June 2012
EMA/CHMP/ICH/126642/2008

ICH guideline S2 (R1) on genotoxicity testing and data interpretation for pharmaceuticals intended for human use

Step 5

Transmission to CHMP	March 2008
Adoption by CHMP for release for consultation	March 2008
End of consultation (deadline for comments)	May 2008
Final adoption by CHMP	December 2011
Date for coming into effect	June 2012

7 Westferry Circus ● Canary Wharf ● London E14 4HB ● United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7418 8416
E-mail ich@ema.europa.eu **Website** www.ema.europa.eu

An agency of the European Union

Table of contents

1. Introduction

1.1. *Objectives of the guideline*

This guidance replaces and combines the ICH S2A and S2B guidelines. The purpose of the revision is to optimize the standard genetic toxicology battery for prediction of potential human risks, and to provide guidance on interpretation of results, with the ultimate goal of improving risk characterization for carcinogenic effects that have their basis in changes in the genetic material. The revised guidance describes internationally agreed upon standards for follow-up testing and interpretation of positive results *in vitro* and *in vivo* in the standard genetic toxicology battery, including assessment of non-relevant findings. This guidance is intended to apply only to products being developed as human pharmaceuticals.

1.2. *Background*

The recommendations from the latest Organization for Economic Co-operation and Development (OECD) guidelines and the reports from the International Workshops on Genotoxicity Testing (IWGT) have been considered where relevant. In certain cases, there are differences from the OECD or IWGT recommendations, which are noted in the text. The following notes for guidance should be applied in conjunction with other ICH guidances.

1.3. *Scope of the guideline*

The focus of this guidance is testing of new “small molecule” drug substances, and the guidance does not apply to biologics. Advice on the timing of the studies relative to clinical development is provided in the ICH M3 (R2) guidance.

1.4. *General principles*

Genotoxicity tests can be defined as *in vitro* and *in vivo* tests designed to detect compounds that induce genetic damage by various mechanisms. These tests enable hazard identification with respect to damage to DNA and its fixation. Fixation of damage to DNA in the form of gene mutations, larger scale chromosomal damage or recombination is generally considered to be essential for heritable effects and in the multi-step process of malignancy, a complex process in which genetic changes might possibly play only a part. Numerical chromosome changes have also been associated with tumorigenesis and can indicate a potential for aneuploidy in germ cells. Compounds that are positive in tests that detect such kinds of damage have the potential to be human carcinogens and/or mutagens. Because the relationship between exposure to particular chemicals and carcinogenesis is established for humans, whilst a similar relationship has been difficult to prove for heritable diseases, genotoxicity tests have been used mainly for the prediction of carcinogenicity. Nevertheless, because germ line mutations are clearly associated with human disease, the suspicion that a compound might induce heritable effects is considered to be just as serious as the suspicion that a compound might induce cancer. In addition, the outcome of genotoxicity tests can be valuable for the interpretation of carcinogenicity studies.

2. The standard test battery for genotoxicity

2.1. Rationale

Registration of pharmaceuticals requires a comprehensive assessment of their genotoxic potential. Extensive reviews have shown that many compounds that are mutagenic in the bacterial reverse mutation (Ames) test are rodent carcinogens. Addition of *in vitro* mammalian tests increases sensitivity for detection of rodent carcinogens and broadens the spectrum of genetic events detected, but also decreases the specificity of prediction; i.e., increases the incidence of positive results that do not correlate with rodent carcinogenicity. Nevertheless, a battery approach is still reasonable because no single test is capable of detecting all genotoxic mechanisms relevant in tumorigenesis.

The general features of a standard test battery are as follows:

i. Assessment of mutagenicity in a bacterial reverse gene mutation test. This test has been shown to detect relevant genetic changes and the majority of genotoxic rodent and human carcinogens.

ii. Genotoxicity should also be evaluated in mammalian cells *in vitro* and/or *in vivo* as follows.

Several *in vitro* mammalian cell systems are widely used and can be considered sufficiently validated: The *in vitro* metaphase chromosome aberration assay, the *in vitro* micronucleus assay (note 1) and the mouse lymphoma L5178Y cell *Tk* (thymidine kinase) gene mutation assay (MLA). These three assays are currently considered equally appropriate and therefore interchangeable for measurement of chromosomal damage when used together with other genotoxicity tests in a standard battery for testing of pharmaceuticals, if the test protocols recommended in this guideline are used.

In vivo test(s) are included in the test battery because some agents are mutagenic *in vivo* but not *in vitro* (note 2) and because it is desirable to include assays that account for such factors as absorption, distribution, metabolism and excretion. The choice of an analysis either of micronuclei in erythrocytes (in blood or bone marrow), or of chromosome aberrations in metaphase cells in bone marrow, is currently included for this reason (note 3). Lymphocytes cultured from treated animals can also be used for cytogenetic analysis, although experience with such analyses is less widespread.

In vitro and *in vivo* tests that measure chromosomal aberrations in metaphase cells can detect a wide spectrum of changes in chromosomal integrity. Breakage of chromatids or chromosomes can result in micronucleus formation if an acentric fragment is produced; therefore assays that detect either chromosomal aberrations or micronuclei are considered appropriate for detecting clastogens. Micronuclei can also result from lagging of one or more whole chromosome(s) at anaphase and thus micronucleus tests have the potential to detect some aneuploidy inducers. The MLA detects mutations in the *Tk* gene that result from both gene mutations and chromosome damage. There is some evidence that MLA can also detect chromosome loss.

There are several additional *in vivo* assays that can be used in the battery or as follow-up tests to develop weight of evidence in assessing results of *in vitro* or *in vivo* assays (see below). Negative results in appropriate *in vivo* assays (usually two), with adequate justification for the endpoints measured, and demonstration of exposure (see section 4.4) are generally considered sufficient to demonstrate absence of significant genotoxic risk.

2.2. Description of the two options for the standard battery

The following two options for the standard battery are considered equally suitable (see note 4):

Option 1

i. A test for gene mutation in bacteria.

ii. A cytogenetic test for chromosomal damage (the *in vitro* metaphase chromosome aberration test or *in vitro* micronucleus test), or an *in vitro* mouse lymphoma *Tk* gene mutation assay.

iii. An *in vivo* test for genotoxicity, generally a test for chromosomal damage using rodent hematopoietic cells, either for micronuclei or for chromosomal aberrations in metaphase cells.

Option 2

i. A test for gene mutation in bacteria.

ii. An *in vivo* assessment of genotoxicity with two different tissues, usually an assay for micronuclei using rodent hematopoietic cells and a second *in vivo* assay. Typically this would be a DNA strand breakage assay in liver, unless otherwise justified (see below; also section 4.2 and note 12).

   There is more historical experience with Option 1, partly because it is based on S2A and B. Nevertheless, the reasoning behind considering Options 1 and 2 equally acceptable is as follows: When a positive result occurs in an *in vitro* mammalian cell assay, clearly negative results in two well conducted *in vivo* assays, in appropriate tissues and with demonstrated adequate exposure, are considered sufficient evidence for lack of genotoxic potential *in vivo* (see section 5.4.1.1 below). Thus a test strategy in which two *in vivo* assays are conducted is the same strategy that would be used to follow up a positive result *in vitro* (see note 4).

Under both standard battery options, either acute or repeat dose study designs *in vivo* can be used. In case of repeated administrations, attempts should be made to incorporate the genotoxicity endpoints into toxicity studies, if scientifically justified. When more than one endpoint is evaluated *in vivo* it is preferable that they are incorporated into a single study. Often sufficient information on the likely suitability of the doses for the repeat-dose toxicology study is available before the study begins and can be used to determine whether an acute or an integrated test will be suitable.

For compounds that give negative results, the completion of either option of the standard test battery, performed and evaluated in accordance with current recommendations, will usually provide sufficient assurance of the absence of genotoxic activity and no additional tests are warranted. Compounds that give positive results in the standard test battery might, depending on their therapeutic use, need to be tested more extensively (see Section 5).

There are several *in vivo* assays that can be used as the second part of the *in vivo* assessment under option 2 (see section 4.2), some of which can be integrated into repeat-dose toxicology studies. The liver is typically the preferred tissue because of exposure and metabolizing capacity, but choice of *in vivo* tissue and assay should be based on factors such as any knowledge of the potential mechanism, of the metabolism *in vivo*, or of the exposed tissues thought to be relevant.

Information on numerical changes can be derived from the mammalian cell assays *in vitro* and from the micronucleus assays *in vitro* or *in vivo*. Elements of the standard protocols that can indicate such potential are elevations in the mitotic index, polyploidy induction and micronucleus evaluation. There is also experimental evidence that spindle poisons can be detected in MLA. The preferred *in vivo* cytogenetic test under Option 2 is the micronucleus assay, not a chromosome aberration assay, to include more direct capability for detection of chromosome loss (potential for aneuploidy).

The suggested standard set of tests does not imply that other genotoxicity tests are generally considered inadequate or inappropriate. Additional tests can be used for further investigation of genotoxicity test results obtained in the standard battery (see sections 4.2 and 5). Alternative species, including non-rodents, can also be used if indicated, and if sufficiently validated.

Under conditions in which one or more tests in the standard battery cannot be employed for technical reasons, alternative validated tests can serve as substitutes provided sufficient scientific justification is given.

2.3. Modifications to the test battery

The following sections describe situations where modification of the standard test battery might be advisable.

2.3.1. Exploratory clinical studies

For certain exploratory clinical studies, fewer genotoxicity assays or different criteria for justification of the maximum dose in vivo might apply (see ICH M3(R2) guidance).

2.3.2. Testing compounds that are toxic to bacteria

In cases where compounds are highly toxic to bacteria (e.g., some antibiotics), the bacterial reverse mutation (Ames) test should still be carried out, just as cytotoxic compounds are tested in mammalian cells, because mutagenicity can occur at lower, less toxic concentrations. In such cases, any one of the *in vitro* mammalian cell assays should also be done, i.e., Option 1 should be followed.

2.3.3. Compounds bearing structural alerts for genotoxic activity

Structurally alerting compounds (Note 5) are usually detectable in the standard test battery since the majority of “structural alerts” are defined in relation to bacterial mutagenicity. A few chemical classes are known to be more easily detected in mammalian cell chromosome damage assays than bacterial mutation assays. Thus negative results in either test battery with a compound that has a structural alert is usually considered sufficient assurance of a lack of genotoxicity. However, for compounds bearing certain specific structural alerts, modification to standard protocols can be appropriate (Note 5). The choice of additional test(s) or protocol modification(s) depends on the chemical nature, the known reactivity and any metabolism data on the structurally alerting compound in question.

2.3.4. Limitations to the use of *in vivo* tests

There are compounds for which many *in vivo* tests (typically in bone marrow, blood or liver) do not provide additional useful information. These include compounds for which data on toxicokinetics or pharmacokinetics indicate that they are not systemically absorbed and therefore are not available to the target tissues. Examples of such compounds are some radioimaging agents, aluminum based antacids, some compounds given by inhalation, and some dermally or other topically applied pharmaceuticals. In cases where a modification of the route of administration does not provide sufficient target tissue exposure, and no suitable genotoxicity assay is available in the most exposed tissue, it might be appropriate to base the evaluation only on *in vitro* testing. In some cases evaluation of genotoxic effects at the site of contact can be warranted, although such assays have not yet been widely used (note 6).

2.4. Detection of germ cell mutagens

Results of comparative studies have shown that, in a qualitative sense, most germ cell mutagens are likely to be detected as genotoxic in somatic cell tests so that negative results of *in vivo* somatic cell genotoxicity tests generally indicate the absence of germ cell effects.

3. Recommendations for *in vitro* tests

3.1. Test repetition and interpretation

Reproducibility of experimental results is an essential component of research involving novel methods or unexpected findings; however, the routine testing of drugs with standard, widely used genotoxicity tests often does not call for replication. These tests are sufficiently well characterized and have sufficient internal controls that repetition of a clearly positive or negative assay is not usually warranted. Ideally it should be possible to declare test results clearly negative or clearly positive. However, test results sometimes do not fit the predetermined criteria for a positive or negative call and therefore are declared “equivocal”. The application of statistical methods can aid in data interpretation; however, adequate biological interpretation is of critical importance. An equivocal test that is repeated might result in (i) a clearly positive outcome, and thus an overall positive result; (ii) a negative outcome, so that the result is not reproducible and overall negative, or (iii) another equivocal result, with a final conclusion that remains equivocal.

3.2. Recommended protocol for the bacterial mutation assay

Advice on the protocols is given in the OECD guideline (1997) and the IWGT report (Gatehouse *et al.*, 1994).

3.2.1. Selection of top dose level

Maximum dose level

The maximum dose level recommended is 5000 µg/plate (or 5 µL/plate for liquid test substance) when not limited by solubility or cytotoxicity.

Limit of solubility

For bacterial cultures, precipitating doses are scored provided precipitate does not interfere with scoring, toxicity is not limiting, and the top concentration does not exceed 5000 µg/plate (or 5 µL/plate for liquid test substance). If no cytotoxicity is observed, then the lowest precipitating dose should be used as the top dose scored. If dose related cytotoxicity or mutagenicity is noted, irrespective of solubility, the top dose scored should be based on cytotoxicity as described below.

Limit of cytotoxicity:

In the Ames test, the doses scored should show evidence of significant toxicity, but without exceeding a top dose of 5000 µg/plate. Toxicity might be detected by a reduction in the number of revertants, and/or clearing or diminution of the background lawn.

3.2.2. Study design / Test protocol

The recommended set of bacterial strains (OECD) includes those that detect base substitution and frameshift mutations as follows:

- *Salmonella typhimurium* TA98;
- *Salmonella typhimurium* TA100;
- *Salmonella typhimurium* TA1535;
- either *Salmonella typhimurium* TA1537 or TA97 or TA97a;
- and either *Salmonella typhimurium* TA102 or *Escherichia coli* WP2 *uvrA* or *Escherichia coli* WP2 *uvrA* (pKM101).

One difference from the OECD and IWGT recommendations is that, based on experience with testing pharmaceuticals, a single bacterial mutation (Ames) test is considered sufficient when it is clearly negative or positive, and carried out with a fully adequate protocol including all strains with and without metabolic activation, a suitable dose range that fulfils criteria for top dose selection, and appropriate positive and negative controls. Also, for testing pharmaceuticals, either the plate incorporation or the pre-incubation method is considered appropriate for this single experiment (note 7). Equivocal or weak positive results might indicate that it would be appropriate to repeat the test, possibly with a modified protocol such as appropriate spacing of dose levels.

3.3. Recommended protocols for the mammalian cell assays

Advice on the protocols is given in the OECD guidelines (1997) and the IWGT publications (e.g., Kirsch-Volders *et al.*, 2003; Moore *et al.*, 2006). Advice on interpretation of MLA results is also given (Moore *et al.*, 2006), including use of a global evaluation factor. Several differences from these recommendations are noted here for testing pharmaceuticals, notably for selection of the top concentration (see details below).

3.3.1. Selection of top concentration

Maximum concentration

The maximum top concentration recommended is 1 mM or 0.5 mg/ml, whichever is lower, when not limited by solubility in solvent or culture medium or by cytotoxicity (note 8).

Limit of solubility

When solubility is limiting, the maximum concentration, if not limited by cytotoxicity, should be the lowest concentration at which minimal precipitate is visible in cultures, provided there is no interference with scoring. Evaluation of precipitation can be done by naked eye or by methods such as light microscopy, noting precipitate that persists or appears during culture (by the end of treatment).

Cytotoxicity

For *in vitro* cytogenetic assays for metaphase chromosome aberrations or for micronuclei, cytotoxicity should not exceed a reduction of about 50% in cell growth (notes 9 and 10). For the MLA, at the top dose there should be 80-90% cytotoxicity as measured by an RTG between 20-10% (note 9).

3.3.2. Study design/Test protocols

For the cytogenetic evaluation of chromosomal damage in metaphase cells *in vitro*, the test protocol should include the conduct of tests with and without metabolic activation, with appropriate positive and negative controls. Treatment with the test articles should be for 3 to 6 hours with a sampling time approximately 1.5 normal cell cycles from the beginning of the treatment. A continuous treatment without metabolic activation up to the sampling time of approximately 1.5 normal cell cycles should be conducted in case of negative or equivocal results for both short treatments, with and without metabolic activation. The same principles apply to the *in vitro* micronucleus assay, except that the sampling time is typically 1.5 to 2 normal cell cycles from the beginning of treatment to allow cells to complete mitosis and enter the next interphase. For both *in vitro* cytogenetic assays, there might be a need to modify the protocol for certain types of chemicals that could be more readily detected by longer treatment, delayed sampling times or recovery periods, e.g., some nucleoside analogues and some nitrosamines. In the metaphase aberration assay, information on the ploidy status should be obtained by recording the incidence of polyploid (including endoreduplicated) metaphases as a percentage of the number of metaphase cells. For MLA, the test protocol should include the conduct of tests with and without metabolic activation, with appropriate positive and negative controls, where the treatment with the test article is for 3 to 4 hours. A continuous treatment without metabolic activation for approximately 24 hours should be conducted in case of a negative or equivocal result for both short treatments, with and without metabolic activation. A standard MLA should include (i) the incorporation of positive controls that induce mainly small colonies, and (ii) colony sizing for positive controls, solvent controls and at least one positive test compound concentration (should any exist), including the culture that gave the greatest mutant frequency.

For mammalian cell assays *in vitro*, built-in confirmatory elements, such as those outlined above (e.g., different treatment lengths, tests with and without metabolic activation), should be used. Following such testing, further confirmatory testing in the case of clearly negative or positive test results is not usually warranted. Equivocal or weak positive results might call for repeating tests, possibly with a modified protocol such as appropriate spacing of the test concentrations.

3.3.3. Positive controls

Concurrent positive controls are important, but *in vitro* mammalian cell tests for genetic toxicity are sufficiently standardized that use of positive controls can generally be confined to a positive control with metabolic activation (when it is done concurrently with the non-activated test) to demonstrate the activity of the metabolic activation system and the responsiveness of the test system.

4. Recommendations for *in vivo* tests

4.1. Tests for the detection of chromosome damage in vivo

Either the analysis of chromosomal aberrations or the measurement of micronucleated polychromatic erythrocytes in bone marrow cells *in vivo* is considered appropriate for the detection of clastogens. Both rats and mice are considered appropriate for use in the bone marrow micronucleus test. Micronuclei can also be measured in immature (e.g., polychromatic) erythrocytes in peripheral blood in the mouse, or in the newly formed reticulocytes in rat blood (note 3). Likewise, immature erythrocytes can be used from any other species which has shown an adequate sensitivity to detect clastogens/aneuploidy inducers in bone marrow or peripheral blood (note 3). Systems for automated analysis (image analysis and flow cytometry) can be used if appropriately validated (OECD, 1997; Hayashi *et al.*, 2000; 2007). Chromosomal aberrations can also be analyzed in peripheral lymphocytes cultured from treated rodents (note 11).

4.2. Other in vivo genotoxicity tests

The same *in vivo* tests described as the second test in the standard battery (option 2) can be used as follow-up tests to develop weight of evidence in assessing results of *in vitro* or *in vivo* assays (notes 11 and 12). While the type of effect seen *in vitro* and any knowledge of the mechanism can help guide the choice of *in vivo* assay, investigation of chromosomal aberrations or of gene mutations in endogenous genes is not feasible with standard methods in most tissues. Although mutation can be measured in transgenes in rodents, this entails prolonged treatment (e.g., 28 days) to allow for mutation expression, fixation and accumulation, especially in tissues with little cell division (see note 12). Thus the second *in vivo* assay will often evaluate a DNA damage endpoint as a surrogate. Assays with the most published experience and advice on protocols include the DNA strand break assays such as the single cell gel electrophoresis ("Comet") assay and alkaline elution assay, the i*n vivo* transgenic mouse mutation assays and DNA covalent binding assays, (all of which may be applied in many tissues, note 12), and the liver unscheduled DNA synthesis (UDS) assay.

4.3. Dose selection for in vivo assays

Typically three dose levels are analyzed (Hayashi *et al.*, 2005).

4.3.1. Short-term studies

For short term (usually 1 to 3 administrations) studies, the top dose recommended for genotoxicity assays is a limit dose of 2000 mg/kg, if this is tolerated, or a maximum tolerated dose defined (for example for the micronucleus assay (OECD)) as the dose producing signs of toxicity such that higher dose levels, based on the same dosing regimen, would be expected to produce lethality. Similar recommendations have been made for the Comet assay (Hartmann *et al.*, 2003) and transgenic mutation assay (Heddle *et al.*, 2000). Suppression of bone marrow red blood cell production should also be taken into account in dose selection. Lower doses are generally spaced at approximately two to three fold intervals below this.

4.3.2. Multiple administration studies

Option 1 Battery: When the *in vivo* genotoxicity test is integrated into a multiple administration toxicology study, the doses are generally considered appropriate when the toxicology study meets the criteria for an adequate study to support human clinical trials; this can differ from dose selection criteria in the OECD guideline for the *in vivo* micronucleus assay. This applies when the *in vitro* mammalian cell test is negative (or "non-relevant positive"; see section 5).

Follow-up studies or Option 2 battery: When carrying out follow-up studies to address any indication of genotoxicity, or when using Option 2 with no *in vitro* mammalian cell assay, several factors should be evaluated to determine whether the top dose is appropriate for genotoxicity evaluation. Any one of the criteria listed below is considered sufficient to demonstrate that the top dose in a toxicology study (typically in rats) is appropriate for micronucleus analysis and for other genotoxicity evaluation:

i. 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; note 13).

ii. Limit dose of 1000 mg/kg for studies of 14 days or longer, if this is tolerated

iii. Maximal possible exposure demonstrated either by reaching a plateau/saturation in exposure or by compound accumulation. In contrast, substantial reduction in exposure to parent drug with time (e.g., ≥ 50% reduction from initial exposure) can disqualify the study (unless a blood sample taken in the first few days is available). If this is seen in one sex, generally the sex with reduced exposure would not be scored at the end of the study, unless there is enhanced exposure to a metabolite of interest.

iv. Top dose is ≥ 50% of the top dose that would be used for acute administration, i.e., close to the minimum lethal dose, if such acute data are available for other reasons. (The top dose for acute administration micronucleus tests is currently described in OECD guidance as the dose above which lethality would be expected; similar guidance is given [e.g. Hartmann *et al.*, 2003] for other *in vivo* assays.)

Selection of a top dose based only on an exposure margin (multiple over clinical exposure) without toxicity is not considered sufficient justification.

4.3.3. Testing compounds that are toxic for blood or bone marrow

Many compounds that induce aneuploidy, such as potent spindle poisons, are detectable in *in vivo* micronucleus assays in bone marrow or blood only within a narrow range of doses approaching toxic doses. This is also true for some clastogens. If toxicological data indicate severe toxicity to the red blood cell lineage (e.g., marked suppression of PCEs or reticulocytes), doses scored should be spaced not more than about 2 fold below the top, cytotoxic dose. If suitable doses are not included in a multi-week study, additional data that could contribute to the detection of aneugens and some toxic clastogens could be derived from any one of the following:

i. Early blood sampling (at 3 – 4 days) is advisable when there are marked increases in toxicity with increasing treatment time. For example, when blood or bone marrow is used for micronucleus measurement in a multiweek study (e.g., 28 days), and reticulocytes are scored, marked hematotoxicity can affect the ability to detect micronuclei; i.e., a dose that induces detectable increases in micronuclei after acute treatment might be too toxic to analyze after multiple treatments (Hamada *et al.*, 2001) . The early sample can be used to provide assurance that clastogens and potential aneugens are detected (but see notes 14 and 15).

ii. an *in vitro* mammalian cell micronucleus assay

iii. an acute bone marrow micronucleus assay

4.4. Demonstration of target tissue exposure for negative in vivo test results

In vivo tests have an important role in genotoxicity test strategies. The value of *in vivo* results is directly related to the demonstration of adequate exposure of the target tissue to the test compound. This is especially true for negative *in vivo* test results when *in vitro* test(s) have shown convincing evidence of genotoxicity, or when no *in vitro* mammalian cell assay is used. Evidence of adequate exposure could include toxicity in the tissue in question, or toxicokinetic data as described in the following section.

4.4.1. When an in vitro genotoxicity test is positive (or not done)

Assessments of *in vivo* exposure should be made at the top dose or other relevant doses using the same species, strain and dosing route used in the genotoxicity assay. When genotoxicity is measured in toxicology assays, exposure information is generally available as part of the toxicology assessment.

Demonstration of *in vivo* exposure should be made by any of the following measurements:

i. Cytotoxicity

    a. For cytogenetic assays: By obtaining a significant change in the proportion of immature erythrocytes among total erythrocytes in the tissue used (bone marrow or blood) at the doses and sampling times used in the micronucleus test or by measuring a significant reduction in mitotic index for the chromosomal aberration assay.

    b. For other *in vivo* genotoxicity assays: Toxicity in the liver or tissue being assessed, e.g., by histopathological evaluation or blood biochemistry toxicity indicators.

ii. Exposure

    a. Measurement of drug related material either in blood or plasma. The bone marrow is a well perfused tissue and levels of drug related materials in blood or plasma are generally similar to those observed in bone marrow. The liver is expected to be exposed for drugs with systemic exposure regardless of the route of administration.

    b. Direct measurement of drug-related material in target tissue, or autoradiographic assessment of tissue exposure.

If systemic exposure is similar to or lower than expected clinical exposure, alternative strategies might be called for such as:

i. Use of a different route of administration;

ii. Use of a different species with higher exposure;

iii. Use of a different tissue or assay (see section 2.3.4, “Limitations to the use of standard *in vivo* tests”).

When adequate exposure cannot be achieved (e.g., with compounds showing very poor target tissue availability) conventional *in vivo* genotoxicity tests have little value.

4.4.2. When in vitro genotoxicity tests are negative

If *in vitro* tests do not show genotoxic potential, *in vivo* (systemic) exposure can be assessed by any of the methods above, or can be assumed from the results of standard absorption, distribution, metabolism and excretion (ADME) studies in rodents done for other purposes.

4.5. Sampling times for in vivo assays

Selection of the sampling time in the *in vivo* MN, chromosomal aberration and UDS test should follow OECD (1997).

When micronucleus analysis is integrated into multi-week studies, sampling of blood or bone marrow can be done the day after the final administration (see recommendation for additional blood sampling time above).

For other genotoxicity assays, sampling time should be selected as appropriate for the endpoint measured; for example, DNA damage/strand break measurements are usually made a few (e.g., 2-6) hours after the last administration for the multiple daily administration. In the case of single administration, two sampling times should be used: a few hours and 24 hours after the treatment.

In principle, studies of any length can be considered appropriate, provided the top dose/exposure is adequate.

4.6. Number of animals analyzed

The number of animals analyzed is determined by current recommendations for the micronucleus assay (OECD) or other genotoxicity assays and generally does not include all the animals treated for a toxicology study. Animals used for genotoxicity analyses should be randomly selected from the group used for the toxicology study.

4.7. Use of male/female rodents in in vivo genotoxicity tests

If sex-specific drugs are to be tested, then the assay can be done in the appropriate sex. *In vivo* tests with the acute protocol can generally be carried out in only one sex. For acute tests, both sexes should be considered only if any existing toxicity, metabolism or exposure (Cmax or AUC) data indicate a toxicologically meaningful sex difference in the species being used. Otherwise, the use of males alone is considered appropriate for acute genotoxicity tests. When the genotoxicity test is integrated into a repeat-dose toxicology study in two sexes, samples can be collected from both sexes, but a single sex can be scored if there is no substantial sex difference evident in toxicity/metabolism. The dose levels for the sex(es) scored should meet the criteria for appropriate dose levels (sections 4.3.2 and 4.3.3).

Similar principles can be applied for other established *in vivo* genotoxicity tests.

4.8. Route of administration

The route of administration is generally the expected clinical route, e.g., oral, intravenous or subcutaneous, but can be modified if appropriate in order to obtain systemic exposure, e.g., for topically applied compounds (see section 2.3.4).

4.9. Use of positive controls for in vivo studies

For *in vivo* studies, it is considered sufficient to treat animals with a positive control only periodically, and not concurrently with every assay, after a laboratory has established competence in the use of the assay (note 16).

5. Guidance on evaluation of test results and on follow-up test strategies

Comparative trials have shown conclusively that each *in vitro* test system generates both false negative and false positive results in relation to predicting rodent carcinogenicity. Genotoxicity test batteries (of *in vitro* and *in vivo* tests) detect carcinogens that are thought to act primarily via a mechanism involving direct genetic damage, such as the majority of known human carcinogens. Therefore, these batteries are not expected to detect non-genotoxic carcinogens. Experimental conditions, such as the limited capability of the *in vitro* metabolic activation systems, can lead to false negative results in *in vitro* tests. The test battery approach is designed to reduce the risk of false negative results for compounds with genotoxic potential. On the other hand a positive result in any assay for genotoxicity does not always mean that the test compound poses a genotoxic/carcinogenic hazard to humans.

Although positive *in vitro* data could indicate intrinsic genotoxic properties of a drug, appropriate *in vivo* data determine the biological significance of these *in vitro* signals in most cases. Also, because there are several indirect mechanisms of genotoxicity that operate only above certain concentrations, it is possible to establish a safe level (threshold) for classes of drugs with evidence for such mechanisms (see 5.2. below, Müller and Kasper, 2000; Scott *et al.*, 1991; Thybaud *et al.*, 2007).

5.1. Assessment of biological relevance

The recommendations below assume that the test has been conducted using appropriate spacing of doses, levels of toxicity etc.

Small increases in apparent genotoxicity *in vitro* or *in vivo* should first be assessed for reproducibility and biological significance. Examples of results that are not considered biologically meaningful include:

i. Small increases that are statistically significant compared with the negative or solvent control values but are within the confidence intervals of the appropriate historical control values for the testing facility

ii. Weak/equivocal responses that are not reproducible

If either of the above conditions applies, the weight of evidence indicates a lack of genotoxic potential, the test is considered negative or the findings not biologically relevant, and no further testing is called for.

5.2. Evaluation of results obtained in in vitro tests

In evaluating positive results, especially for the microbial mutagenicity test, the purity of the test compound should be considered, to determine whether the positive result could be attributable to a contaminant.

5.2.1. Evaluation of positive results obtained in vitro in a bacterial mutation assay

Since positive results in the Ames test are thought to indicate DNA reactivity, extensive follow-up testing to assess the *in vivo* mutagenic and carcinogenic potential would be warranted to assess the potential risk for treatment of patients, unless justified by appropriate risk-benefit analysis.

There are some well characterized examples of artifactual increases in colonies that are not truly revertants. These can occur due to contamination with amino acids (i.e. providing histidine for *Salmonella typhimurium* strains or tryptophan for *Escherichia coli* strains), so that the bacterial reversion assay is not suitable for testing a peptide that is likely to degrade. Certain cases exist where positive results in bacterial mutation assays might be shown not to indicate genotoxic potential *in vivo* in humans, for example when bacterial-specific metabolism occurs, such as activation by bacterial nitroreductases.

5.2.2. Evaluation of positive results obtained in vitro in mammalian cell assays

Recommendations for assessing weight of evidence and follow-up testing for positive genotoxicity results are discussed in IWGT reports (e.g., Thybaud *et al.*, 2007). In addition, the scientific literature gives a number of conditions that can lead to a positive *in vitro* result of questionable relevance. Therefore, any *in vitro* positive test result should be evaluated based on an assessment of the weight of evidence as indicated below. This list is not exhaustive, but is given as an aid to decision-making.

i. The conditions do not occur *in vivo* (pH; osmolality; precipitates)

(Note that the 1 mM limit avoids increases in osmolality, and that if the test compound alters pH it is advisable to adjust pH to the normal pH of untreated cultures at the time of treatment)

ii. The effect occurs only at the most toxic concentrations.

In the MLA increases at ≥80% reduction in RTG

For *in vitro* cytogenetics assays when growth is suppressed by ≥50%

If any of the above conditions apply the weight of evidence indicates a lack of genotoxic potential; the standard battery (option 1) can be followed. Thus, a single *in vivo* test is considered sufficient.

5.2.3. Evaluation of in vitro negative results

For *in vitro* negative results further testing should be considered in special cases, such as (the examples given are not exhaustive, but are given as an aid to decision-making): The structure or known metabolism of the compound indicates that standard techniques for *in vitro* metabolic activation (e.g., rodent liver S9) might be inadequate; the structure or known activity of the compound indicates that the use of other test methods/systems might be appropriate.

5.3. Evaluation of results obtained from in vivo tests

In vivo tests have the advantage of taking into account absorption, distribution and excretion, which are not factors in *in vitro* tests, but are potentially relevant to human use. In addition metabolism is likely to be more relevant *in vivo* compared to the systems normally used *in vitro*. If the *in vivo* and *in vitro* results do not agree, then the difference should be considered/explained on a case-by-case basis, e.g., a difference in metabolism; rapid and efficient excretion of a compound *in vivo*.

In vivo genotoxicity tests also have the potential to give misleading positive results that do not indicate true genotoxicity. As examples:

(i) Increases in micronuclei can occur without administration of any genotoxic agent, due to disturbance in erythropoiesis (Tweats *et al.*, 2007 I);

(ii) DNA adduct data should be interpreted in the light of the known background level of endogenous adducts;

(iii) Indirect, toxicity-related effects could influence the results of the DNA strand break assays (e.g., alkaline elution and Comet assays).

Thus it is important to take into account all the toxicological and haematological findings when evaluating the genotoxicity data (note 15). Indirect effects related to toxicological changes could have a safety margin and might not be clinically relevant.

5.4. Follow-up strategies for positive results

5.4.1. Follow-up to findings in vitro in mammalian cell tests

The following discussion assumes negative results in the Ames bacterial mutation assay.

5.4.1.1. Mechanistic/in vivo follow-up

When there is insufficient weight of evidence to indicate lack of relevance, recommended follow-up for positive mammalian cell assays would be to provide experimental evidence, either by additional *in vitro* studies (i, below) *or* by carrying out two appropriate *in vivo* assays (ii, below), as follows:

i. Mechanistic information that contributes to a weight of evidence for a lack of relevant genotoxicity is often generated *in vitro*, for example evidence that a test compound that induces chromosome aberrations or mutations in the MLA is not a DNA damaging agent (e.g., other negative mutation/DNA damage tests in addition to the Ames test; structural considerations), or evidence for an indirect mechanism that might not be relevant *in vivo* or might have a threshold (e.g., inhibition of DNA synthesis, reactive oxygen species produced only at high concentrations) (Galloway *et al.*, 1998; Scott *et al.*, 1991; Müller and Kasper, 2000). Similar studies can be used to follow up a positive result in the *in vitro* micronucleus assay, or in this case evidence can include a known mechanism that indicates chromosome loss/aneuploidy, or centromere staining experiments (note 17) that indicate chromosome loss. Polyploidy is a common finding in chromosome aberration assays *in vitro*. While aneugens can induce polyploidy, polyploidy alone does not indicate aneugenic potential and can simply indicate cell cycle perturbation; it is also commonly associated with increasing cytotoxicity. If polyploidy, but no structural chromosome breakage, is seen in an *in vitro* assay, generally a negative *in vivo* micronucleus assay with assurance of appropriate exposure would provide sufficient assurance of lack of potential for aneuploidy induction.

If the above mechanistic information and weight of evidence supports the lack of relevant genotoxicity, only a single *in vivo* test with appropriate evidence of exposure is called for in order to establish the lack of genotoxic activity. This is typically a cytogenetic assay, and the micronucleus assay *in vivo* is called for when following up potential for chromosome loss.

If there is not sufficient weight of evidence or mechanistic information to rule out relevant genotoxic potential, two *in vivo* tests are generally called for, with appropriate endpoints and in appropriate tissues (usually two different tissues), and with an emphasis on obtaining sufficient exposure in the *in vivo* models.

Or

ii. Two appropriate *in vivo* assays are done, usually with different tissues, and with supporting demonstration of exposure.

In summary, negative results in appropriate *in vivo* assays, with adequate justification for the endpoints measured and demonstration of exposure (see section 4.4.1) are considered sufficient to demonstrate absence of significant genotoxic risk.

5.4.1.2. Follow-up to an in vitro positive result that is dependent upon S9 activation

When positive results are seen only in the presence of the S9 activation system, it should first be verified that metabolic activation is responsible and not some other difference in conditions (e.g., low or no serum in the S9 mix, compared with ≥10% serum in the non-activated incubations). The follow-up strategy is then aimed at determining the relevance of the results *in vitro* to conditions *in vivo*, and will generally focus on *in vivo* studies in liver (note 18).

5.4.2. Follow-up to a positive *in vivo* micronucleus assay

If there is an increase in micronuclei *in vivo*, all the toxicological data should be evaluated to determine whether a non-genotoxic effect could be the cause or a contributing factor (note 15). If non-specific effects of disturbed erythropoiesis or physiology (such as hypo/hyperthermia) are suspected, an *in vivo* assay for chromosome aberrations might be more appropriate. If a “real’ increase is suspected, strategies should be used to demonstrate whether the increase is due to chromosome loss or chromosome breakage (note 17). There is evidence that aneuploidy induction, e.g., with spindle poisons, follows a non-linear dose response. Thus, it might be possible to determine that there is a threshold exposure below which chromosome loss is not expected and to determine whether an appropriate safety margin exists compared with clinical exposure.

In conclusion, the assessment of the genotoxic potential of a compound should take into account the totality of the findings and acknowledge the intrinsic values and limitations of both *in vitro* and *in vivo* tests.

5.5. Follow-up genotoxicity testing in relation to tumour findings in a carcinogenicity bioassay

Additional genotoxicity testing in appropriate models can be conducted for compounds that were negative in the standard test battery but which have shown increases in tumours in carcinogenicity bioassay(s) with insufficient evidence to establish a non-genotoxic mechanism. To help understand the mode of action, additional testing can include modified conditions for metabolic activation in *in vitro* tests or can include *in vivo* tests measuring genetic damage in target organs of tumour induction, such as DNA strand break assays (e.g., comet or alkaline elution assays), liver UDS test, DNA covalent binding (e.g., by ^{32}P-postlabelling), mutation induction in transgenes, or molecular characterization of genetic changes in tumour-related genes (Kasper *et al.*, 2007).

6. Notes

1. The *in vitro* micronucleus assay has been widely evaluated in international collaborative studies (Kirsch-Volders *et al.*, 2003), is validated by ECVAM (Corvi *et al.*, 2008), and is the subject of an OECD guideline 487 (2010).
2. There is a small but significant number of genotoxic carcinogens that are reliably detected by the bone marrow tests for chromosomal damage but have yielded negative/weak/conflicting results in the *in vitro* tests outlined in the standard battery options. Carcinogens such as procarbazine, hydroquinone, urethane and benzene fall into this category. Some other examples from a survey of companies are described by Tweats *et al.*, 2007, II.
3. In principle, micronuclei in hematopoietic cells can be evaluated in bone marrow from any species, and in blood from species that do not filter out circulating micronucleated erythrocytes in the spleen. In laboratory mice, micronuclei can be measured in polychromatic erythrocytes in blood, and mature (normochromatic) erythrocytes can be used when mice are treated continuously for about 4 weeks or more. Although rats rapidly remove micronucleated erythrocytes from the circulation, it has been established that micronucleus induction by a range of clastogens and aneugens can be detected in rat blood reticulocytes (Wakata *et al.*, 1998; Hamada *et al.*, 2001). Rat blood can be used for micronucleus analysis provided methods are used to ensure analysis of the newly formed reticulocytes (Hayashi *et al.*, 2007; MacGregor *et al.*, 2006), and the sample size is sufficiently large to provide appropriate statistical sensitivity given the lower micronucleus levels in rat blood than in bone marrow (Kissling *et al.*, 2007). Whichever method is chosen, bone marrow or blood, automated or manual analysis, each laboratory should determine the appropriate minimum sample size to ensure that scoring error is maintained below the level of animal-to-animal variation.

   Some experience is now available for micronucleus induction in the dog and rhesus monkey (Harper *et al.*, 2007; Hotchkiss *et al.*, 2008). One example where such alternative species might be useful would be in evaluation of a human metabolite that was not sufficiently represented in rodents but was formed in the dog or monkey.
4. While the two options in the battery are equally suitable, specific knowledge about an individual test compound can indicate that one option is preferable. For example, if systemic exposure in animal models is equal to or less than anticipated clinical exposure, in vitro assays should be employed: Option 1 (see also sections 2.3.4. and 4.4.1). On the other hand Option 2, including a test in liver, is recommended in cases where short-lived reactive metabolites are expected to be generated in the liver.
5. Certain structurally alerting molecular entities are recognized as being causally related to the carcinogenic and/or mutagenic potential of chemicals. Examples of structural alerts include alkylating electrophilic centers, unstable epoxides, aromatic amines, azo-structures, *N*-nitroso groups, and aromatic nitro-groups (Ashby and Paton 1994). For some classes of compounds with specific structural alerts, it is established that specific protocol modifications/additional tests are important for optimum detection of genotoxicity (e.g., molecules containing an azo-group, glycosides, compounds such as nitroimidazoles requiring nitroreduction for activation, compounds such as phenacetin requiring a different rodent S9 for metabolic activation).
6. There is some experience with *in vivo* assays for micronucleus induction in skin and colon (Hayashi *et al.*, 2007), and DNA damage assays in these tissues can also be an appropriate substitute.

7. A few chemicals are more easily detectable either with plate-incorporation or with pre-incubation methods, though differences are typically quantitative rather than qualitative (Gatehouse *et al.*, 1994). Experience in the pharmaceutical industry where drugs have been tested in both protocols has not resulted in different results for the two methods, and, in the IWGT report (Gatehouse *et al.*, 1994), the examples of chemical classes listed as more easily detectable in the pre-incubation protocol are generally not pharmaceuticals and are positive in *in vivo* genotoxicity tests in liver. These include short chain aliphatic nitrosamines; divalent metals; aldehydes (e.g., formaldehyde, crotonaldehyde); azo dyes (e.g., butter yellow); pyrrolizidine alkaloids; allyl compounds (allylisothiocyanate, allyl chloride), and nitro (aromatic, aliphatic) compounds.

8. The rationale for a maximum concentration of 1 mM for *in vitro* mammalian cell assays includes the following: The test battery includes the Ames test and an *in vivo* assay. This battery optimizes the detection of genotoxic carcinogens without relying on any individual assay alone. There is a very low likelihood of compounds of concern (DNA damaging carcinogens) that are not detected in Ames test or *in vivo* genotoxicity assay, but are detectable in an *in vitro* mammalian assay only above 1 mM. Second, a limit of 1 mM maintains the element of hazard identification, being higher than clinical exposures to known pharmaceuticals, including those that concentrate in tissues (Goodman & Gilman, 2001), and is also higher than the levels generally achievable in preclinical studies *in vivo*. Certain drugs are known to require quite high clinical exposures for therapeutic effect, e.g., nucleoside analogs and some antibiotics. While comparison of potency with existing drugs can be of interest to sponsors, perhaps even above the 1 mM limit, it is ultimately the *in vivo* tests that determine relevance for human safety. For pharmaceuticals with unusually low molecular weight (e.g., less than 200) higher test concentrations should be considered.

9. Although some genotoxic carcinogens are not detectable in *in vitro* genotoxicity assays unless the concentrations tested induce some degree of cytotoxicity, DNA damaging agents are generally detectable with only moderate levels of toxicity (Greenwood *et al.*, 2004). As cytotoxicity increases, mechanisms other than direct DNA damage by a compound or its metabolites can lead to ‘positive’ results that are related to cytotoxicity and not genotoxicity. Such indirect induction of DNA damage secondary to damage to non-DNA targets is more likely to occur above a certain concentration threshold. The disruption of cellular processes is not expected to occur at lower, pharmacologically relevant concentrations.

   In cytogenetic assays, even weak clastogens that are known to be carcinogens are positive without exceeding a 50% reduction in cell counts. On the other hand, compounds that are not DNA damaging, mutagenic or carcinogenic can induce chromosome breakage at toxic concentrations. For both *in vitro* cytogenetic assays, the chromosome aberration assay and the *in vitro* micronucleus assay, a limit of about 50% growth reduction is considered appropriate.

   For cytogenetic assays in cell lines, measurement of cell population growth over time (by measuring the change in cell number during culture relative to control, e.g., by the method referred to as population doubling (PD; note 10), has been shown to be a useful measure of cytotoxicity, as it is known that cell numbers can underestimate toxicity. For lymphocyte cultures, an inhibition of proliferation not exceeding about 50% is considered sufficient; this can be measured by mitotic index (MI) for metaphase aberration assays and by an index based on cytokinesis block for *in vitro* micronucleus assays. In addition, for the *in vitro* micronucleus assay, since micronuclei are scored in the interphase subsequent to a mitotic division, it is important to verify that cells have progressed through the cell cycle. This can be done by use of cytochalasin B to allow nuclear division but not cell division, so that micronuclei can be scored in binucleate cells (the preferred method for lymphocytes). For cell lines, other methods to demonstrate cell proliferation, including cell population growth over time (PD) as described above, can be used (Kirsch-Volders *et al.*, 2003).

For MLA, appropriate sensitivity is achieved by limiting the top concentration to one with close to 20% Relative Total Growth (RTG) (10-20%) both for soft agar and for microwell methods (Moore *et al.*, 2002). Reviews of published data using the current criteria found very few chemicals that were positive in MLA only at concentrations with less than 20% RTG and that were rodent carcinogens, and convincing evidence of genotoxic carcinogenesis for this category is lacking. The consensus is that caution is appropriate in interpreting results when increases in mutation are seen only below 20% RTG, and a result would not be considered positive if the increase in mutant fraction occurred only at ≤ 10% RTG.

In conclusion, caution is appropriate in interpreting positive results obtained as reduction in growth/survival approaches or exceeds 50% for cytogenetics assays or 80% for MLA. It is acknowledged that the evaluation of cells treated at these levels of cytotoxicity/clonal survival can result in greater sensitivity but bears an increased risk of non-relevant positive results. The battery approach for genotoxicity is designed to ensure appropriate sensitivity without relying on single *in vitro* mammalian cell tests at high cytotoxicity.

To obtain an appropriate toxicity range, a preliminary range-finding assay over a broad range of concentrations is useful, but in the genotoxicity assay it is often critical to use multiple concentrations that are spaced quite closely (less than two–fold dilutions). Extra concentrations can be tested but not all concentrations need be evaluated for genotoxicity. It is not intended that multiple experiments be carried out to reach exactly 50% reduction in growth, for example, or exactly 80% reduction in RTG.

10. For *in vitro* cytogenetic assays it is appropriate to use a measure of relative cell growth to assess toxicity, because cell counts can underestimate toxicity (Greenwood *et al.*, 2004). Using calculated population doublings (see glossary) to estimate the 50% growth reduction level, it was demonstrated that the frequency of positive results with compounds that are not mutagenic or carcinogenic is reduced, while agents that act via direct interaction with DNA are reliably positive.

11. In certain cases it can be useful to examine chromosome aberrations at metaphase in lymphocytes cultured from test animals after one or more administrations of test compound, just as bone marrow metaphase cells can be used. Since circulating lymphocytes are not replicating, agents that require replication for their genotoxic effect (e.g., some nucleoside analogs) are not expected to be detected in this cell type. Because some lymphocytes are relatively long-lived, in principle there is the potential for accumulation of un-repaired DNA damage *in vivo* that would give rise to aberrations when the cells are stimulated to divide *in vitro*. The *in vivo* lymphocyte assay can be useful in following up indications of clastogenicity, but in general another tissue such as liver is a more informative supplement to the micronucleus assay in hematopoietic cells because exposure to drug and metabolite(s) is often higher in liver.

12. The inclusion of a second *in vivo* assay in the battery is to provide assurance of lack of genotoxicity by use of a tissue that is well exposed to a drug and/or its metabolites; a small number of carcinogens that are considered genotoxic gave positive results in a test in liver but

were negative in a cytogenetics assay *in vivo* in bone marrow. These examples likely reflect a lack of appropriate metabolic activity or lack of reactive intermediates delivered to the hematopoietic cells of the bone marrow.

Assays for DNA strand breaks, DNA adducts, and mutations in transgenes have the advantage that they can be applied in many tissues. Internationally agreed protocols are not yet in place for all the *in vivo* assays, although considerable experience and published data and protocol recommendations exist for DNA strand break assays (Comet and alkaline elution assays), DNA adduct (covalent binding) measurements, and transgenic rodent mutation assays, in addition to the UDS assay. For a compound that is positive in vitro in the MLA and induces predominantly large colonies, and is also shown not to induce chromosome breakage in an in vitro metaphase assay, an in vivo assay for mutation, such as a transgenic mouse mutation assay, should be considered in preference to a DNA strand break assay. The UDS assay is considered useful mainly for compounds that induce bulky DNA adducts or are positive in the Ames test. Because cytotoxicity induces DNA strand breakage, careful cytotoxicity assessment is needed to avoid confounding the results of DNA strand break assays. This has been well characterized for the *in vitro* alkaline elution test (Storer *et al.*, 1996) but not yet fully validated for the Comet assay. In principle the DNA strand break assays can be used in repeat-dose toxicology assays with appropriate dose levels and sampling times.

Since liver of mature animals is not a highly mitotic tissue, often a non-cytogenetic endpoint is used for the second assay, but when dividing hepatocytes are present, such as after partial hepatectomy, or in young rats (Hayashi *et al.*, 2007), micronucleus analysis in liver is possible, and detects known genotoxic compounds.

13. For common vehicles like aqueous methyl cellulose this would usually be appropriate, but for vehicles such as Tween 80, the volume that can be administered could be as much as 30 fold lower than that given acutely.

14. Caution is appropriate if the toxicological study design includes additional blood sampling, e.g., for measurement of exposure. Such bleeding could perturb the results of micronucleus analysis since erythropoiesis stimulated by bleeding can lead to increases in micronucleated erythrocytes.

15. Increases in micronuclei can occur without administration of any genotoxic agent, due to disturbance in erythropoiesis (such as regenerative anemia; extramedullary hematopoiesis), stress, and hypo- and hyperthermia (reviewed by Tweats *et al.*, 2007, I). In blood, changes in spleen function that affect clearance of micronucleated cells from the blood could lead to small increases in circulating micronucleated red blood cells.

16. Positive controls for either short-term or repeat dose genotoxicity studies:

    For micronucleus (and other cytogenetic) assays, the purpose of the positive control is to verify that the individuals scoring the slides can reliably detect increases in micronuclei. This can be accomplished by use of samples from periodic studies (every few months) of small groups of animals (one sex) given acute treatment with a positive control. For manual scoring such slides can be included in coded slides scored from each study. Positive control slides should not be obvious to readers based on their staining properties or micronucleus frequency. For automated scoring, appropriate quality control samples should be used with each assay.

    For other *in vivo* genotoxicity assays, the purpose of positive controls is to demonstrate reliable detection of an increase in DNA damage/mutagenicity using the assay in the chosen species, tissue and protocol. After a laboratory has demonstrated that it can consistently detect appropriate positive control compounds in multiple independent experiments, carrying out positive control experiments periodically is generally sufficient provided experimental conditions are not changed. However, currently it is considered that for the Comet assay concurrent positive controls are advisable.

17. Determination of whether micronucleus induction is due primarily to chromosome loss or to chromosome breakage could include staining micronuclei *in vitro* or *in vivo* to determine whether centromeres are present, e.g., using fluorescent in situ hybridization (FISH) with probes for DNA sequences in the centromeric region, or a labeled antibody to kinetochore proteins. If the majority of induced micronuclei are centromere positive, this suggests chromosome loss. (Note that even potent tubule poisons like colchicine and vinblastine do not produce 100% kinetochore positive micronuclei, but more typically 70 to 80%, and are accepted as primarily aneugens for assessing risk). An alternative approach is to carry out an *in vitro* or *in vivo* assay for metaphase structural aberrations; if negative this would imply that micronucleus induction is related to chromosome loss.

18. Standard induced S9 mix has higher activation capacity than human S9, and lacks phase two detoxification capability unless specific cofactors are supplied. Also, non-specific activation can occur *in vitro* with high test substrate concentrations (see Kirkland *et al.*, 2007). Genotoxicity testing with human S9 or other human-relevant activation systems can be helpful. Analysis of the metabolite profile in the genotoxicity test incubations for comparison with known metabolite profiles in preclinical species (in uninduced microsomes or hepatocytes, or *in vivo*) or in preparations from humans can also help determine the relevance of test results (Ku *et al.*, 2007), and follow-up studies will usually focus on *in vivo* testing in liver. A compound that gives positive results *in vitro* with S9 might not induce genotoxicity *in vivo* because the metabolite is not formed, is formed in very small quantities, or is metabolically detoxified or rapidly excreted, indicating a lack of risk *in vivo*.

7. Glossary

Alkaline elution assay: see *DNA strand break assay.*

Aneuploidy: numerical deviation of the modal number of chromosomes in a cell or organism.

Base substitution: the substitution of one or more base(s) for another in the nucleotide sequence. This can lead to an altered protein.

Cell proliferation: the ability of cells to divide and to form daughter cells.

Centromere/kinetochore: structures in chromosomes essential for association of sister chromatids and for attachment of spindle fibres that move daughter chromosomes to the poles and ensure inclusion in daughter nuclei.

Clastogen: an agent that produces structural breakage of chromosomes, usually detectable by light microscopy.

Cloning efficiency: the efficiency of single cells to form clones. It is usually measured after seeding low numbers of cells in a suitable environment.

Comet assay: see DNA strand break assay.

Culture confluency: a quantification of the cell density in a culture by visual inspection.

Cytogenetic evaluation: chromosome structure analysis in mitosis or meiosis by light microscopy or micronucleus analysis.

DNA adduct: product of covalent binding of a chemical to DNA.

DNA repair: reconstitution of the original DNA sequence after DNA damage.

DNA strand breaks: single or double strand scissions in the DNA.

DNA strand break test: alkaline treatment that converts certain types of DNA lesions into strand breaks that can be detected by the alkaline elution technique, measuring migration rate through a filter, or by the single cell gel electrophoresis or Comet test (in which cells embedded in a thin layer of gel on a microscope slides are subjected to electric current, causing shorter pieces of DNA to migrate out of the nucleus into a “Comet tail”). The extent of DNA migration is measured visually under the microscope on stained cells.

Frameshift mutation: a mutation (change in the genetic code) in which one base or two adjacent bases are added to (inserted in) or deleted from the nucleotide sequence of a gene. This can lead to an altered or truncated protein.

Gene mutation: a detectable permanent change within a single gene or its regulating sequences. The changes can be point mutations, insertions, or deletions.

Genetic endpoint: the precise type or class of genetic change investigated (e.g., gene mutations, chromosomal aberrations, DNA strand breaks, DNA repair, DNA adduct formation, etc).

Genotoxicity: a broad term that refers to any deleterious change in the genetic material regardless of the mechanism by which the change is induced.

Micronucleus: particle in a cell that contains nuclear DNA; it might contain a whole chromosome(s) or a broken centric or acentric part(s) of chromosome(s).

Mitotic index: percentage of cells in the different stages of mitosis amongst the cells not in mitosis (interphase) in a preparation (slide).

Numerical chromosome changes: chromosome numbers different from the original haploid or diploid set of chromosomes; for cell lines, chromosome numbers different from the modal chromosome set.

Plasmid: genetic element additional to the normal bacterial genome. A plasmid might be inserted into the host chromosome or form an extra-chromosomal element.

Point mutations: changes in the genetic codes, usually confined to a single DNA base pair.

Polychromatic erythrocyte: an immature erythrocyte in an intermediate stage of development that still contains ribosomes and, as such, can be distinguished from mature normochromatic erythrocytes (lacking ribosomes) by stains selective for RNA.

Polyploidy: Numerical deviation of the modal number of chromosomes in a cell, with approximately whole multiples of the haploid number. Endoreduplication is a morphological form of polyploidy in which chromosome pairs are associated at metaphase as “diplochromosomes”.

Population doubling or culture growth: This can be calculated in different ways; one example of an appropriate formula is: Population doublings (PDs) = the log of the ratio of the final count (N) to the starting (baseline) count (Xo), divided by the log of 2. That is: PD = [log(N ÷ Xo)] ÷ log 2.

Recombination: breakage and balanced or unbalanced rejoining of DNA.

RTG (relative total growth): This measure of cytotoxicity takes the relative suspension growth (based on cell loss and cell growth from the beginning of treatment to the second day post-treatment) and multiplies it by the relative plating efficiency at the time of cloning for mutant quantization.

Single Cell Gel Electrophoresis assay: Comet assay. *See DNA strand break assay*

Survival (in the context of mutagenicity testing): proportion of living cells among dead cells, usually determined by staining or colony counting methods after a certain treatment interval.

Transgene: an exogenous or foreign gene inserted into the host genome, either into somatic cells or germ line cells.

Unscheduled DNA synthesis (UDS): DNA synthesis that occurs at some stage in the cell cycle other than S-phase in response to DNA damage. It is usually associated with DNA excision repair.

8. Referemces

Ashby, J., D. Paton, "The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures," *Mutation Research* 286:3-74, 1994.

Corvi, R., S. Albertini, T. Hartung, S. Hoffmann, D. Maurici, S. Pfuhler, J. van Benthem, P. Vanparys, "ECVAM Retrospective Validation of the *in vitro* micronucleus test (MNT)", *Mutagenesis*, 23:271-283, 2008.

Gatehouse, D., S. Haworth, T. Cebula, E. Gocke, L. Kier, T. Matsushima, C. Melcion, T. Nohmi, T. Ohta, S. Venitt, E. Zeiger, "Report from the working group on bacterial mutation assays: International Workshop on Standardisation of Genotoxicity Test Procedures," *Mutation Research* 312: 217-233, 1994

Goodman & Gilman "The Pharmacological Basis of Therapeutics". J. G. Hardman, L E. Limbird, A. G. Gilman (Eds.), McGraw-Hill Professional, New York; 10th edition (August 13, 2001).

Greenwood, S.K., R.B. Hill, J.T. Sun, M.J. Armstrong, T.E. Johnson, J.P. Gara, S.M. Galloway, "Population Doubling: A simple and more accurate estimation of cell growth suppression in the *in vitro* assay for chromosomal aberrations that reduces irrelevant positive results," *Environmental and Molecular Mutagenesis*, 43:36-44, 2004.

Hamada S., S. Sutou, T. Morita, A. Wakata, S. Asanami, S. Hosoya, *et al.*, "Evaluation of the rodent micronucleus assay by a 28-day treatment protocol: summary of the 13th collaborative study by the collaborative study group for the micronucleus test (CSGMT)/Environmental Mutagen Society of Japan (JEMS)–Mammalian Mutagenicity Study Group (MMS)," *Environmental and Molecular Mutagenesis*, 37:93-110, 2001.

Hartmann A., E. Agurell, C. Beevers, S. Brendler-Schwaab, B. Burlinson, P. Clay, A. Collins, A. Smith, G. Speit, V. Thybaud, R.R. Tice, "Recommendations for conducting the in vivo alkaline Comet assay", *Mutagenesis* 18:45-51, 2003.

Hayashi, M., J.T. MacGregor, D.G. Gatehouse, I. Adler, D.H. Blakey, S.D. Dertinger, G. Krishna, T. Morita, A. Russo, S. Sutou, "*In vivo* Rodent Erythrocyte Micronucleus Assay. II. Some Aspects of Protocol Design Including Repeated Treatments, Integration With Toxicity Testing, and Automated Scoring," *Environmental and Molecular Mutagenesis*, 35:234-252, 2000.

Hayashi, M., J.T. MacGregor, D.G. Gatehouse, D.H. Blakey, S.D. Dertinger, L. Abramsson-Zetterberg, G. Krishna, T. Morita, A. Russo, N. Asano, H. Suzuki, W. Ohyama, D. Gibson, "*In vivo* erythrocyte micronucleus assay. III. Validation and regulatory acceptance of automated scoring and the use of rat peripheral blood reticulocytes, with discussion of non-hematopoietic target cells and a single dose-level limit test," *Mutation Research*, 627:10-30, 2007.

Heddle, J.A., S. Dean, T. Nohmi, M. Boerrigter, D. Casciano, G. R. Douglas, B.W. Glickman, N.J. Gorelick, J.C. Mirsalis, H-J Martus, T.R. Skopek, V. Thybaud, K.R. Tindall, N. Yajima, "In vivo transgenic mutation assays", *Environmental and Molecular Mutagenesis*, 35:253-259, 2000.

Hotchkiss, C. E. , M. E. Bishop, S. D. Dertinger, W. Slikker, M. M. Moore, J.T. MacGregor, Flow Cytometric Analysis of Micronuclei in Peripheral Blood Reticulocytes IV: An Index of Chromosomal Damage in the Rhesus Monkey (Macaca mulatta), *Toxicological Sciences* 102:352-358, 2008.

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

Kenelly, J.C., R. Waters, J. Ashby, P.A. Lefevre, B. Burlinson, D.J. Benford, S.W. Dean, I de G. Mitchell, "*In vivo* rat liver UDS assay. In: Supplementary Mutagenicity Tests, UKEMS Recommended Procedures, ed Kirkland D.J. and Fox M., Cambridge University Press, pp 52-77, 1993

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

Kirsch-Volders, M., T. Sofuni, M. Aardema, S. Albertini, D. Eastmond, M. Fenech, M. Ishidate, S. Kirchner, E. Lorge, T. Morita, H. Norppa, J. Surralles, A. Vanhauwaert, A. Wakata, "Report from the *in vitro* micronucleus assay working group", *Mutation Research*, 540:153-163, 2003.

Kissling, GE., S.D. Dertinger, M. Hayashi, J.T. MacGregor., "Sensitivity of the erythrocyte micronucleus assay: Dependence on number of cells scored and inter-animal variability", *Mutation Research* 634:235–240, 2007.

Ku, W.W., A. Bigger, G. Brambilla, H. Glatt, E. Gocke, P.J. Guzzie, A. Hakura, M. Honma, H-J. Martus, R.S. Obach, S. Roberts, "Strategy for genotoxicity testing-Metabolic considerations,"*Mutation Research*, 627:59-77, 2007.

MacGregor J.T., M.E. Bishop, J.P. McNamee, M. Hayashi, N. Asano, A. Wakata, M. Makajima, J. Saito, A. Aidoo, M.M. Moore, S.D. Dertinger, "Flow Cytometric analysis of micronuclei in peripheral blood reticulocytes: II. An efficient method of monitoring chromosomal damage in the rat", *Toxicological Sciences,* 94:92-107, 2006.

Moore, M.M. , M.Honma, J. Clements, K.Harrington-Brock, T.Awogi,G.Bolcsfoldi, M.Cifone, D.Collard, M.Fellows, K.Flanders, B.Gollapudi,P.Jenkinson,P.Kirby, S.Kirchner, J.Kraycer, S.McEnaney, W.Muster, B.Myhr, M.O'Donovan, J.Oliver, M-C.Ouldelhkim, K.Pant, R.Preston,C.Riach, R.San, H.Shimada, L.F. Stankowski, Jr. "Mouse Lymphoma Thymidine Kinase Locus Gene Mutation Assay: Follow-Up International Workshop on Genotoxicity Test Procedures, New Orleans, Louisiana, April 2000," *Environmental Molecular Mutagenesis,* 40, 292-299, 2002.

Moore, M. M., M. Honma, J. Clements, G. Bolcsfoldi, B. Burlinson, M. Cifone, J. Clarke, R. Delongchamp, R. Durward, M. Fellows, B. Gollapudi, S. Hou, P. Jenkinson, M. Lloyd, J. Majeska, B. Myhr, M. O'Donovan, T. Omori, C. Riach, R. San, L. F. Stankowski, Jr., A. K. Thakur, F. Van Goethem, S. Wakuri, I. Yoshimura, "Mouse Lymphoma Thymidine Kinase Gene Mutation Assay: Follow-up Meeting of the International Workshop on Genotoxicity Testing–Aberdeen, Scotland, 2003–Assay Acceptance Criteria, Positive Controls, and Data Evaluation," *Environmental and Molecular Mutagenesis*, 47:1-5, 2006.

Müller, L, P. Kasper, "Human biological relevance and the use of threshold arguments in regulatory genotoxicity assessment: Experience with pharmaceuticals." *Mutation Research,* 464: 9-34, 2000.

OECD Guidelines for Genetic Toxicology (1997), www.oecd.org/dataoecd.

Scott, D., S.M. Galloway, R.R. Marshall, M. Ishidate Jr., D. Brusick, J. Ashby, B.C. Myhr, "Genotoxicity under extreme culture conditions. A report from ICPEMC Task Group 9", *Mutation Research*, 257:147-204, 1991.

Storer, R.D., T.W. McKelvey, A.R. Kraynak, M.C. Elia, J.E. Barnum, L.S. Harmon, W.W. Nichols, J.G. DeLuca, "Revalidation of the *in vitro* alkaline elution/rat hepatocyte assay for DNA damage: improved criteria for assessment of cytotoxicity and genotoxicity and results for 81 compounds," *Mutation Research*, 368:59-101, 1996.

Suzuki, H., N. Ikeda, K. Kobayashi, Y. Terashima, Y. Shimada, T. Suzuki, T. Hagiwara, S.Hatakeyama, K.Nagaoka, J. Yoshida, Y. Saito, J. Tanaka, M. Hayashi, "Evaluation of liver and peripheral blood micronucleus assays with 9 chemicals using young rats. A study by the Collaborative Study Group for the Micronucleus Test (CSGMT)/Japanese Environmental Mutagen Society (JEMS) - Mammalian Mutagenicity Study Group (MMS)," *Mutation Research,* 583: 133-145, 2005

Thybaud, V., M. Aardema, J. Clements, K. Dearfield, S. Galloway, M. Hayashi, D. Jacobson-Kram, D. Kirkland, J. T. MacGregor, D. Marzin, W. Ohyama, M. Schuler, H. Suzuki, E. Zeiger, "Strategy for genotoxicity testing: Hazard identification and risk assessment in relation to *in vitro* testing," *Mutation Research*, 627:41-58, 2007.

Tweats, D. J., D. Blakey, R. H. Heflich, A. Jacobs, S. D. Jacobsen, T. Morita, T. Nohmi, M. R. O'Donovan, Y. F. Sasaki, T. Sofuni, R. Tice, "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," *Mutation Research*, 627:78-91, 2007, I.

Tweats, D. J., D. Blakey, R. H. Heflich, A. Jacobs, S. D. Jacobsen, T. Morita, T. Nohmi, M. R. O'Donovan, Y. F. Sasaki, T. Sofuni, R. Tice, "Report of the IWGT working group on strategy/interpretation of regulatory *in vivo* tests. II. Identification of *in vivo*-only positive compounds in the bone marrow micronucleus test," *Mutation Research*, 627:92-105, 2007, II.

Wakata, A., Y. Miyamae, S. Sato, T. Suzuki, T. Morita, N. Asano, T. Awogi, K. Kondo, M. Hayashi, "Evaluation of the Rat Micronucleus Test with Bone Marrow and Peripheral Blood: Summary of the 9th Collaborative Study by CSGMT/JEMS • MMS," *Environmental and Molecular Mutagenesis*, 32:84-100, 1998.