Kasper P, Hayashi M, Vanparys P, Carmichael P, Dertinger S, Eastmond D, Elhajouji A, Krul C, Rothfuss A, Schoening G, Smith A, Speit G, Thomas C, van Benthem J, Corvi R, 2009. Reduction of Use of Animals in Regulatory Genotoxicity Testing: Identification and Implementation Opportunities - Report from an ECVAM Workshop. Mutat. Res. 680, 31- 42.

Rothfuss A, O'Donovan M, De Boeck M, Brault D, Czich A, Custer L, Hamada S, Plappert-Helbig U, Hayashi M, Howe J, Kraynak A, van der Leede B, Nakajima M, Priestley C, Thybaud V, Saigo K, Sawant S, Shi J, Storer R, Struwe M, Vock E, Galloway S, 2010a. Collaborative study on 15 compounds in the rat liver Comet Assay integrated into 2- and 4-week repeat-dose studies. Mutat. Res. 702, 40-69.

Rothfuss A, Honma M, Czich A, Aardema MJ, Burlinson B, Galloway S, Hamada S, Kirkland D, Heflich RH, Howe J, Nakajima M, O'Donovan M, Plappert-Helbig U, Priestley C, Recio L, Schuler M, Uno Y, Martus H-J, 2010b. Report of the IWGT Working Group. Improvement of *in vivo* genotoxicity assessment:

combination of acute tests and integration into standard toxicity testing. Mutat. Res. December 2010, [Epub ahead of print]. doi:10.1016/j.mrgentox.2010.12.005.

Vasquez MZ, 2010. Combining the *in vivo* comet and micronucleus assays: a practical approach to genotoxicity testing and data interpretation. Mutagenesis 25, 187-199.

D. APPENDIX: WORK ONGOING IN OTHER GROUPS

In Vitro Genetic Toxicity Testing (IVGT) Project

The Emerging Issues Subcommittee of the ILSI Health and Environmental Sciences Institute (HESI) started the project on Relevance and Follow-up of Positive Results from *In Vitro* Genetic Toxicity Testing (IVGT) in June 2006. The mission of the IVGT project committee is to improve the scientific basis of the interpretation of results from *in vitro* genetic toxicology tests for purposes of more accurate human risk assessment; to develop follow-up strategies for determining the relevance of *in vitro* test results to human health; and to provide a framework for integration of *in vitro* testing results into a risk-based assessment of the effects of chemical exposures on human health.
The report of the Review Subgroup (Dearfield et al., 2011) provides a comprehensive evaluation of existing tests and guidelines, and presents a decision tree for follow-up strategies to *in vitro* positives. The report of the New/Emerging Technologies Workgroup (Lynch et al., 2011) is a summary of various novel and emerging technologies in genetic toxicology. The Quantitative Subgroup continues its work on collecting and evaluating dose response genetic toxicity data. Through this subgroup, IVGT is also involved in the validation of the *Pig-a* assay (see chapter 8.2.2) via inter-laboratory trials.

For the coming years, new initiatives of the project have been started. An Improving Existing Assays Workgroup, providing research and data for consideration in the 1 mM versus 10 mM debate about highest concentrations for testing, as well as evaluating several commonly used cell lines for genomic integrity, a New Approaches workgroup, hosting a workshop for presentations and discussions on new models that are not currently used in genotoxicity testing but could be applicable, and a Nano-genotoxicology Working Group providing a forum for evaluating the genetic toxicity of nanoparticles/nanomaterials have been initiated. The latter workgroup organised a workshop during the 2010 annual meeting of the American Environmental Mutagenicity Society. The results of this workshop will be published soon.

International Workshops on Genotoxicity Testing (IWGT)

The International Workshops on Genotoxicity Testing (IWGT) is an initiative of a number of scientists to discuss current issues on genotoxicity. The recommendations of the four earlier IWGT workshops have been highly influential in shaping revisions of OECD guidelines and the recommendations in the ICH S2A and B guidances.

The aim of the last IWGT meeting in 2009 was to revisit some "old" topics in regulatory genotoxicity testing such as cytotoxicity endpoints for *in vitro* tests and photogenotoxicity. New topics arising from changes in regulatory guidances will include discussions about top concentrations needed for *in vitro* tests, integration of genotoxicity endpoints into standard rodent toxicity studies and predictive alternatives to *in vivo* tests. As usual, invited experts, such as scientists from academia, government and industry from across the world, will participate to provide focused discussion and to give conclusive recommendations. The success of IWGT has been largely due to getting powerful representation of all global stakeholders around one table, and striving for data-driven consensus.

For the coming period, the following topics are being addressed: (1) suitable top concentration for tests with mammalian cells, (2) photogenotoxicity testing requirements, (3) *in vitro* test approaches with better predictivity, (4) improvement of *in vivo* genotoxicity assessment; the link to standard toxicity testing, (5) use of historical control data for the interpretation of positive results, and (6) suitable follow-up risk assessment testing for *in vivo* positive result. The conclusions of topics 3 and 6 are already published (Pfuhler et al., 2010: Thybaud et al., 2010) and those of the other topics will follow.

Organisation for Economic Cooperation and Development (OECD)

The Organisation for Economic Cooperation and Development (OECD) will start in March 2011 an extensive project in which all existing OECD guidelines for genotoxicity testing will be re-evaluated. At the moment Canada, France and the Netherlands are lead countries. The proposal will be to archive those guidelines which are outdated and hardly ever used. From the remaining ones, those which are commonly used in all known testing strategies may be revised and/or updated. For the *in vitro* guidelines using mammalian cells, the revision may comprise a lowering of the top dose concentration, a recommendation for the cell type to be used and a recommendation how to determine cytotoxicity. For gene mutation in mammalian cells *in vitro*, a separate guideline will be developed for the mouse lymphoma assay in addition to the one for other endpoints like *hprt* or *aprt.* Guidelines for germ cell genotoxicity tests will be revised and/or updated as a separate group. In addition to archiving or revision, new test guidelines will be developed for the *in vivo* gene mutation assay with transgenic animals, the Comet assay, and DNA adducts.

Japanese Centre for the Validation of Alternative Methods (JACVAM) initiative

JaCVAM (the Japanese Centre for the Validation of Alternative Methods) is coordinating an International Validation Study on the Comet assay *in vivo* and *in vitro*. While the experimental phase for the Comet assay *in vivo* will be finalised during 2011, the Comet assay *in vitro* will require more time. As soon as the validation study report for the Comet assay *in vivo* will be available, the validation study will be peer reviewed, while an OECD Test Guideline will be drafted (http://jacvam.jp/en)

The European Cosmetics Association (COLIPA) project

The OECD guidelines dealing with genotoxicity tests in mammalian cells allow the use of a variety of cell lines, strains or primary cells including human cells. However, one of the sources of unexpected results obtained in these *in vitro* genotoxicity tests may be the cell type used. The European Cosmetics Association (COLIPA) started an initiative to investigate the importance of the cell type used (Fowler et al., 2011 manuscript submitted). This question was also addressed at the 5th International Workshop on Genotoxicity Testing (IWGT) in August 2009.

In the COLIPA study (Fowler et al., 2011, submitted), the micronucleus induction in three *p53*-deficient rodent cell lines (V79, CHL and CHO cells) is compared with the induction in two human cell lines (TK6 and HepG2 cells) and human peripheral blood lymphocytes (HuLy) which are all *p53*-proficient. In the study, 19 substances that were accepted as producing false positive results in *in vitro* mammalian cell assays (Kirkland et al., 2008) are investigated. These chemicals are all negative in the Ames test and in *in vivo* genotoxicity studies and are either non-carcinogens or rodent carcinogens with a non-mutagenic mode of action. The study clearly demonstrated that the rodent cell lines were more susceptible to both cytotoxicity and micronucleus induction than *p53*-competent cells and, consequently, more susceptible to giving false positive results. Positive responses were mostly found in V79 cells, frequently in CHL and CHO cells, less frequently in TK6 cells, rarely in human lymphocytes and almost never in HepG2 cells. The authors concluded that a careful selection of the cell type for genotoxicity testing may lead to a reduction in the percentage of false positive results without decreasing the sensitivity of the assays.

These findings were confirmed in other laboratories. During the IWGT workshop (Pfuhler et al., 2010) a comparison of several cell lines used at Novartis, Switzerland, was reported. Results from the *in vitro* micronucleus test in different cell types (V79, 65 substances; L5178Y, 51 substances; TK6 cells, 80 substances) were compared with *in vitro* MN or chromosome aberration induction in human peripheral blood lymphocytes. It was reported that all cell lines detected the positives from the primary human lymphocyte studies whereas particularly the rodent cell lines (V79 and L5178Y cells) had a low specificity (around 60%). The *p53*-proficient TK6 cells had the best overall concordance (81%) with a specificity of 80%.

In the same workshop data were shown from a multi-laboratory exercise (10 different laboratories), which were used for the finalization of the *in vitro* micronucleus OECD guideline. Eleven chemicals, all relevant *in vivo* genotoxic carcinogens, were tested in the *in vitro* micronucleus test in five different cell lines (CHL, V79, CHO, L5178Y and TK6 cells). With one exception (2-aminoanthracene in CHO cells in one laboratory), all chemicals were as expected positive in the *in vitro* micronucleus test in all cell lines in all laboratories at concentrations of approximately 50% toxicity as measured by relative population doublings. Apparently, all cell lines have a comparable sensitivity, although a low specificity was reported earlier for the rodent cell lines. Moreover, it demonstrates that an increase of specificity as found in the human *p53*-proficient cell and cell lines does not come at the cost of a decreased sensitivity of the assays.

Both in the COLIPA project and at the IWGT workshop, it was recommended to avoid the use of *p53*-compromised cells but instead to use *p53*-competent and preferably human cells in *in vitro* mammalian genotoxicity tests. At the workshop also results obtained in the same COLIPA project were presented concerning the genetic stability of several commonly used cell lines over 50 passages in continuous culture. TK6 cells maintained a stable number of chromosomes whereas the modal chromosome number for CHL cells decreased by 2 and for CHO cells increased by 1. Apparently as time in culture increases, the commonly used rodent cell lines are more prone to genomic instability which may partially explain the higher frequency of positive responses. The IWGT further recommends to adhere to good cell practice, characterise all new cells, check regularly for drift, and work from low passage stocks. It was emphasised that a common genotoxicity cell bank with fully characterised stocks of all cells would be very useful.

References

Dearfield KL, Thybaud V, Cimino MC, Custer L, Czich A, Harvey JS, Hester S, Kim JH, Kirkland D, Levy DD, Lorge E, Moore MM, Ouédraogo-Arras G, Schuler M, Suter W, Sweder K, Tarlo K, van Benthem J, van Goethem F, Witt KL, 2011. Follow-up actions from positive results of in vitro genetic toxicity testing. Environ. Mol. Mutagen. 52: 177-204.

Fowler P, Smith K, Young J, Jeffrey L, Kirkland D, Pfuhler S, Carmichael P, 2011. Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. I. Choice of cell type. Publication submitted.

Kirkland D, Kasper P, Müller L, Corvi R, Speit G, 2008. Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: A follow-up to an ECVAM workshop, Mutat. Res. 653, 99-108.

Lynch AM, Sasaki JC, Elespuru R, Jacobson-Kram D, Thybaud V, De Boeck M, Aardema MJ, Aubrecht J, Benz RD, Dertinger SD, Douglas GR, White PA, Escobar PA, Fornace A Jr, Honma M, Naven RT, Rusling JF, Schiestl RH, Walmsley RM, Yamamura E, van Benthem J, Kim JH, 2011. New and emerging technologies for genetic toxicity testing. Environ. Mol. Mutagen. 52, 205-223.

Pfuhler S, Kirst A, Aardema M, Banduhn N, Goebel C, Araki D, Costabel-Farkas M, Dufour E, Fautz R, Harvey J, Hewitt NJ, Hibatallah J, Carmichael P, Macfarlane M, Reisinger K, Rowland J, Schellauf F, Schepky A, Scheel J, 2010. A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: genotoxicity. A COLIPA analysis. Regul. Toxicol. Pharmacol. 57, 315-324. Erratum in: Regul. Toxicol. Pharmacol. 58, 544.

Thybaud V, Macgregor JT, Müller L, Crebelli R, Dearfield K, Douglas G, Farmer PB, Gocke E, Hayashi M, Lovell DP, Lutz WK, Marzin D, Moore M, Nohmi T, Phillips DH, Van Benthem J, 2010. Strategies in case of positive *in vivo* results in genotoxicity testing. Mutat. Res. 2010. [Epub ahead of print]. doi:10.1016/j.mrgentox.2010.09.002.

3007

3008 **Glossary [and/or] abbreviations**

3009 **To be prepared**

3010