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 may be appropriate to base the evaluation only on *in vitro* testing. In some cases evaluation of genotoxic effects at the site of contact may 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 need replication. These tests are sufficiently well characterized and have sufficient internal controls that repetition of a negative assay is not usually needed. 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 may 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 assays

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 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 5000μg/plate. There is some evidence that dose-related genotoxic activity can be detected when testing certain compounds in the insoluble range in bacterial genotoxicity tests. On the other hand, heavy precipitates can interfere with scoring colonies or render the test compound unavailable to enter cells and interact with DNA. 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 is based on cytotoxicity as described below.

Limit of cytotoxicity:

In the bacterial reverse mutation test, the doses scored should show evidence of significant toxicity, but without exceeding a top dose of 5000 μg/plate. Toxicity may 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; TA100; TA1535; either TA1537 or TA97 or TA97a ; and either 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 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

fulfills 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 appropriate for this single experiment (note 7). Equivocal or weak positive results may indicate the need 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 (Kirsch-Volders et al 2003; Moore et al 2006). Several differences from these recommendations are noted here for testing pharmaceuticals, notably for selection of the top concentration, related to the maximum concentration, cytotoxicity and solubility (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 or 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 should be done by methods such as light microscopy, noting precipitate that persists, or appears during culture (by the end of treatment).

Cytotoxicity

Cytotoxicity should approach but not exceed a 50% reduction in cell growth (notes 9,10) for *in vitro* cytogenetic assays for metaphase chromosome aberrations or for micronuclei, or should approach a reduction of about 80% in RTG (relative total growth) for the mouse lymphoma *tk* mutation assay (note 9).

3.3.2 Study design/Test protocols

For the cytogenetic evaluation of chromosomal damage in metaphase cells *in vitro*, the test protocol includes the conduct of tests with and without metabolic activation, with appropriate positive and negative controls. Treatment with the test articles is 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 is needed 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, certain chemicals may 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) cells as a percentage of the number of metaphase cells. An elevated mitotic index (MI) or an increased incidence of polyploid cells may give an indication of the potential of a compound to induce aneuploidy. For the mouse lymphoma tk assay, the test protocol includes the conduct of tests with and without metabolic activation, with appropriate positive and negative controls, where the treatment with the test articles is for 3 to 4 hours. A continuous treatment without metabolic activation for approximately 24 hours is needed in case of a negative or equivocal result for both short treatments, with and without metabolic activation. An appropriate mouse lymphoma tk assay includes (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), are used. Following such testing, further confirmatory testing in the case of clearly negative or positive test results is not usually needed. Equivocal or weak positive results may require 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 for chromosome aberration and MLA assays can be confined to a positive control with metabolic activation (provided 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 appropriate for the detection of clastogens. Both rats and mice are appropriate for use in the bone marrow micronucleus test. Micronuclei may 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). Chromosomal aberrations can also be analyzed in peripheral lymphocytes cultured from treated rodents (note 11).

Note that when no *in vitro* mammalian cell assay is conducted, (Option 2), the micronucleus test *in vivo* is recommended, not the metaphase chromosome aberration assay, to include more direct capability for detection of chromosome loss (potential for aneuploidy).

4.2 Automated analysis of micronuclei

Systems for automated analysis (image analysis and flow cytometry) can be used if appropriately justified and validated (OECD, 1997; Hayashi et al 2000; 2007).

4.3 Other *in vivo* genotoxicity tests

The same *in vivo* tests described as the second test in the standard battery can be used as follow-up tests to develop weight of evidence in assessing results of *in vitro* or *in vivo* assays (notes 4 and 11). 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; while mutation can be measured in transgenes in rodents this entails prolonged treatment (e.g., 28 days). Thus the second *in vivo* assay will often evaluate a surrogate (DNA damage) endpoint. 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 *in vivo* transgenic mouse mutation assays and DNA covalent binding assays, (all of which may be applied in many tissues, note 4), in addition to the liver unscheduled DNA synthesis (UDS) assay.

4.4 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 by the acute protocol may generally be carried out in only one sex (note 12). For acute tests both sexes should be considered only if any existing toxicity/metabolism data indicate a substantial sex difference in the species being used. Otherwise, males alone are appropriate for acute genotoxicity tests. When the genotoxicity test is integrated into a repeat-dose toxicology study in two sexes, both sexes are scored except when there is no substantial sex difference evident in toxicity/metabolism, when a single sex may be scored. The dose levels for the sex(es) scored should meet the criteria for appropriate dose levels (sections 4.7.2 and 4.7.3).

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

4.5 Use of multiple administrations in genotoxicity assays *in vivo* and integration into toxicology studies

4.5.1 Sampling times

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 below).

When blood or bone marrow is used for micronucleus measurement in a multiweek study (e.g., 28 days), marked hematotoxicity may affect the ability to detect micronuclei, i.e., a dose that induces detectable increases in micronuclei after acute treatment may be too toxic to analyze after multiple treatments. It can be useful to obtain an additional sample blood on day 2 to 4 of dosing (Hamada et al, 2001); see section 4.7.3). The early sample can be used if needed to provide assurance that clastogens and potential aneugens are detected (but see note 13).

For other genotoxicity assays, sampling time is selected as appropriate for the endpoint measured; for example DNA damage/strand break measurements are usually made a few (e.g., 4) hours after the last administration.

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

4.5.2 Numbers of animals analyzed

The number of animals analyzed is determined by current recommendations for the micronucleus assay (OECD) or other genotoxicity assay and generally does not include all the animals treated for a toxicology study.

4.6 Route of administration

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

4.7 Dose selection for *in vivo* assays

4.7.1 Short-term studies

For short term (usually 1 to 2 administration) protocols, the top dose recommended for genotoxicity assays is a limit dose of 2000 mg/kg if this is tolerated, or maximum tolerated dose defined, for example for the micronucleus assay (OECD 474) 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 may also be taken into account in dose selection. Lower doses are generally spaced at approximately two to three fold intervals below this.

4.7.2 Multiple administration studies

In the Option 1 battery, when the *in vitro* mammalian cell assay is negative (or "non-relevant positive" ,(see section 5), if 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. However, 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 demonstrate that the top dose is appropriate for genotoxicity evaluation, as follows:

Recommendations for determining whether the top dose in a toxicology study (typically in rats) is appropriate for micronucleus analysis and for other genotoxicity evaluation (any one of the following):

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 14).

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

iii. Exposure:

   a. Plateau/saturation in exposure

   b. Accumulation

   Substantial reduction in exposure to parent drug with time (e.g., ³ 50% reduction from initial exposure) would usually disqualify the study. If this is seen in one sex, generally the sex with reduced exposure

would not be scored, unless there is enhanced exposure to a metabolite of interest.

iv Top dose is ³ 50% of the top dose that would be used for acute administration, (close to the minimum lethal dose) if such acute data are available for other reasons. (The top dose for acute administration micronucleus test 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.

If dose levels/exposure are not appropriate, acute *in vivo* assays should be performed to maximize exposure or obtain the appropriate toxicity range, (preferably conducting two genotoxicity assays in the same study) or an *in vitro* mammalian cell assay should be done if not already completed.

4.7.3 Additional guidance on dose selection for multiple administration studies

Compounds that induce aneuploidy, such as spindle poisons, are typically 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 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 may be required to ensure detection of aneugens and some toxic clastogens; these could be derived from any one of the following:

a. 2 -4 day blood sampling from the multiweek study before substantial hematotoxicity developed
b. an *in vitro* mammalian cell micronucleus assay
c. An acute bone marrow micronucleus assay

4.8 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.

4.8. 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. Bioavailability
   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. 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 may be needed 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".

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

4.8.2 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.

4.9 Use of positive controls for *in vivo* studies

For *in vivo* studies, it is not necessary to include concurrent treatments with positive controls in every study, after a laboratory has established competence in the use of the assay (note 15).

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, whereas a positive result in any assay for genotoxicity does not necessarily mean that the test compound poses a genotoxic/carcinogenic hazard to humans.

Although positive *in vitro* data may 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 many 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 historical control range for the testing facility

ii. Weak/equivocal responses that are not reproducible

If any of the above conditions apply 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 required.

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 may be attributable to a contaminant.

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

There are some well characterized examples of

artefactual increases in colonies that are not truly revertants. These may occur due to contamination with amino acids, (providing histidine for *Salmonella* strains or tryptophan for *E. 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 may 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 may 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. Conditions that 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 and no additional testing beyond the standard battery (option 1) with one negative *in vivo* test would be needed.

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) may be inadequate; the structure or known activity of the compound indicates that the use of other test methods/systems may 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., difference in metabolism; rapid and efficient excretion of a compound may occur *in vivo*, etc.)

In vivo genotoxicity tests also have the potential to give misleading positive results that do not indicate true genotoxicity. For example, increases in micronuclei can occur without administration of any genotoxic agent, due to disturbance in erythropoeisis (Tweats et al, 2007 I), DNA adduct data should be interpreted in the light of the known background level of endogenous adducts, and indirect, toxicity-related effects can 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 hematological findings when evaluating the genotoxicity data (note 17). Indirect effects related to toxicological changes may have a safety margin and may not to 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

To evaluate *in vitro* mammalian cell assay positive results for which there is insufficient weight of evidence to indicate lack of relevance, recommended follow-up for mammalian cell assays would be to provide experimental evidence, either by additional *in vitro* studies <u>or</u> by carrying

out two appropriate *in vivo* assays, 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/threshold mechanism not relevant *in vivo* (e.g., inhibition of DNA synthesis, reactive oxygen species produced only at high concentrations, etc, (Galloway et al, 1998; Scott et al, 1991; Muller 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 18) that indicate chromosome loss.

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

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

In summary, if the results of the *in vitro* mammalian cell assay are positive and there is not sufficient weight of evidence or mechanistic information to rule out relevant genotoxic potential, two *in vivo* tests are required, with appropriate endpoints and in appropriate tissues (usually two different tissues), and with an emphasis on obtaining sufficient exposure in the *in vivo* models.

Negative results in appropriate *in vivo* assays, with adequate justification for the endpoints measured, and demonstration of exposure (see section 4.8.1) is sufficient to demonstrate absence of genotoxic activity.

5.4.1.2 Follow-up to an *in vitro* positive result that is dependent upon S-9 activation

When positive results are seen only in the presence of the S-9 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 S-9 mix, compared with ³10% serum in the non-activated incubations). The follow-up strategy is then aimed at determining the relevance of any reactive metabolites produced *in vitro* to conditions *in vivo*, and will generally focus on *in vivo* studies in liver (note 16).

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 may be the cause or a contributing factor (note 17). If non-specific effects of disturbed erythropoeisis or physiology (such as hypo/hyperthermia) are suspected, an *in vivo* assay for chromosome aberrations may be more appropriate. If a "real' increase is suspected, strategies would be needed to demonstrate whether the increase is due to chromosome loss or chromosome breakage (note 18). There is evidence that aneuploidy induction, e.g., with spindle poisons, follows a non-linear dose response. Thus, it may

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 tumor findings in a carcinogenicity bioassay

Additional genotoxicity testing in appropriate models may be conducted for compounds that were negative in the standard test battery but which have shown increases in tumors 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 tumor-related genes (Kasper IWGT).

6. NOTES

1. The *in vitro* micronucleus assay has been widely evaluated in international collaborative studies (Kirsch-Volders et all, 2003), is considered validated by ECVAM (Corvi et al, 2008), and an OECD guideline is in preparation.

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 hematopoeitic cells may be evaluated in bone marrow from many 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 may 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 minimum sample size required 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. 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.

4. 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 mutation 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 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. 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 alkaline elution assay (Storer et al, 1996) but not yet fully validated for the Comet assay. In principle the DNA strand break assays may 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 with special protocols, or in young rats, (Suzuki et al 2005) micronucleus analysis in liver is possible, and detects known genotoxic compounds.

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, liver 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, IWGT, 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 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. Viewing the battery as a whole means that it is not necessary to detect in the mammalian cell assay every compound considered to be a genotoxic carcinogen. There is a low likelihood of such 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's, 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, e.g., nucleoside analogs and some antibiotics. While comparison of potency with existing drugs may 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.

9. Although some genotoxic carcinogens are not detectable in *in vitro* genotoxicity assays unless the concentrations tested induce some degree of cytotoxicity, particularly when measured by colony forming assays, DNA damaging agents are generally detectable with only moderate levels of toxicity (e.g., 30% reduction in growth measured at the time of sampling in the chromosome aberration assay, 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 are 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 the *in vitro* micronucleus assay, a limit of about 50% is also 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 mitotic index (MI) not exceeding about 50% is considered sufficient. 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). Other methods to demonstrate cell proliferation, including cell population growth over time (PD) as described above, may be used for cell lines (Kirsch-Volders et al 2003).

For the mouse lymphoma assay, appropriate sensitivity is achieved by limiting the top concentration to one with close to 20% Relative Total Growth (RTG) both for soft agar and for microwell methods (IWGT). Reviews of published data using the current criteria described by Moore et al (2006) 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 (Moore et al, 2006) is that caution is needed 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.

Caution is appropriate in interpreting positive results obtained as reduction in growth/survival approaches or exceeds 50% for cytogenetics assays or 80% for the mouse lymphoma assay. It is acknowledged that the evaluation of cells treated at these levels of cytotoxicity/clonal survival may 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 the need to rely 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 may be tested but not all 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. Cell growth assessment

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 to estimated 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 true DNA damaging agents are reliably positive.

11. In certain cases it may 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 may be used. 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 may 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 hematopoeitic cells because exposure to drug and metabolite(s) is often higher in liver.

12. Extensive studies of the activity of known clastogens in the acute mouse bone marrow micronucleus test have shown that in general male mice are more sensitive than female mice for micronucleus induction. Quantitative differences in micronucleus induction have been identified between the sexes, but no qualitative differences have been described. Where marked quantitative differences exist, there is invariably a difference in toxicity between the sexes. Thus males alone can be appropriate for acute *in vivo* micronucleus tests. When the assay is integrated into a repeat-dose toxicology study, because both sexes are usually available for study, samples can be collected from both sexes, and both sexes scored unless there is no substantial sex difference in toxicity/metabolism.

13. Caution is required 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 erythropoeisis stimulated by bleeding can lead to increases in micronucleated erythrocytes.

14. 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.

15. Positive controls

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 of small groups of positive control animals (one sex). For manual scoring such slides can be included in coded slides scored from each study, or used for periodic demonstration of ability of readers to recognize positive responses. 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 reliably detect appropriate positive control compounds in multiple independent experiments, it is no longer necessary to carry out concurrent controls with every assay using that protocol, but controls can be tested periodically.

16. Standard induced S-9 mix has higher activation capacity than human S-9, 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 S-9 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 S-9 may 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*.

17. Increases in micronuclei can occur without administration of any genotoxic agent, due to disturbance in erythropoeisis (such as regenerative anemia; extramedullary hematopoeisis), stress, hypo- and hyperthermia (reviewed by Tweats et al 2007I, IWGT). In blood, changes in spleen function that affect clearance of

micronucleated cells from the blood are expected to lead to increases in circulating micronucleated red blood cells.

18. 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%, but 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 infer that micronucleus induction is related to chromosome loss.

医薬品医療機器等レギュラトリーサイエンス総合研究事業研究班
平成10年度総合施設会

改訂遺伝毒性ガイドラインとその背景

本間　正充
国立医薬品食品衛生研究所・変異遺伝部

ICH Guidance Documents of S2A and S2B

(http://www.ich.org/cache/compo/276-254-1.html)

INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED TRIPARTITE GUIDELINE

GUIDANCE ON SPECIFIC ASPECTS OF REGULATORY GENOTOXICITY TESTS FOR PHARMACEUTICALS

S2A

Current Step 4 version

dated 16 July 1966

INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED TRIPARTITE GUIDELINE

GENOTOXICITY: A STANDARD BATTERY FOR GENOTOXICITY TESTING OF PHARMACEUTICALS

S2B

Current Step 4 version

dated 16 July 1997

Process of Revision of the Genotoxicity Guideline in ICH

June 2006 in Yokohama, Japan	The ICH Steering Committee (SC) agreed to initiate a revision of the genotoxicity guideline.
October 2006 in Chicago, USA	The ICH Expert Working Group (EWG) discuss revision and plan the approach to a revised guideline.
May 2007 In Brussels, Belgium	The EWG reached a consensus of the revised issues and agree to make a draft guideline according to the consensus.
October 2007 In Yokohama, Japan	Drafting new guideline for Step 2.

S2 EWG Members

Health Authorities

MHLW **Makoto Hayashi** (Rapporteur), NIHS
Masamitsu Honma, NIHS
Osamu Fueki, PMDA

EU **Peter Kasper**, BfArM (D)
J. W. van der Laan, RIVM (NL)

FDA **David Jacobson-Kram**, CDER

Industry

JPMA **Akihiro Wakata**, Astellas Pharma
Shigeki Sawada, Eisai Co.

EFPIA **Lutz Müller**, F. Hoffmann-La Roche
Veronique Thybaud, Sanofi-aventis

PhRMA **Jerry D. Frantz**, BMS
Sheila M. Galloway, Merck

Observer

EFTA	B.P. Schmid
Canada	T. Matula
WSMI	B. Steinfoff
IGPA	N.F. Cappuccino
BIO	M. L. Garriott

Background Issues Initiating Revision of the Current Genotoxicity Guideline

- **Too many positives in the *in vitro* mammalian cell assay systems that may not be relevant to human risk**
- **Taking into consideration of 3R's for genotoxicity assays whenever possible "without impacting" the scientific value of the tests and the evaluation of the human risk.**

3R's in animal experiments
- **Replacement**
- **Reduction**
- **Refinement**

ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Mutation Research 584 (2005) 1–256

Genetic Toxicology and Environmental Mutagenesis
www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens
I. Sensitivity, specificity and relative predictivity☆

David Kirkland[a,*], Marilyn Aardema[b], Leigh Henderson[c,1], Lutz Müller[d]

[a] Cerrance Laboratories Limited, Otley Road, Harrogate HG3 1PY, UK
[b] The Procter and Gamble Company, Mount Vallee Laboratories, PO Box 538707, Cincinnati, OH 45253-8707, USA
[c] Henderson Scientific Consultancy, Nantwich CW5 7LZ, UK
[d] Non-Clinical Drug Safety, F. Hoffmann-La Roche Ltd., PRBN-A Bldg 733/116, CH-4070 Basel, Switzerland

Received 3 November 2004; received in revised form 7 February 2005; accepted 25 February 2005

Performance of Individual Genotoxicity Tests in Detecting Rodent Carcinogens

Reports		Ames	MLA	Chrom. Ab.	MN (Vitro)	MN (Vivo)
Krikland et al., Mutat. Res. 584, 1, 2005	Sens.	58.8	73.1	65.6	78.7	
	Spec.	73.9	39.0	44.9	30.8	
Zeiger Reg. Tox. Pharm. 28, 85, 1998	Sens.	54.0	74.0	52.0		28.0
	Spec.	79.0	32.0	68.0		82.0

Sens. (Sensitivity): % of positive results among rodent carcinogens.
Spec. (Specificity): % of negative results among rodent non-carcinogens.

High Frequency of Positive Results of *in vitro* Mammalian Cell Genotoxicity Test

- The positive results are generally weak and not relevant under *in vivo* condition.
- The positive results are due to un-physiological experimental conditions (high/low pH, high osmolality, high cytotoxicity, insolubility), but not to be true genotoxicity.
- The positive results lead to a great deal of follow-up testing (*in vivo*) to assess whether there is any genotoxic risk.

↓

Need actions to reduce the high frequency of in vitro test and avoid the wasteful follow-up tests and use of animals.

Remedies for the Problems Related to Non-Relevant Positives in Mammalian Cell Tests

(proposed in Chicago Meeting)

Option 1
Keep testing approach as it is, but improve recommendations for WOE assessment and for follow-up testing.

Option 2
Reduction top concentration to reduce non-relevant results.

Option 3
Reduction cytotoxicity at top concentration to reduce non-relevant results.

Option 4
Remove in vitro mammalian tests from the battery without substitution (false negatives??)

Option 5
Remove in vitro mammalian tests from the battery, but require 2 in vivo assays.

Summary of Major Points of the Revisions

- S2A and S2B guidances merged into one
- Options provided for the test battery
  - Battery with *in vitro* mammalian cell assay
  - Battery without *in vitro* mammalian cell assay but two *in vivo* endpoints
- In vitro mammalian cell assay
  - Reduction in top concentration from 10 mM to 1 mM
  - Tightened acceptable cytotoxicity limits
  - No longer require testing of precipitating concentrations
- In vitro bacterial mutation assay no longer requires duplicate assay

Summary of Major Points of the Revisions, Continued

- Integration of genotoxicity endpoints into routine toxicology studies
  - Stringent criteria defined for acceptability of top dose
- Advice on choice of second *in vivo* genotoxicity endpoint
  - includes Comet assay, decreases emphasis on UDS assay)
- Provided advice on weight of evidence and data evaluation to determine relevance of positive findings

Two Options in the Revised Guideline

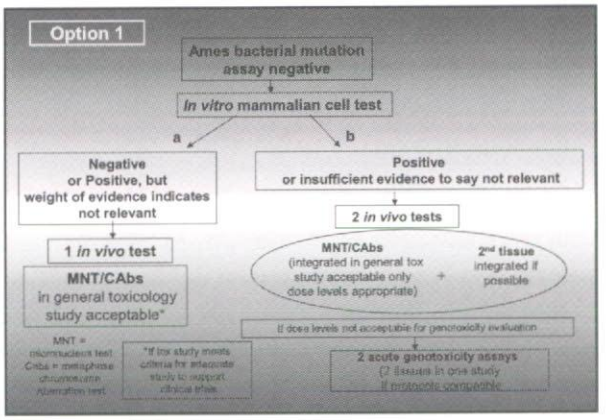
- Option 1: two *in vitro* tests only
  - in the past this led to many *in vivo* assays before FIM, to follow up positive *in vitro* results
- Option 2: one *in vitro* test (Ames) and two *in vivo* endpoints, preferably integrated into toxicology study

Revised Battery

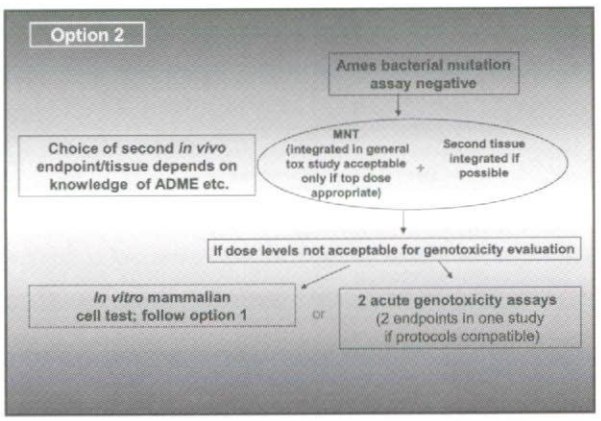
Current	Revisions proposed in Brussels May 2007	
	Option 1	Option 2
Ames microbial mutagenesis & repeat	Ames (one complete assay)	Ames (one complete assay)
In vitro mammalian cell assay: Chromosome aberrations, OR *tk* mutations in mouse lymphoma cells	*In vitro* mammalian cell assay (1 mM top conc): Chromosome aberrations, OR: *tk* mutations in mouse lymphoma cells OR: in vitro micronucleus assay	
In vivo cytogenetic assay	*In vivo* cytogenetic assay integrated *	*In vivo* cytogenetic assay (integrated if top dose appropriate) 2nd *in vivo* endpoint Integrated, if possible

* integrated provided tox study is adequate to support clinical trials and sampling within a day of last day of dosing

Option 1



Option 2



Criteria for Acceptable Dose/exposure in (sub)Chronic Study

- Maximum feasible dose
- Limit dose (1000 mg/kg for ≥14 days)
- Plateau/saturation in exposure/accumulation
  - ≥50% reduction from initial exposure would disqualify study
- Top dose ≥50% of top dose that would be used for acute administration
- Margin of exposure not sufficient

Benefits of Revisions: The 3 R's

- No longer require concurrent positive controls in every *in vivo* assay
- Integration of genotoxicity into toxicology assays
- Reduction in "non-relevant" *in vitro* results will reduce number of follow-up *in vivo* assays

Benefits of Revisions:

- Incorporates accumulated knowledge specific to testing of pharmaceuticals
- Takes advantage of new technologies
- More options in the test battery
- Reduction in delays caused by dealing with "non-relevant" *in vitro* positive genotoxicity results
- More efficient use of resources

S2 Revisions Allow for:

- Shorter development timelines
- More efficient use of resources, drug, and animals
- Better interpretation of genotoxic results while allowing continued safety for patients and volunteers

19

Next Step

January 28 2008 Teleconference
- final draft accepted by EWG for postal sign off (step 2)

厚生労働科学研究費補助金（医薬品・医療機器等レギュラトリーサイエンス総合研究事業）
平成19年度分担研究報告書

－抗悪性腫瘍薬の非臨床安全性試験方法等の国際的標準化に関する研究－

分担研究者：中江　　大（東京都健康安全研究センター　参事研究員）
協力研究者：小野寺博志（医薬品医療機器総合機構）
笛木　　修（医薬品医療機器総合機構）
浦野　　勉（医薬品医療機器総合機構）
込山　則行（医薬品医療機器総合機構）
佐神　文郎（エーザイ株式会社）
甲斐　修一（ブリストルマイヤーズ）
西村　千尋（日本化薬）

研究要旨

本研究は、日米EU医薬品規制調和国際会議（ICH）における抗悪性腫瘍薬の非臨床安全性試験方法に関するガイドライン策定を支援する目的で、国内の関係組織（独立行政法人医薬品医療機器総合機構[PMDA]および日本製薬工業協会[JPMA]）から研究協力者の参加を得て、関連情報の収集・解析を行い、併せてICHの場での議論に資するため国内の意思統一も図る。本年度は、米国食品医薬品局によるICH用ドラフトガイドラインおよびEUの欧州医薬品調査庁によるガイダンスを入手・和訳し、それらと平成16－18年度厚生労働科学研究において作成したガイドライン案を比較・検討した。その結果、日本のガイドライン案は抗悪性腫瘍薬に特異的な内容のみの記載に止める立場であったのに対し、米国のドラフトガイドラインは既存のガイドライン記載内容も採り入れて単独で成立させることを目指したものであり、EUのガイダンスはそれらの中間でやや日本寄りの内容であることが判明した。本研究グループは、この成果を基に、日本側の意思を統一してICHにおける議論に望み、平成19年10月のICH横浜会議より開始された抗悪性腫瘍薬の非臨床安全性試験方法ガイドライン策定に向けた専門家ワーキンググループ（S9 EWG）に参画した。S9 EWGは、平成20年６月のICHポートランド会議においてガイドライン骨子をまとめ、同11月のICHブリュッセル会議においてstep 2に達することを目標としている。本研究グループは、この議論に参加しつつ、関連する諸問題について随時検討を行っている。本年度は、S9 EWGの懸案事項のひとつである長期反復投与毒性試験の最長期間についてPMDA・JPMAの保有する過去の申請データを検討した結果、３ヶ月の試験で観察されなかった毒性が６ヶ月またはそれより長期の試験で観察される例のあることが判明した。本研究グループとしては、M3ガイドラインとの整合性を考慮する観点も含め、なお検討が必要ながら、抗悪性腫瘍薬の長期反復投与毒性試験の期間として３ヶ月間をもって最長としてもよい場合もあるが、臨床投与時に長期連続投与が想定されているような比較的毒性が弱いタイプの抗悪性腫瘍薬などを中心とした多くの場合に、３ヶ月間を越えるある程度の期間（日本側の主張としては６ヶ月間）の反復投与毒性試験が必要であると考える。以上の結果より、本

研究グループは、抗悪性腫瘍薬の非臨床安全性試験方法の国際的標準化について、日本として規制当局と製薬業界の意志を統一し、ICH S9 EWGにおける議論に参加し、その中で発生する諸問題について検討を行い、一定の成果を得たものである。S9ガイドラインは、なお骨子の策定途上であり、来年度以降も粛々と研究を進めていく予定である。

キーワード：抗悪性腫瘍薬、非臨床安全性試験、試験法ガイドライン、国際標準化、ICH S9 EWG

A．研究目的

抗悪性腫瘍薬の非臨床安全性試験方法については、日米両極とも正式に制定されたガイドラインが存在せず、EUに比較的古いガイダンスが存在している状態である。日本においては、平成16－18年度厚生労働科学研究費補助金の医薬品・医療機器等レギュラトリーサイエンス総合研究事業による「ワクチンや抗がん剤など特殊な成分の医薬品における非臨床安全性試験の実施手法等に関する研究」班においてガイドライン案(別添１）およびQ&A（別添２）が作成され、現在、それらの英語版と併せて公表が準備されている。本研究は、日米EU医薬品規制調和国際会議（ICH）における抗悪性腫瘍薬の非臨床安全性試験方法に関するガイドライン（S9ガイドライン）策定を支援する目的で、国内の関係組織（独立行政法人 医薬品医療機器総合機構[PMDA]および日本製薬工業協会[JPMA]）から研究協力者の参加を得て、関連情報の収集・解析を行い、併せてICHの場での議論に資するため国内の意思統一も図る。

本研究においては、さらに、必要に応じて、発がん性試験法などの改良と国際標準化に関しても対象に加える。

B．研究方法

本研究は、PMDAより小野寺 博志・笛木 修・浦野 勉・込山 則行 各氏、JPMAより佐神 文郎・甲斐 修一・西村 千尋 各氏に研究協力者として参加していただいて遂行した。

1．３極のガイドライン案/ガイダンスの検討：

本研究グループは、米国食品医薬品局（FDA）によるICH用ドラフトガイドライン（以下、米国ガイドライン案）およびEUの欧州医薬品調査庁（EMEA）によるガイダンス（以下、EUガイダンス）を入手・和訳し、それらと平成16－18年度厚生労働科学研究において作成したガイドライン案（以下、日本ガイドライン案）を比較・検討した。

2．ICHにおける抗悪性腫瘍薬の非臨床安全性試験方法ガイドライン策定に向けた専門家ワーキンググループ（S9 EWG）への参画：

本研究グループは、S9 EWGに参画し、そこでの議論に対応した。

3．S9 EWGでの議論に関連する諸問題の検討：

本研究グループは、S9 EWGの議論に参加しつつ、関連する諸問題について随時検索を行っている。本年度は、S9 EWGの懸案事項のひとつである長期反復投与毒性試験の最長期間についてPMDA・JPMAの保有する過去の申請データを検討した。

C．研究結果

1．３極のガイドライン案/ガイダンスの検討：

抗悪性腫瘍薬の非臨床安全性試験法について、FDAは、平成19年10月８日付にてICH用のガイドライン案を策定した。一方、EUにおいては、平成10年６月23日付のガイダンスがEMEAより公表されている。本研究グループは、S9 EWGにおける議論の開始に先立ってこれらの米国ガイドライン案（別添３）およびEUガイダンス（別添４）を入手して和訳し、日本ガイドライン案と比較・検討した。その比較結果は、表１－３に概要を記した通りである。

表1　抗がん剤の非臨床安全性試験に関する日米EUのガイドライン案ないしガイダンスの比較1

			日本ガイドライン案	日本ガイドライン案Q&A	参考：バイオ医薬品(S6)ガイドライン	EUガイダンス	米国ガイドライン案
1 緒言	1.1 背景		①より適正な試験が適切な時期に実施されることを目標に、抗がん剤の非臨床安全性試験の基本的考え方を示した				Ⅰ.Ⅱ. Introduction/Background ・抗がん剤開発における非臨床毒性、薬効試験が適切にデザインされるようにアシストすることが目的である ・ここでの推奨事項は最低限であり、多くの開発化合物はこれ以上の試験がなされるだろう ・本推奨事項はINDs, NDAs, BLAsに適用される ・他のガイダンスも参考にすること　開発の早期に当局に相談することを奨励する
	1.2 目的		①GL案は基本的考え方の提示であり、固守するよう求めたものではない　得られる結果が臨床上有用であれば新たな実験の実施は必要である		1)ヒトへの初回投与量と増量計画を設定すること 2)標的臓器の特定と変化の可逆性 3)臨床でのモニタリング項目の特定	本ガイダンス以外にも必要に応じて追加試験が必要	・本ガイダンスは、抗がん剤開発における多くの問題点を全て包括するものではない ・薬理学的、毒性学的特長が出せるような試験デザインが必要であり、薬剤あるいはヒト適用に関連した新たな特長を明確にするために標準的な非臨床試験をモディファイする必要もある ・IND開始にあたっての非臨床試験の目的は、1)適切な開始用量の設定、2)ヒトへの安全性に影響を与える所見の同定であり、更に3)薬理作用の評価、4)毒性標的器官の評価、5)用量-毒性相関の明確化、6)毒性の可逆性の評価である　また、7)臨床試験での最大用量の予測の手助けになる場合もある
	1.3 適用範囲		①本GL案は細胞毒性型、非細胞毒性型の区別なく抗がん剤全般に対応する ②バイオ医薬品も含めているが、薬剤の特性上「S6GL」を考慮した上で本GL案を参考にすること	①-1. 本GL案は、がんの進行、転移、延命や症状緩和に有効性を示す抗がん剤全てに適用される(Q2) ①-2. 既存の抗がん剤の効果を増強させるもの、副作用を軽減させるもの、がんの予防薬には適用されない(Q2) ①-3. アジュバント療法剤あるいはそれに適用拡大する薬剤は、本GL案に合致しない部分もある(Q2) ①-4. 併用療法を前提として開発する場合、併用投与による毒性試験は必要ない(Q5)　しかし、代謝や排泄に相互作用がある場合や共通の細胞や機能を標的とする場合など併用により相乗的な影響が想定される場合は、併用投与による毒性試験の実施が必要 ②-1. 遺伝子医薬品(アンチセンス、リボザイム等)も本GL案の適用範囲に含まれるが、動物にターゲットとなる特定遺伝子がない場合がある　動物種の適切性が重要(Q3) ②-2. がん細胞特異的な抗体に抗がん剤を結合させた抗体薬剤複合体では、1)結合させる抗がん剤、2)抗体、3)それらの複合体、それぞれについて毒性試験の必要性を明確にする(Q4)　結合させる抗がん剤が既承認薬の場合は、抗がん剤そのものの毒性試験は省略できる場合もある　複合体が安定な場合、複合体のみで毒性評価すればよい場合もある　複合体あるいは抗体については動物の適切性からS6GLに準じた対応が必要	(適用)細菌、酵母、昆虫、植物及び哺乳動物細胞を含む種々の発現系を用いて産生される医薬品に適用される　有効成分としてタンパク質、ペプチド、それらの誘導体及びそれらを構成成分とする製品 (除外)抗生物質、アレルゲンエキス、ヘパリン、ビタミン、細胞性血液成分、細菌もしくはウイルスワクチン、DNAワクチン、細胞治療、遺伝子治療 開発ステージ毎ではなくNDAまでに必要とされる試験の考え方を記載 非臨床安全性試験では、以下の点を考慮すること 1)適切な動物種の選択、2)歳、3)生理的状態、4)投与量、投与経路、投与方法等を含めた投与計画、5)使用条件下での被験物質の安定性 動物種の選択 適切な動物種とは、受容体又はエピトープが発現されていて、薬剤が薬理学的活性を示す種　正当な理由が示されれば、1種の適切な動物で良い　更に短期試験で2種の動物種が必要な場合でも、長期試験では1種の動物種のみで実施可能な場合もある(例えば、短期試験で2種の動物種が同様な毒性プロファイルを示した場合など)　適切な動物種が存在しない場合は、ヒト型受容体を発現させたトランスジェニック動物あるいはその動物にとっての相同タンパク質などを用いることも検討すべき 用法/用量の設定 投与経路、投与回数は臨床で予定される投与方法に可能な限り近い形にすべき　しかし、動物で有効成分の消失速度がヒトに比して速い場合、溶解性が低い場合など、これを補うために投与回数を臨床試験での投与計画に比べて増やすことも必要 投与量は、毒性用量及びNOAELを含み、用量-反応関係に関する情報が得られるように設定する 曝露評価 曝露量及び投与量に基ずく安全域を予測するため、IND前に適切な動物種での吸収、血中濃度及びクリアランスに関する情報がある程度得られている必要がある 動物種差に起因する薬物動態の差を知ることは、毒性評価上重要である　単回、反復投与の薬物動態、TK、組織分布は有用 代謝については、バイオ医薬品の場合、ペプチド、アミノ酸への分解である　その代謝経路は一般によく判っており、代謝試験の実施は必要ない	腫瘍細胞に直接的な作用を及ぼす細胞傷害及び細胞増殖抑制薬に適用 単剤治療の開発に焦点を絞っている	・本ガイダンスはがん患者を治療することを意図した全ての薬剤を対象としている(投与ルートに係わらず、また低分子及びBiologicsも含む) ・また、抗がん剤の副作用軽減を意図した薬剤、がんそのものではなく症状緩和を意図した薬剤も対象とする ・ただし、がん予防薬、健常人を対象とした試験、放射線治療薬、ワクチン、細胞/遺伝子治療は対象外である
2 基本的考え方	2.1 一般原則		①基本は「毒性試験法GL」や「M3GL」であり、それらとの考え方の差異について記載 ②TKについては言及せず　TKガイダンスに準拠する				Ⅳ. Nonclinical study recommendations A. General comments on study design 1. Route of administration ・臨床予定投与ルートと同じルートで実施すべき ・臨床投与ルートと同じルートでの実施が不可能な場合は、PKデータを含めた科学的妥当性があれば代替ルートで実施可　その場合、臨床での毒性が予見できるように全身あるいは局所曝露量が類似になるような投与ルートとする 2. Dose groups ・Dose-rangeは、毒性学的特長が質的量的に詳細に判るだけでなく、MTDが求められるように設定すべき ・通常、溶媒対照と用量反応曲線が明確になるように最低3用量で構成される ・最高用量では明確な毒性が見られるように、最低用量では何も見られないか最低限の毒性が見られるように設定する　毒性が弱い薬剤では、投与可能最大量か臨床予定用量あるいは曝露量の適切な倍数でよい ・少なくても1試験で毒性の可逆性を検討する回復期間を設定すべきである ・動物数は、非げっ歯類で少なくとも3匹/sex/群とし、更に対照群と高用量群では回復性のために2匹/sex/群とすべき ・一般的には両性で実施すべきで、片性での実施には妥当性が必要 3. Test species ・使用動物種については、適切な妥当性が必要 ・Biologicsでの適切な動物種の選択はAppendix 1に記述 4. Schedule ・投与スケジュールと期間に関する推奨例をTable 1に示した ・投与スケジュールは、予定されている臨床での投与スケジュールを反映させるべき　通常、非臨床と臨床で1:1となるように設定 ・動物でクリアランスが早い、あるいは中和抗体が産生される、受容体の飽和、低い溶解性などを補うために投与頻度をヒトより多くすることが必要な場合がある ・Continuous infusionの場合は、毒性試験でも同様の手法を採用すべき ・1日数回投与では、動物に2回/日での投与が望ましい
	2.2 がん患者を対象とするPⅠまで	2.2.1 単回投与毒性	①2種の哺乳動物で実施(注1,2) 注1:PⅠ初回投与量設定がげっ歯類の結果で科学的に設定可能であれば、非げっ歯類での試験は必要ない 注2:既に海外でPⅠなどのヒトでの投与実績がある場合は、それらの結果も併せて本邦での初回投与量を設定する ②少なくとも1種では雌雄での実施、概略の致死量を求める	①-1:PⅠ初回投与量の設定に用いる毒性指標は、単回投与でのLD10、STD10あるいは反復投与でのMTDやNOAELなど薬剤の特性、臨床試験計画からCase by Caseで判断すべき(Q6)　ただし、健康成人でのPⅠではNOAELを根拠として設定すべき ②-1. 副作用が強く休薬期間後に再投与するサイクル投与が想定される抗がん剤の場合、単回投与での毒性プロファイル(標的器官、DLT、安全域、回復性など)がヒトへの安全性を検討する上で有用な場合がある(Q7)　そのような場合は、薬剤の特性により適切な時期による血液検査、血液化学的検査、病理組織学的検査の実施を個別に判断する必要がある	用量と全身毒性又は局所毒性との関連性を明らかにするために、単回投与毒性試験の実施は有用　しかし、薬理試験又は動物モデルでの効力試験の一部を単回投与毒性としても可 安全性薬理のパラメータを設定することを考慮すること	重度の毒性症状又は死亡がでる用量での評価をげっ歯類で実施する マウスでMTDを求め、毒性と体表面積との関連性が線形であるかをラットで確認する　線形性がない場合、初回投与量は感受性の高い動物種に基づき決定させる 初回投与量は単回投与MTDの1/10、作用機序が新規の場合は、非げっ歯類でも概略のMTDを求める	6. Evaluation a. General test battery and evaluations ・一般毒性試験の標準protocolには、一般状態、体重、血液・生化学、器官重量、剖検、病理組織学的検査が含まれるべき　病理組織学的検査は、臨床開始用量算出に使われる動物種での試験の対照群と高用量群で実施すべき　高用量群で多くの死亡例が出た場合は、臨床開始用量の設定に用いた用量群の病理組織学的検査を実施する　高用量群の剖検あるいは病理組織学的検査で所見が見られた器官は全群で実施する　しかし、低、中用量群の検査結果は、INDまでには必須ではない b. Recovery groups ・INDまでに、少なくとも1試験で(通常、14から28日あるいはそれ以上の)休薬期間をとり、所見の可逆性、継続性を検討する　通常、対照群と高用量群(場合によっては中用量群にも)に設定する　なるべく長い試験で設定する c. Toxicity related to similar compounds ・類似の構造や機能(低分子とBiologics間でも)を持った薬剤は、その薬剤特有の毒性にも気をつける必要がある　ex. チュブリン阻害剤の多くは神経毒性をもつので、それらが検出できるように試験系を構成する必要がある d. Special considerations for biological products i. Immunogenicity testing ・Biologicsでの試験で重要な考慮点は、免疫原性の評価を含めることである　免疫反応(IgG抗体の存在で確認される)は薬効や毒性に大きく影響を与えることから、反復投与毒性評価の制限因子となっている ・不純物やadjuvant活性、glycosylationなどの修飾、蛋白凝集形成など多くの因子が免疫原性に影響を与えることから、物理化学的特性だけからは免疫原性の予測は不可能である　このことから非臨床や臨床での検討が必要とされる ・Cytokine release syndrom(infusion reaction)もBiologics投与に関連して発現する特長的な副作用である　非臨床試験でinfusion reactionを見出すことはほとんどない　しかし、もし見られた場合は、臨床開始用量やinfusion時間、頻度を決めるのに有用であろう　In Vitro assayとして、ヒトPBMCsでのcytokine release測定が予測データとして有用かもしれない ii. Tissue cross-reactivity ・組織交差反応性試験はmAb医薬品での未知の標的器官の検索や動物種における薬理反応の確認に用いられる　この試験常にPⅠ前に実施されるべきである　適切な組織交差反応性試験に関しては、FDA 1997 'Points to Consider'を参照のこと ・免疫組織学的試験での薬剤の組織分布は、ターゲット部位の評価に有用である