Guidance Document on Revisions to OECD Genetic Toxicology Test Guidelines

TABLE OF CONTENTS

1 GENERAL INTRODUCTION (PREAMBLE)

An Introduction document to the Genetic Toxicology Test Guidelines (TGs) was first published in 1987 (OECD, 1987). Following a global update of the Genetic Toxicology TGs, which was completed in 2015, the present Guidance Document is written to provide succinct and useful information to individuals new to genetic toxicology testing as well as experienced individuals wishing to obtain an overview of the recent changes that were made to the TGs during the recent round of revisions. It provides: (1) general background and historical information on the OECD genetic toxicology TGs; (2) a brief overview of the important types of genetic damage evaluated by these tests; and 3) a description of the specific tests for which new or revised TGs where developed in 2014-2015, and the issues and changes addressed therein.

1.1 General Background

Since the 1980s, the view on the relative importance of the various tests for which a TG exists, has changed. Simultaneously, there has been an increase in our knowledge of the mechanisms leading to genetic toxicity as well as the experience with the use of the tests. Our interpretation of test results has evolved, as has our identification of the critical steps in the different tests, as well as the strengths and weaknesses of the different tests. Moreover, it has become clear that tests which detect the types of genetic damage that can be transmitted (gene mutations, structural chromosomal damage and numerical chromosomal abnormalities) in mammalian cells, should be considered to be the most relevant for evaluating chemicals for their potential to induce mutation.

There have also been significant economic changes since the OECD genetic toxicology TGs were first established. The number of newly developed substances to be tested has increased; furthermore, the European Registration, Evaluation, Authorization and Restriction of Chemical substances (REACH) initiative requires the testing of an unprecedented large number of substances already in commerce, but not previously evaluated for genotoxicity. Consequently, testing must be more efficient, faster and cheaper; whereas, at the same time the quality, sensitivity and specificity of the tests should not suffer. In addition, most regulatory authorities have increased their commitment to a reduction in the use of animals in toxicology testing. For some genetic toxicology testing strategies the number of required tests has been reduced from several to 3, or even as few as 2, *in vitro* tests., In line with the basic principles of humane animal experimentation (Reduce, Refine, and Replace, *i.e.* 3Rs) it has been recommended that *in vitro* tests should be followed up with as few as possible *in vivo* tests. Regulatory authorities have established various ways to do this, including a prohibition of *in vivo* tests in the European Union (EU) directive for cosmetic ingredients. Importantly, an evaluation of new data supports significant reductions in animal use. This can be, accomplished for example, by the incorporation of some *in vivo* genetic toxicity endpoints into repeated dose toxicity tests, or by eliminating the need for concurrent positive controls in *in vivo* genetic toxicology tests, thus reducing the total number of animals used in evaluating a particular test substance.

At its 22nd meeting in March 2010, the OECD Workgroup of National Coordinators for Test Guidelines (WNT) formed an Expert Workgroup that would review all of the genetic toxicology

TGs and make decisions to, retire or update the various TGs, and to develop new TGs. Subsequently, taking advantage of experience with the tests, the TG revisions were made, which provided increased knowledge concerning the features of the various tests and the technical conduct of the tests. In addition, the revision process provided an opportunity to harmonize, as appropriate, the recommendations across all of the genetic toxicology TGs. This harmonization led, for instance, to a common approach to the interpretation of test results.

1.2 History and Status of TGs

The history and current status of the different TGs is summarized in Table 1. Since the last TG revision in 1997 new TGs have been adopted: TG 487 (*in vitro* mammalian micronucleus test) in 2010; TG 488 (transgenic rodent somatic and germ cell gene mutation assays) in 2011; TG489 (*in vivo* mammalian alkaline comet assay) in 2014; and finally TG 490 (*in vitro* gene mutation assays using the thymidine kinase (*tk*) locus [Mouse Lymphoma Assay (MLA) and TK6] approved in 2015. Because of the acceptance of a new TG (TG 490) that includes both the MLA and TK6, TG 476 was revised and updated and now includes only the *in vitro* mammalian cell tests using the hypoxanthine guanine phosphoribosyl transferase (*hprt*) locus and xanthine-guanine phosphoribosyl transferase transgene (*xprt*)

A decision to delete some TGs was made based on the observation that these tests are rarely used in the various legislative jurisdictions, and on the availability of newer tests showing a better performance for the same endpoint. Moreover, the assays conducted in mammalian cells are preferred because they are considered more relevant. TGs that were deleted include these: TG 477 (sex-linked recessive lethal test in *Drosophila melanogaster*); TG 480 (*Saccharomyces cerevisiae,* gene mutation test); TG 481 (*Saccharomyces cerevisiae,* mitotic recombination test); TG 482 (DNA damage and repair, Unscheduled DNA synthesis in mammalian cells *in vitro*); and, TG 484 (mouse spot test). In addition, TG 479 (*in vitro* sister chromatid exchange test in mammalian cells) was also deleted because of a lack of understanding of the mechanism(s) of action of the effect detected by the test.

The tests described in the deleted TGs should not be used for new testing, and are no longer a part of the set of OECD recommended tests. However, data previously generated from these deleted TGs can still be used in regulatory decisions. Therefore, the TGs will be available on the OECD public website (http://www.oecd.org/env/testguidelines - bottom section ("TGs that have been cancelled and/or replaced with updated TGs"), because it may be useful to consult these TGs in the context of the assessment of substances based on old study reports.

In addition, it is recognized that two tests have limitations that result in their being less widely used and less favoured by some regulatory authorities than in the past. These include TG 485, (the mouse heritable translocation test which requires 500 first generation males per dose level) and TG486 (the *in vivo* unscheduled DNA synthesis test). Although both of these tests fulfil most of the criteria for deletion, the decision was made to neither delete nor update these TGs because they were still viewed as having regulatory utility.

A decision was made not to update TG 471 (bacterial reverse mutation test) during this round of revisions.

Table 1: Current status of the Test Guidelines for genetic toxicology

TG	Title	Adopted	Revised	Deleted
	Recently Revised Test Guidelines			
473	*In vitro* mammalian chromosomal aberration test	1983	1997 / 2014	
474	Mammalian erythrocyte micronucleus test	1983	1997 / 2014	
475	Mammalian bone marrow chromosomal aberration	1984	1997 / 2014	
476	*In vitro* mammalian cell gene mutation test[1]	1984	1997 / 2015	
487	*In vitro* mammalian cell micronucleus test	2010	2014	
478	Rodent dominant lethal assay	1984	2015	
483	Mammalian spermatogonial chromosome aberration	1997	2015	
	Recently Added Test Guidelines			
488	Transgenic rodent somatic and germ cell gene	2011		
489	*In vivo* mammalian alkaline Comet assay	2014		
490	*In vitro* gene mutation assays using the *tk* locus	2015		
	Archived/Deleted Test Guidelines			
472	Genetic Toxicology: *Escherichia coli*, Reverse	1983		1997
477	Sex-linked recessive lethal test in *Drosophila*	1984		2013
479	*In vitro* sister chromatid exchange assay in	1986		2013
480	*Saccharomyces cerevisiae,* gene mutation assay	1986		2013
481	*Saccharomyces cerevisiae,* mitotic recombination	1986		2013
482	DNA damage and repair, Unscheduled DNA	1986		2013
484	Mouse spot test	1986		2013

[1]After the revision, TG 476 is only for the mammalian cell gene mutation test using the *hprt* or *xprt* locus

2 AIM OF GENETIC TOXICOLOGY TESTING

The purpose of genotoxicity testing is to identify substances that can cause genetic alterations in somatic and/or germ cells and to use this information in a variety of regulatory decisions. Compared to most other types of toxicity, genetic alterations may result in effects manifested after unusually long periods following exposure. Furthermore, the disease endpoint can be caused by DNA damage that occurs in a single cell at low exposures. Rather than destroying that cell, the genetic alteration can result in a phenotype that not only persists, but can be amplified, as the cell divides, creating an expanding group of dysfunctional cells within a tissue or organ. Genetic alterations in somatic cells may cause cancer if they occur in genes (*i.e.* proto-oncogenes, tumour suppressor genes and/or DNA damage response genes); alternatively, they may be responsible for a variety of other (non-cancer) genetic diseases (Erickson, 2010). Accumulation of DNA damage in somatic cells has also been related to degenerative conditions such as accelerated aging, immune dysfunction, cardiovascular and neurodegenerative diseases (Hoeijmakers *et al.*, 2009; Slatter and Gennery, 2010; De Flora and Izzotti, 2007; Frank, 2010). In germ cells, DNA damage is associated with spontaneous abortions, infertility or heritable damage in the offspring and/or subsequent generations resulting in genetic diseases.

2.1 Genetic Toxicology Endpoints

Two types of genetic toxicology studies are considered important: 1) those measuring direct, irreversible damage to the DNA that is transmissible to the next cell generation, (*i.e.* mutagenicity); and 2) those measuring early, potentially reversible effects to DNA or on mechanisms involved in the preservation of the integrity of the genome (genotoxicity). These terms are defined in the context of the OECD TGs, as follows.

Mutagenicity is a component of genotoxicity. Mutagenicity results in events that alter the DNA and/or chromosomal structure and that are passed to subsequent generations. Thus mutations include the following: (1) changes in a single base pairs, partial, single or multiple genes, or chromosomes; (2) breaks in chromosomes that result in the stable (transmissible) deletion, duplication or rearrangement of chromosome segments; (3) a change (gain or loss) in chromosome number (*i.e.* aneuploidy) resulting in cells that have not an exact multiple of the haploid number; and, (4) mitotic recombination.

Genotoxicity is a broader term. It includes mutagenicity (described above), and it includes DNA damage, which may or may not result in permanent alterations in the structure or information content in a cell or its progeny. Thus, tests for genotoxicity also include those tests that evaluate induced damage to DNA (but not direct evidence of mutation) via effects such as unscheduled DNA synthesis (UDS), DNA strand breaks and DNA adduct formation.

3 TEST GUIDELINES FOR GENETIC TOXICOLOGY

A full evaluation of a chemical's ability to induce the possible types of genetic damage involved in adverse human health outcomes (cancer, non-cancer diseases involving somatic cell mutation, and heritable disease) includes tests that can detect gene mutation, chromosomal damage and aneuploidy. To adequately cover all the genetic endpoints, one must use multiple tests (*i.e.*, a test battery); as no individual test can provide information on all endpoints. Complete assessment of genotoxic potential through the detection of gene mutations, structural chromosomal aberrations, and numerical chromosomal abnormalities can be achieved in a variety of ways. However, the selection of: 1) which tests to use, 2) how to combine them into test batteries, 3) whether to use them for initial screening or to follow up previously generated results, and 4) how to interpret the hazard identified (or not), or to make decisions about further testing or regulatory action, is beyond the purview of the OECD TGs and this document. Recommended batteries of tests are described in other regional or international regulatory documents for various each types of chemicals (*e.g.* Cimino, 2006a, 2006b; Eastmond *et al.*, 2009).

Indicator tests detect primary DNA damage (*i.e*. the first in the chain of events leading to a permanent change), but not the consequences of this genetic damage. They are called indicator tests because the measured endpoint does not always lead to mutation, a change that can be passed on to subsequent generations. The DNA damage measured in the comet assay, or the unscheduled DNA synthesis test, may lead to cell death, or it may initiate DNA repair, which can return the DNA either to its original state or result in mutation. When evaluating potential genotoxicants, more weight should be given to the measurement of permanent DNA changes

than to DNA damage events that are reversible. In general, indicator tests should not be used in isolation and a substance should not be considered mutagenic (or nonmutagenic) on the results of indicator tests alone. As long as these limitations are recognized, indicator tests can be useful for: 1) preliminary screening; 2) as part of *in vivo* follow up of *in vitro* positive results; 3) for mechanistic studies, *e.g.* for the detection of oxidative DNA damage; and, 4) as an exposure biomarker demonstrating that the test chemical, or its metabolic or reactive products, have reached a target tissue and can damage the DNA.

The guidance provided in the TGs has been developed specifically for the routine evaluation of test materials, in particular for hazard identification. When a test is being used for more detailed experimentation, or for other regulatory purposes, alterations to the test protocol may be necessary. For instance, if the goal is to conduct a more detailed dose response evaluation, perhaps at low doses/concentrations to assess whether there is a no-effect level, or to better define a point of departure, or to understand the response at particular levels of exposure, it is likely that a larger number of test concentrations/doses would be required, and/or a different strategy for concentration/dose selection (than indicated in the TGs) would be needed. If the goal is to evaluate whether a chemical that is a mutagen and a carcinogen is inducing specific tumors via a mutagenic mode of action, it would be desirable to tailor the concentration/dose selection and possibly the length of exposure and timing of sampling to optimally address the question(s) being addressed for that specific chemical. There are a number of references that provide additional information for designing experiments that go beyond simple hazard identification: (Cao *et al.*, 2014; Gollapudi *et al.*, 2013; Johnson *et al.*, 2014; MacGregor *et al.*, 2015a; MacGregor *et al.*, 2015b; Moore *et al.*, 2008; Parsons *et al*., 2013; and Manjanantha *et al*., 2015)

The individual TGs provide specific information describing the tests and the detailed recommendations for their conduct. The tests, for which there are TGs, are briefly discussed below. This section is divided into *in vitro* and *in vivo* tests and further divided based on the principal genetic endpoint detected by the test.

3.1 *In vitro* genetic toxicology tests

3.1.1 Tests for gene mutation.

TG 471: Bacterial Reverse Mutation Test. The bacterial reverse mutation test (commonly called the "Ames test") identifies substances that induce gene mutations that are point mutations (*i.e.*, base-pair substitutions and frameshift mutations resulting from small insertions and deletions). This test uses specific strains of two species of bacteria, *Salmonella typhimurium* and *Escherichia coli.* Each strain contains identified mutations in amino acid (*i.e*., histidine [His] or tryptophan [Tryp], respectively) biosynthesis gene as the reporter locus. Those mutations prevent bacterial growth in the absence of the amino acid in the growth medium. Exposure to mutagens may induce a second mutation (a reversion) that will restore the wild type DNA sequence and the functional capability of the bacteria to synthesize the essential amino acid, and, thus, to grow on medium without the required amino acid. Cells in which this second, function-restoring mutation (reversion) has occurred are called revertants. Consequently, the Ames test is termed a "reverse mutation test".

There is a panel of specific strains that is used for the bacterial mutation test, which are each sensitive to a different mechanism of mutation (*e.g.* base substitution at GC pairs, base substitution at AT pairs or a single base insertion or deletion). A positive result in any one strain is considered relevant, and positive results in additional strains do not necessarily increase the level of confidence in the mutagenic response. The strains that can be reverted by the test chemical, and, therefore, the types of mutation(s) induced by the test chemical, provide information on the chemical's mechanism of action.

An advantage of the bacterial test is the relatively large number of cells exposed (about 10^8) with a background mutant frequency that is both low and stable enough to allow a large dynamic range between the background and the highest mutant frequencies usually detected. This combination of wide dynamic range and stable background allows for relatively sensitive and reliable detection of compounds that induce a weak response.

S. typhimurium and *E. coli* are prokaryotic cells that differ from mammalian cells in factors such as cellular uptake, metabolism, chromosome structure and DNA repair processes. As such they may be less predictive of effects in for humans. There have been developments to automate the test and to minimize the amount of test substance required (Claxton *et al*., 2001; Fluckiger-Isler *et al*., 2004). While widely used for screening, the miniaturized versions of the Ames test have not been universally accepted as replacements for standard regulatory testing, although they are described in TG 471.

TG 476: *In vitro* mammalian cell gene mutation tests using the *hprt* or *xprt* genes. These *in vitro* mammalian cell gene mutation tests identify substances that induce gene mutations at the *hprt* (hypoxanthine-guanine phosphoribosyl transferase) or *xprt* (xanthine-guanine phosphoribosyl transferase) reporter locus. Unlike the Ames assay, this test is a forward mutation test because the mutation inactivates the function of the gene product rather than reversing a previous inactivating mutation. Gene mutations are evaluated as mutant colonies that can grow in medium containing a selective agent such as 6-thioguanine, a metabolic poison which allows only cells deficient in *hprt* to grow and form colonies. Because the *hprt* gene is on the X-chromosome in humans and rodents, only one copy of *hprt* gene is active per cell. Thus, a mutation involving only the single active *hprt* locus will result in a cell with no functional HPRT enzyme. The test can be performed using a variety of established cell lines. The most commonly used cells for the *hprt* test include the CHO, CHL and V79 lines of Chinese hamster cells, L5178Y mouse lymphoma cells, and TK6 human lymphoblastoid cells. The non-autosomal location of the *hprt* gene (X-chromosome) means that the types of mutations detected in this assay are point mutations, including base pair substitutions and frameshift mutations resulting from small insertions and deletions.

For the *xprt* assay, the *gpt* transgene codes for XPRT protein, which is a bacterial analogue of the mammalian HPRT protein. The only cells suitable for the *xprt* test are AS52 cells containing the bacterial *gpt* transgene (and from which the *hprt* gene was deleted). The autosomal location of the *gpt* locus allows the detection of certain genetic events, such as large deletions and loss of heterozygosity (LOH), not readily detected at the hemizygous *hprt* locus on X-chromosomes.

Both tests involve treating cells with the test substance, followed by an incubation period that provides sufficient time (termed the expression time) for the newly induced mutants to lose their functional HPRT enzyme. The cell population is cloned in the presence and absence of the selective agent 6-thioguanine for the enumeration of mutant cells and the measurement of cloning efficiency, respectively, in order to calculate a mutant frequency. This mutant selection can be performed either using petri dishes (for monolayer cultures) or microtiter culture plates (for suspension cell cultures). The soft agar cloning method has also been used successfully for the *hprt* assay in L5178Y mouse lymphoma cells (Moore and Clive, 1982).

The cell density in mutant selection culture plates should be limited in order to avoid metabolic co-operation between mutant and non-mutant cells, which would alter mutant selection. This is particularly important for cells growing in monolayer such as cultures of V79 or CHO cells (COM 2000), but is less of an issue for cells growing in suspension. A sufficient number of plates/cells must be evaluated in order to reach an adequate statistical power based on the spontaneous mutant frequency.

TG 490: *In vitro* mammalian cell gene mutation tests using the thymidine kinase gene. This new TG describes two distinct assays that identify substances that cause gene mutations at the thymidine kinase (*tk*) reporter locus. The two assays use two specific *tk* heterozygous cells lines: L5178Y $tk^{+/-}$3.7.2C cells for the mouse lymphoma assay (MLA) and TK6 $tk^{+/-}$ cells for the TK6 assay; these are forward mutation assays. The mouse lymphoma assay (MLA) and TK6 assay using the *tk* locus were originally described in TG 476. Since the last revision of TG476, the MLA Expert Workgroup of the International Workshop for Genotoxicity Testing (IWGT) has developed internationally harmonized recommendations for assay acceptance criteria and data interpretation for the MLA (Moore *et al.* 2003, 2006) and this new TG was written to accommodate these recommendations. While the MLA has been widely used for regulatory purposes, the TK6 has been used much less frequently. It should be noted that in spite of the similarity between the endpoints, the two cell lines are not interchangeable, and regulatory programs may validly express a preference for one over the other for a particular regulatory use. For instance, the validation of the MLA demonstrated its appropriateness for detecting not only gene mutation, but also, the ability of a test substance to induce structural chromosomal damage.

The autosomal and heterozygous nature of the thymidine kinase gene in the two cell lines enables the detection of cells deficient in the enzyme thymidine kinase following mutation from $tk^{+/-}$ to $tk^{-/-}$. This deficiency can result from genetic events that are compatible with cell survival while they affect the *tk* gene. Genetic events detected using the *tk* locus include both gene mutations (point mutations, frameshift mutations, small deletions) and chromosomal events (large deletions, chromosomal rearrangements and mitotic recombination). The latter events are expressed as loss of heterozygosity (LOH), which is a common genetic change of tumor suppressor genes in human tumorigenesis.

Tk mutants include normal growing and slow growing mutants. These are recognized as "large colony" and "small colony" mutants in the MLA, and as "early appearing colony" and "late appearing colony" mutants in the TK6 assay. Normal growing and slow growing mutants are scored simultaneously and differentiated by size and shape in the MLA. Normal growing and

slow growing mutants are scored at different incubation times in the TK6 assay. Scoring of slow growing colonies in the TK 6 assay requires cell refeeding with the selective agent and growth mediums (Liber *et al.*, 1989). Normal growing and slow growing mutants must be enumerated as separate mutant frequencies. Scoring of normal and slow growing mutants can give an indication as to whether the *tk* mutants resulted from point mutations and/or chromosomal mutations. Normal growing colonies are considered indicative (but not exclusively predictive) of chemicals inducing point and other small-scale mutations whereas slow growing colonies are predictive of chemicals that induce chromosomal damage. Slow growing colonies consist of cells that have suffered damage impacting genes adjacent to *tk*. Their doubling time is prolonged and thus the size of the colony is smaller than for a normal growing one.

The test involves treating cells with the test substance, providing sufficient expression time for the newly induced mutants to lose their functional *tk* enzyme. Then the cell population is cloned in the presence and absence of the selective agent triflurothymidine for the enumeration of mutant cells and the measurement of cloning efficiency, respectively, in order to calculate of a mutant frequency. This mutant selection can be performed using soft agar cloning medium in petri dishes or liquid medium in microwell culture plates.

3.1.2 Tests for chromosomal aberrations.

There are basically two types of endpoints that can be used to determine if a substance can cause chromosomal damage and/or aneuploidy: chromosomal aberrations and micronuclei. Both can be visualized under a microscope. Most chromosomal aberrations are not viable when, for example, the deficiency comprises essential gene(s) and, thus, they are not transmitted to daughter cells. Although micronuclei are visualized in cells following the first cell division, these are not retained in all subsequent generations. Based on many genetic studies of the chromosomal basis of heritable genetic effects in humans, and other species, it can be assumed that compounds able to induce chromosomal aberrations and micronuclei in those tests are also able to induce transmissible chromosome mutations (*e.g*. reciprocal translocations, stable translocations and aneuploidy) in humans.

TG 473: *In vitro* mammalian chromosomal aberration test. The *in vitro* chromosomal aberration test identifies substances that induce structural chromosomal aberrations (deletions and rearrangements) in cultures of established cell lines [*e.g*. CHO, V79, Chinese Hamster Lung (CHL/IU), TK6] or primary cell cultures, including human or other mammalian peripheral blood lymphocytes. Structural chromosomal aberrations may be of two types, chromosome or chromatid; depending on the mechanism of action; the chromatid-type is most often observed. Most chromosomal aberrations are observed only in metaphases of the first or second mitotic cell division, because cells containing these aberrations are lost in subsequent cell divisions. Individual cells are viewed by microscope and the information on the types of chromosomal aberrations seen in that cell is recorded. Chromosome mutations occur if DNA strand breaks are mis-repaired by non-homologous end joining. This repair system involves removal of a few nucleotides to allow somewhat inaccurate alignment of the two ends for rejoining followed by addition of nucleotides to fill in gaps.

Fluorescence *in situ* hybridization (FISH), or chromosome painting, can provide additional information through enhanced visualization of translocations that are not readily visible in the standard chromosomal aberration test; this technique is not required for hazard assessment.

The standard design of the chromosomal aberration test is not optimised for the detection of aneuploidy. An increased incidence of polyploidy or chromosome endoreduplication could suggest an induction of numerical chromosomal abnormalities. However, it does not necessarily indicate an aneugenic mechanism as these observations may also result from failed cytokinesis or from cytotoxicity.

TG 487: *In vitro* mammalian cell micronucleus test. The *in vitro* micronucleus test identifies substances that induce chromosomal breaks and aneuploidy. Micronuclei are formed when either a chromosome fragment or an intact chromosome is unable to migrate to a mitotic pole during the anaphase stage of cell division and is left out of the daughter nuclei. The test, thus, detects both structural chromosomal breaks (caused by clastogens) or numerical chromosomal abnormalities or chromosome loss (caused by aneugens). The use of FISH and centromere staining can provide additional mechanistic information and can help in differentiating clastogens (micronuclei without centromeres) from aneugens (micronuclei with centromeres).

The test can be conducted using cultured primary human or other mammalian peripheral blood lymphocytes and a number of cell lines such as CHO, V79, CHL/IU, L5178Y and TK6. There are other cell lines that have been used for the micronucleus assay (*e.g.* HT29, Caco-2, HepaRG, HepG2 and primary Syrian Hamster Embryo cells) but these have not been extensively validated, and the TG recommends that they be used only if they can be demonstrated to perform according to the described requirements.

The scoring of micronuclei is generally conducted in the first division daughter cells. Cytochalasin B can be used to block cytoplasm division/cytokinesis and generate binucleate cells during or after test substance exposure. This may be desirable, because it can be used to measure cell proliferation and allows the scoring of micronuclei in dividing cells only. The use of cytochalasin B is mandatory for mixed cell cultures such as whole blood cultures in order to identify the dividing target cell population; but for cell lines the test can be conducted either with or without cytochalasin B.

The scoring of micronuclei has a potential for automation, *i.e.,* flow cytometry or image analysis. Automated systems that can measure micronucleated cell frequencies include, but are not limited to, flow cytometers (Torous *et al.,* 2000; De Boeck *et al*., 2005; Dertinger *et al.,* 2011), image analysis platforms (Parton *et al.* 1996; Asano *et al.,* 1998), and laser scanning cytometers (Styles *et al.,* 2011).

The *in vitro* micronucleus test has been shown to be as sensitive as the chromosomal aberration test for the detection of clastogens, and has the additional advantage of detecting aneugenic substances (Corvi *et al.,* 2008). However, the *in vitro* micronucleus test detects only chromosome fragments, the visible chromosomal aberration test detects translocations and other complex chromosomal rearrangements that may provide additional mechanistic information.

3.2 *In vivo* genetic toxicology tests

3.2.1 Tests for gene mutations.

For the *in vivo* tests, gene mutations are measured in "reporter" genes that are not involved in carcinogenesis or other identified diseases; however, these "reporter" genes are assumed to be mutated through similar molecular mechanisms to mutations resulting in disease. Moreover, positive selection systems have been developed to select, visualize, and enumerate the clones/colonies resulting from mutant cells.

TG 488: Transgenic rodent somatic and germ cell gene mutation assays. The transgenic rodent gene mutation test identifies substances that induce gene mutations in transgenic reporter genes in somatic and germ cells. The test uses transgenic rats or mice that contain multiple copies of chromosomally integrated phage or plasmid shuttle vectors which harbour the transgenic reporter genes in each cell of the body, including germ cells. Therefore, mutagenicity can be detected in virtually all tissues of an animal that yields sufficient DNA, including specific site of contact tissues and germ cells. The reporter genes are used for detection of gene mutations and/or chromosomal deletions and rearrangements resulting in DNA size changes (the latter specifically in the *lacZ* plasmid and *Spi*$^-$ test models) induced *in vivo* by the test substances (OECD, 2009, OECD, 2011; Lambert *et al.*, 2005). Briefly, genomic DNA is extracted from tissues, transgenes are extracted from genomic DNA, and transfected into bacteria where the mutant frequency is measured using specific selection systems. The transgenes are genetically neutral in the animals, *i.e.*, their presence or alteration has no functional consequence to the animals that harbour them. These transgenes respond to treatment in the same way as endogenous genes in rats or mice with a similar genetic background, especially with regard to the detection of base pair substitutions, frameshift mutations, and small deletions and insertions (OECD, 2009). These tests, therefore, circumvent many of the existing limitations associated with the study of *in vivo* gene mutation in endogenous genes (*e.g.*, limited tissues suitable for analysis) that can be used to readily enumerate mutant cells). Because the target genes are functionally neutral, mutations can accumulate over time allowing increased sensitivity for detection of mutations when tissues receive repeated administrations of the test chemical. A 28-day treatment is recommended for somatic tissues.

DNA sequencing of mutants is not required, but it is often helpful to confirm that the mutational spectra or type of mutations seen following treatment are different from those found in the untreated animal/tissue, to calculate the frequency of the different specific types of mutations, and to provide mechanistic data. Sequencing is also used to estimate the amount of clonal expansion of the originally mutated cell to more accurately estimate the actual mutation frequency by adjusting the frequency of mutants detected by positive selection.

3.2.2 Tests for chromosomal damage

As described in Section 3.1.2, there is strong evidence for linking chromosomal aberrations with adverse human health outcomes.

TG 475: Mammalian bone marrow chromosomal aberration test. The mammalian bone marrow chromosomal aberration test identifies substances that induce structural chromosomal

aberrations in bone marrow cells. While rodents are usually used, other species may, in some cases, be appropriate, if scientifically justified. Structural chromosomal aberrations may be of two types, chromosome- or chromatid-type depending on the mechanism of action. The chromatid-type is more often observed. Chromosomal aberrations are observed only in metaphase of the first or second mitotic cell division because cells containing aberrations are usually lost in subsequent cell divisions.

Although chromosomal aberrations can potentially be measured in other tissues, TG 475 describes detection of effects in bone marrow cells, only. Because of this tissue limitation, the test may not provide useful information for some organ-specific substances. Individual cell information on the various types of chromosomal aberrations is visualized in individual cells using microscopy.

FISH can provide additional information through enhanced visualization of translocations that are not readily visible in the standard chromosomal aberration test; although, this technique is not required for general hazard assessment.

The standard design of this test is not optimized for the detection of aneuploidy. Polyploidy (including endoreduplication) could arise in chromosomal aberration tests *in vivo*. Although, increased incidence of polyploidy may be seen as an indication for numerical chromosomal abnormalities, an increase in polyploidy *per se* does not indicate aneugenic potential; rather, it may simply indicate cell cycle perturbation or cytotoxicity. Because of the nature of the damage, it can only be detected within days of its occurrence. Longer exposures, thus, do not increase the sensitivity of these tests.

TG 474: Mammalian erythrocyte micronucleus test. The mammalian erythrocyte micronucleus test identifies substances that induce micronuclei in erythroblasts sampled from bone marrow (usually measured in immature erythrocytes) or peripheral blood (measured in reticulocytes) of animals. Usually rodents are used, but other species (dogs, primates, humans) have been studied. When a bone marrow erythroblast develops into an immature erythrocyte (sometimes also referred to as a polychromatic erythrocyte, or reticulocyte) and then migrates into the peripheral blood, the main nucleus is extruded. Subsequently, any micronuclei that have been formed may remain behind in the cytoplasm. Thus, visualisation or detection of micronuclei is facilitated in erythrocytes because they lack a main nucleus.

Micronuclei may originate from acentric chromosomes, lagging chromosome fragments or whole chromosomes, and, thus, the test has the potential to detect both clastogenic and aneugenic substances. The use of FISH and centromere staining can provide additional mechanistic information, and help differentiate clastogens (resulting in micronuclei without centromeres) from aneugens (resulting in micronuclei with centromeres). Automated systems that can measure micronucleated erythrocyte frequencies include, but are not limited to, flow cytometers (Torous *et al.*, 2000; De Boeck *et al*., 2005; Dertinger *et al.*, 2011), image analysis platforms (Parton *et al.* 1996; Asano *et al.*, 1998), and laser scanning cytometers (Styles *et al.*, 2011).

Micronuclei can be measured in other tissues, provided that the cells have proliferated before tissue collection and can be properly sampled (Uno 2015a and b). However, this TG is restricted

to measurement of effects in the bone marrow or the peripheral blood because of the lack of validation of tests applied to other tissues. These limitations restrict the usefulness of the micronucleus test for the detection of organ-specific genotoxic substances. As with TG 475, because of the nature of the damage, it can only be detected within days of its occurrence. Longer exposures do not increase the sensitivity of these tests.

TG 478: Rodent dominant lethal assay. The rodent dominant lethal test identifies substances that induce genetic damage causing embryonic or fetal death resulting from inherited dominant lethal mutations induced in germ cells of an exposed parent, usually male rats or mice (Bateman, 1984; Generoso and Piegorsch, 1993) or, predominantly, in the zygote after fertilization (Marchetti *et al*, 2004). Usually male rats are treated; occasionally, females are treated; however, females appear less suitable in a system where fertilization of the eggs is essential and where embryonic death is evaluated (Green *et al.*, 1985).

Dominant lethality is generally a consequence of structural and/or numerical chromosomal aberrations, but gene mutations and toxic effects cannot be excluded. Because it requires a large number of animals, this test is rarely used. Since death of a conceptus is the event detected, the dominant lethal test does not necessarily assess a biological endpoint that reflects a potential health risk to future generations. While the endpoint is a reproductive health effect, the dominant lethal assay should be considered an indicator test for inherited congenital malformations. However, it is a reproductive health effect. Moreover, the majority of chemicals that are positive in the dominant lethal tests also are positive in the heritable translocation test (TG 485) and specific locus test (Yauk *et al*., 2015). These latter two tests do measure mutational events that affect the health of the offspring. Furthermore, chemicals that cause dominant lethality also cause F1 congenital malformations (i.e the viable equivalent of dominant lethality (Anderson *et al.*, 1998).

TG 483: Mammalian spermatogonial chromosomal aberration test. The spermatogonial chromosomal aberration test identifies substances that induce structural chromosomal aberrations in male germ cells and is predictive for the induction of heritable mutations, usually sexually mature Chinese hamsters or mice are used. Chromosomal aberrations in spermatogonial cells are readily observed in metaphases of the first or second mitotic cell division of spermatogenesis. Cytogenetic preparations for analysis of spermatogonia metaphases at 24 and 48 hr after exposure allows the analysis of chromosomal aberrations in spermatocytes. A measure of cytotoxicity, and thus of exposure of the target cells, can be obtained by measuring the ratio between spermatogonia metaphases to either meiotic metaphases or interphase cells. The standard design of the test is not suitable for detection of aneuploidy. Although, increased incidence of polyploidy may be seen as an indication for numerical chromosomal abnormalities, an increase in polyploidy *per se* does not indicate aneugenic potential because it can result from cell cycle perturbation or cytotoxicity.

FISH (or chromosome painting) can provide additional information through enhanced visualization of translocations and other rearrangements that are not readily visible in the standard chromosomal aberration test; this technique is not, however, required for general hazard identification.

TG 485: Mouse heritable translocation assay. The mouse heritable translocation assay identifies substances that induce structural chromosome changes in the first generation progeny of exposed males. The test is performed in mice, because of the ease of breeding and cytological verification. Sexually mature animals are used. The average litter-size of the strain should be greater than 8, and it should be relatively constant. The type of chromosome change detected in this test system is reciprocal translocation. Carriers of translocation heterozygotes and XO-females show reduced fertility. This method is used to select first generation progeny for cytogenetic analysis. Translocations are cytogenetically observed as quadrivalents, which are comprised of two sets of homologous chromosomes (or bivalents) in meiotic cells at the diakinesis stage of metaphase I of F1 male progeny. To analyze for translocation heterozygosity one of two possible methods is used: 1) fertility testing of first generation progeny; or 2) cytogenetic analysis of all male first generation progeny. Monitoring of the litter size of the F1 can provide indication that dominant lethality is also occurring. The mouse heritable translocation test requires a large number of animals and is consequently rarely used. Moreover, expertise for the performance of the mouse heritable translocation test is no longer readily available.

3.2.3 Indicator tests.

TG 486: Unscheduled DNA synthesis (UDS) test with mammalian liver cells *in vivo.* The unscheduled DNA synthesis (UDS) test identifies substances that induce DNA damage and subsequent repair (measured as unscheduled DNA synthesis *vs.* normal S-phase scheduled synthesis) in liver cells of animals, commonly rats. However, this test does not detect the mutagenic consequences of the unrepaired genetic damage. Accordingly, the UDS test may be an appropriate test to detect DNA damage after exposure to substances, that specifically target the liver and that were positive in the Ames test in the presence of metabolic activation. The test responds positively only to substances that induce the type of DNA damage that is repaired by nucleotide excision repair. The test is based on the incorporation of tritium-labelled thymidine into the DNA by repair synthesis after excision and removal of a stretch of DNA containing a region of damage.

To conduct the UDS, the compound is administered *in vivo* by the appropriate route, the liver cells are collected, generally by liver perfusion, and put into culture. The incorporation of tritium-labelled thymidine into the liver cell DNA is conducted *in vitro,* and this is scored following autoradiography.

The UDS test should not be considered as a surrogate test for a gene mutation test (Kirkland and Speit, 2008).

TG 489: *In vivo* mammalian alkaline comet assay. The comet assay identifies substances that induce DNA damage. An alternate name is the alkaline single-cell gel electrophoresis assay. Under alkaline conditions (> pH 13), the comet assay can detect single and double strand breaks in eukaryotic cells, resulting, for example, from direct interactions with DNA, alkali labile sites, or as a consequence of transient DNA strand discontinuities resulting from DNA excision repair. These strand breaks may be: 1) repaired, resulting in no persistent effect; 2) lethal to the cell; or

3) fixed as a mutation resulting in a permanent heritable change. Therefore, the alkaline comet assay detects primary DNA damage that may (or may not) lead to gene mutations and/or chromosomal aberrations.

The comet assay can be applied to any tissue of an animal from which good quality single cell or nuclei suspensions can be made, including specific site of contact tissues and germ cells (Tice *et al.,* 1990); cell division is not required. This makes the comet assay useful in assessing exposure to target tissues and possible target tissues, and it provides an indication as to whether the chemical or its metabolites can cause primary DNA damage in that tissue. This test does not detect aneuploidy, structural chromosomal damage, or mutation. Like many of the *in vivo* tests, it can be integrated into repeat dose toxicity studies designed for other purposes, but because the damage it measures usually does not persist, longer exposures do not result in increased sensitivity.

It should be noted that the standard alkaline comet assay as described in TG 489, this guideline is not considered appropriate to measure DNA strand breaks in mature germ cells (*i.e.,* sperm). Genotoxic effects may be measured in testicular cells at earlier stages of differentiation. However, as in males the gonads contain a mixture of somatic and germ cells, positive results in the whole gonad (testis) are not necessarily reflective of germ cell damage; nevertheless, positive results indicate that the test chemical has reached the gonad.

The alkaline comet assay is most often performed in rodents, although it can be applied to other species. Further modifications of the assay allow more efficient and specific detection of DNA cross-links or, certain oxidized bases (by addition of lesion-specific endonucleases). The test guideline does not include procedures for the conduct of these modifications of the test.

Fragmentation of the DNA can be caused not only by chemically-induced genotoxicity, but also during the process of cell death, *i.e.,* apoptosis and necrosis. It is difficult to distinguish between genotoxicity and apoptosis/necrosis by the shape of the nucleus and comet tail after electrophoresis, *e.g.,* by scoring "hedgehogs" (Guerard *et al.,* 2014; Lorenzo *et al,* 2013). Consequently, for positive results, it is recommended that tissue samples be collected for histopathological examination to determine if apoptosis/necrosis could have resulted in DNA breaks via a non-genotoxic mechanism.

4 OVERVIEW OF ISSUES ADDRESSED IN THE 2014-2015 REVISION OF THE GENETIC TOXICOLOGY TEST GUIDELINES

As indicated in the introductory paragraphs, the Expert Workgroup undertook an extensive revision of the genetic toxicology TGs including a comprehensive harmonization of recommendations across the TGs. Therefore, this Guidance Document provides an amplification of important issues considered to be important for test conduct and data interpretation and also an overview of the new recommendations.

4.1 Issues specific to *in vitro* TGs

4.1.1 Cells

In the revision to the *in vitro* TGs there is new guidance concerning the characterization and handling/culturing of cells that are used in the individual tests. For many of the widely used mammalian cell lines, a new cell repository has been recently established and stocked with cells that are as close as possible to the original source and available for distribution (Lorge *et al.*, xxxx in prep).

4.1.2 Cytotoxicity and selection of highest concentration for cytotoxic chemicals

For the *in vitro* assays, cytotoxicity is used as a primary means for selecting test concentrations. The *in vitro* assays are conducted using measurements of cytotoxicity that have been developed and are specific to the individual assays. Since the last revisions of the TGs, the importance of cytotoxicity and the possibility that biologically irrelevant positive results can be obtained at high levels of cytotoxicity has been recognized (REFS). The recommendations for measuring cytotoxicity and the appropriate levels of cytotoxicity are now clearly emphasized in the individual TGs and are summarized here. These changes were implemented to standardize interpretation of assay results and guide the conduct of testing to increase the reliability and acceptability of the data by providing clearer standards for measuring cytotoxicity and ensuring that the most appropriate limit concentration is used when testing cytotoxic materials. It should be noted that for *in vitro* assays, the treatment period is relatively short (generally 3 to 24 hours). It is not possible to conduct longer term *in vitro* exposures. This is because some of the endpoints, particularly the cytogenetic endpoints, do not accumulate with time, and the gene mutations (especially the slow growing/small colony *tk* mutants) decrease with time. Therefore, in order to detect genetic damage, the *in vitro* assays may utilize somewhat higher test concentrations (and therefore higher levels of cytotoxicity) than would be found in typical *in vivo* exposures. That is, the need to attain the recommended level of cytotoxicity is a consequence of the need to use very short exposures.

4.1.2.1 *in vitro cytogenetic assays*

The proper conduct of the *in vitro* cytogenetic assays requires assuring that the cells have, in fact undergone cell division. The reduction of cell proliferation is, thus, usually used to evaluate cytotoxicity. Two new measures of cytotoxicity for the *in vitro* cytogenetic assays, the Relative Increase in Cell Count (RICC) and Relative Population Doubling (RPD), have been developed and are now recommended in the revised TGs 473 and 487 for use with cell lines. Previously recommended methods such as Relative Cell Counts and optical evaluation of confluence or cell density are no longer recommended. It should be noted that RPD is thought to underestimate cytotoxicity in cases of long-term treatment as stated in the revised TGs. For the micronucleus assay, the Cytokinesis Blocked Proliferation Index (CBPI), or the replication index (RI) continue to be acceptable measures of cytotoxicity. Mitotic index continues to be recommended for the chromosomal aberrations assay when using primary cultures of lymphocytes for which, in contrast to immortalized cell lines, RPD and RICC may be cumbersome/impossible to measure.

The top level of cytotoxicity for assay acceptance has been more explicitly defined for the *in vitro* cytogenetic assays. It is now recommended that if the maximum concentration is based on cytotoxicity, the highest concentration should aim to achieve 55 ± 5% cytotoxicity using the recommended cytotoxicity parameters (*i.e.* reduction in RICC, RPD, CBPI, RI, or MI to 45± 5% of the concurrent negative control). It is emphasized that care should be taken in interpreting positive results only found in the higher end of this 55 ± 5% cytotoxicity range.

4.1.2.2 *In vitro gene mutation assays*

The *in vitro* gene mutation assays require that the cells grow through the expression phase of the assay and also during the cloning for mutant selection. Therefore, they can only be conducted using concentrations that are compatible with cell survival and proliferation. For the MLA, TG 490 now clearly articulates that only the Relative Total Growth (RTG), originally defined by Clive and Spector (1975) should be used as the measure for cytotoxicity. RTG was developed to take into consideration the relative (to the negative control) cell growth of the treated cultures during the treatment and expression periods and the cloning efficiency at the time of mutant selection. For the other *in vitro* gene mutations assays (TK6, *hprt* and *xprt*) the relative survival (RS) should be used. RS is the plating efficiency immediately after treatment and corrected to include any cell loss during treatment. That is, RS should not be based solely on the plating efficiency of those cells that survive the treatment. The appropriate calculations to correct for cell loss during treatment are included in the TGs. In addition, for assays using RS as the measure of cytotoxicity, the cells used to determine the cloning efficiency immediately after treatment should be a representative sample from each of the respective untreated and treated cell cultures. For the *in vitro* gene mutation assays, if the maximum concentration is based on cytotoxicity, the highest concentration should aim to achieve between 20 and 10% RTG for the MLA, and between 20 and 10% RS for the TK6, *hprt* and XPRT assays. The revised TGs indicate that care should be taken when interpreting positive results only found between 20 and 10% RTG/RS and a result would not be considered positive if the increase in MF occurred only at or below 10% RTG/RS.

4.1.3 Selection of highest concentration tested for poorly soluble and/or non-cytotoxic chemicals

In this revision of the *in vitro* TGs, new recommendations are made for chemicals that are poorly soluble and/or non-cytotoxic. For poorly soluble test chemicals that are not cytotoxic at concentrations below the lowest insoluble concentration, and even if cytotoxicity occurs above the lowest insoluble concentration, it is required to test at only one concentration producing turbidity or with a visible precipitate because artefactual effects may result from the precipitate. Turbidity or a precipitate visible by eye, or with the aid of an inverted microscope, should be evaluated at the end of the treatment with the test chemical. Although it is not specifically included in the TGs, care should be taken in interpreting a positive result that is only seen at the precipitating concentration.

Until the recent revision of the TGs, the recommended top concentration, in the absence of cytotoxicity/solubility issues, was 10 mM or 5000 µg/ml (whichever is lower). The 10 mM limit

was defined originally as a limit low enough to avoid artifactual increases in chromosome damage and/or mutations due to excessive osmolality and appeared high enough to ensure detection (Scott *et al.*, 1991). Based on data from a number of independent reports (Goodman and Gilman, 2002; Kirkland *et al.*, 2007, 2008; Parry *et al.*, 2010; Kirkland and Fowler, 2010; Morita *et al.*, 2012), there was unanimous agreement during the recent revision discussions that the top concentration could be lowered. The reduction should result in an improvement of the specificity of the tests without losing sensitivity. An analysis of the data set generated by Parry *et al.* (2010) suggests that 10 mM is still required to detect biologically relevant effects from lower molecular weight non-cytotoxic substances and that test sensitivity at 10 mM is more similar to 2000 than to 5000 µg/ml (Brookmire *et al.*, 2013). Based on extensive discussion, the decision was made that if toxicity and solubility are not limiting factors the combination 10 mM or 2000 µg/ml, whichever is lower, represents the best balance between mM and µg/ml concentrations. A document was prepared that details the analysis that was conducted and provides the rationale for this new recommendation for top concentration (in the absence of cytotoxicity or issues of solubility) (OECD, 2014c).

However, when the composition of the test substance is not defined [*e.g.* substance of unknown or variable composition, complex reaction products, biological materials (*i.e.* UVCBs), environmental extracts], the top concentration in the absence of sufficient cytotoxicity may need to be higher (*e.g.* 5 mg/ml) to increase the concentration of each of the components.

TG 471 was not updated in the current round of revisions. Therefore, the top concentration for the Ames test remains at 5000 µg/plate in the absence of cytotoxicity at lower concentrations or problems with pH or solubility.

4.1.4 Treatment duration and sampling time.

The treatment durations and sampling times for each of the *in vitro* assays is clarified in the revised/new TGs. For both the chromosomal aberration and micronucleus assays, the cells should be exposed for 3 to 6 hours without and with metabolic activation. The cells should be sampled for scoring at a time that is equivalent to about 1.5 normal cell cycle lengths after the beginning of treatment for the chromosomal aberration assay and 1.5 to 2.0 normal cell cycle lengths after the beginning of treatment for the micronucleus assay. In addition, an experiment should be conducted in which cells should be continuously exposed without metabolic activation until they are sampled at a time equivalent to about 1.5 normal cell cycle lengths for the chromosomal aberration assay and 1.5 to 2.0 normal cell cycle lengths for the micronucleus assay. The reason for the difference in sampling time for chromosomal aberration and micronucleus analysis is that more time is needed for the cells to divide so as to see micronuclei in the daughter cells. The gene mutation assays have different recommendations depending upon the locus used. For assays using the *hprt* or *xprt* gene, TG 476 indicates that 3 to 6 hours of exposure (both with and without metabolic activation) is usually adequate. For the *tk* gene (TG 490), 3 to 4 hours of exposure (both with and without metabolic activation) is usually adequate. There is a new recommendation for the MLA that if the short-term treatment yields negative results and there is information suggesting the need for a longer treatment (*e.g.* nucleoside analogs or poorly soluble substances) that consideration should be given to conducting the test

with a longer treatment (*i.e.* 24 hours without S9). Consistent with the 1997 version of TG 476, there, is however, no requirement for the MLA that the longer treatment be routinely conducted if the short treatment is negative. Following treatment, the newly induced gene mutations require time to lose the normal enzyme (HPRT, XPRT or TK) before they can be successfully recovered as selective agent resistant colonies. Therefore, the cells are cultured for a period of time that has been shown to provide for optimal phenotypic expression. For both *hprt* and *xprt* the recommendation is to allow a minimum of 7 to 9 days post treatment for expression. Newly induced *tk* mutants express much faster than *hprt* or *xprt* mutants and because the small colony/slow growing mutants have doubling times much longer than the rest of the cell population, their mutant frequency actually declines once they are expressed. Therefore, it is important that the recommended expression periods of 2 days (post treatment) for the MLA and 2 to 3 days (post treatment) for TK6 are followed.

4.1.5 Concentration selection and minimum number of test concentrations/cultures

The revised/new *in vitro* TGs include updated recommendations on the selection and minimum number of test cultures meeting the acceptability criteria (appropriate cytotoxicity, number of cells, appropriate background frequency, *etc*) that should be evaluated. There is also additional guidance identifying situations where it may be advisable to use more than the minimum number of concentrations. The decision was made to continue to recommend at least three analysable test concentrations for the cytogenetic assays, and four for the *in vitro* gene mutations assays. For all assays, the solvent and positive control cultures are to be performed in addition to the minimum number of test substance concentrations. It is now recommended that, while the use of duplicate cultures is advisable, either replicate or single treated cultures may be used at each concentration tested. For the cytogenetic assays (as discussed further in Section 4.3) the most important point is that the total number of cells scored at each concentration should provide for adequate statistical power. Therefore, the results obtained for replicate cultures at a given concentration should be reported separately, but they must be pooled for data analysis. For test substances demonstrating little or no cytotoxicity, concentration intervals of approximately 2 to 3 fold will usually be appropriate. Where cytotoxicity occurs, concentrations should be selected to cover the cytotoxicity range from that producing the top level of cytotoxicity recommended for the particular assay (see Section 4.1.2) and including concentrations at which there is moderate and little or no cytotoxicity. Many test substances exhibit steep concentration response curves. Accordingly, in order to cover the whole range of cytotoxicity or to study the concentration response in detail, it may be necessary to use more closely spaced concentrations and more than three/four concentrations. It may also be useful to include more than three/four concentrations when single cultures are used, or when it is necessary to perform a repeat experiment.

4.1.6 Metabolic Activation

While some substances are reactive and able to directly interact with the DNA and to exert their genotoxic and/or mutagenic effects, many need to be metabolized and transformed into the reactive metabolites that interact with DNA. Unfortunately, most commonly used cell lines lack the ability to metabolize substances. The most commonly used activating system is the S9 fraction prepared from the homogenized livers of rats pretreated with PCBs, or other agents,

which induce the P450 mixed function oxidase ("phase I") system. This practice arose due to the prevalence of studies available in the 1970s suggesting that oxidative metabolism of pro-mutagens to metabolites capable of nucleophilic covalent modification of DNA was the major source of concern. In the absence of practical alternatives, this remains the most common approach, particularly when screening compounds for which there are no preliminary data on which to base an alternative approach. Preparation of S9 liver homogenates from other animals is possible. Screening using mouse or hamster S9 was commonly performed, particularly in the older literature and continues to be acceptable substitutes for rat preparations. Human liver S9 preparations are sometimes used (Cox *et al.*, 2015). S9 fractions prepared from homogenates of other organs have been used, with kidney and lung being the most common reported in the literature. Metabolism other than phase I metabolism is required to metabolize some pro-carcinogens to the mutagenic form. For example, phase II sulfation of phase I oxidation products is known to activate some aromatic amines and alkenyl benzenes to carcinogenic metabolites. While the phase II enzymes are present in the S9 fraction, the co-factors required for phase II metabolism are generally not added because phase II metabolism is known to transform many phase I metabolites into non-mutagenic metabolites. Moreover, there are no widely accepted protocols for studying the genotoxicity of the products of other metabolic pathways. The use of metabolically competent cells (*e.g.* HepG2 liver cell lines) has been reported but there are no widely accepted protocols for their use. Use of non-standard metabolic systems is always appropriate provided scientific justification and appropriate control experiments are used.

4.2 Issues specific to *in vivo* TGs

4.2.1 Dose Selection

Measurement of toxicity is used for two objectives: (1) to better define the doses to be used, and (2) to demonstrate sufficient exposure of the target tissues.

4.2.1.1 *Limit Dose*

If toxicity and solubility are not limiting factors, and if genetic toxicity is not expected based on data from structurally related substances, then use of a single dose at the limit dose may be sufficient. This limit dose is 2000 mg/kg bw/day for a treatment period of < 14 days and 1000 mg/kg bw/day for a treatment period > 14 days. The limit dose was retained in the revised *in vivo* TGs in order to strike a balance between the need to prevent false negative results and the humane treatment of animals.

4.2.1.2 *Range-finding study*

Dose levels should be based on the results of a dose range-finding study measuring general toxicity which is conducted by the same route of exposure that will be used in the main experiment, or on the results of pre-existing sub-acute toxicity studies. It should be performed with due consideration to minimizing the number animals used. Substances with specific biological activities at low non-toxic doses (such as hormones and mitogens), and substances that exhibit saturation of toxicokinetic properties may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis. The dose levels selected should cover a range from little or no toxicity up to the Maximum Tolerated Dose (MTD).

4.2.1.3 Maximum Tolerated Dose (MTD)

When toxicity is the limiting factor, the top dose is usually the MTD, which is defined as the highest dose that will be tolerated without evidence of study-limiting toxicity such that higher dose levels, based on the same dosing regimen, would be expected to produce lethality or evidence of pain, suffering or distress necessitating humane euthanasia (OECD, 2000). The MTD is established in range-finding studies by measuring clinical effects and mortality, but it can also be identified from other toxicity studies in the same animal strain. In keeping with the 3Rs principles, animal use in determination of the MTD should be minimized; accordingly, a tiered range-finding study is recommended. The study should start with the most likely dose to cause toxicity, using a small number of animals (*e.g.* 2 per sex). If the MTD is not defined, a further group of animals should be exposed to a higher or lower dose depending on the clinical effects of the first dose. This strategy should be repeated until the appropriate MTD is found. Animals should be monitored for clinical signs of distress and excess toxicity should be euthanatized prior to completion of the test period in this, and all other, phases of the compete study (OECD, 2000).

4.2.1.4 Dosing and route of administration

In general, the anticipated route of human exposure should be used; however, other routes of exposure (such as, drinking water, subcutaneous, intravenous, topical, inhalation, intratracheal, dietary, or implantation) may be acceptable where they can be justified. It should be noted that intraperitoneal injection is specifically listed as not recommended in the revised TGs because it is not a physiologically relevant route of human exposure. The maximum volume of liquid that can be administered by gavage or injection at one time depends on the size of the test animal. The volume should not exceed 2 mL/100g body weight. In rare cases, the use of volumes greater than this may be appropriate and should be justified. Except for irritating or corrosive substances, which will normally reveal exacerbated effects at higher concentrations, variability in test volume should be minimized by adjusting the concentration to ensure a constant volume at all dose levels.

4.2.2 Proof of exposure (bioavaiability)

One of the more complex issues *in vivo* genetic toxicology testing is the determination that target tissues have received sufficient exposure when negative results have been obtained, particularly for a non-toxic substance tested at the limit dose. This information is required in order to establish that there has been sufficient exposure to a tissue (*i.e.* bioavailability) to justify a conclusion that the test chemical is non-genotoxic or /non-mutagenic.

For studies investigating genotoxic effects in the blood, bone marrow, or other well-perfused tissues, indirect evidence of target tissue exposure is generally sufficient to infer tissue exposure. Examples are absorption, distribution, metabolism, and excretion (ADME) data collected in the current or a concurrent experiment, or from clinical signs (such as coloured urine, ataxia, *etc*.) However, due consideration should be given to the possibility that short-lived metabolites may not reach the tissue being investigated, even when the chemical or metabolites are present in the circulatory system (Cliet *et al.*, 1993). In such cases, it may be necessary to actually determine

presence of test substance and/or metabolites in samples of the target tissue(s). Consequently, without the demonstration of bioavailability the value of a negative test is limited. If there is evidence that the test substance(s), or its metabolite(s) will not reach the target tissue it is not appropriate to use the particular *in vivo* test.

Direct evidence of target tissue exposure may be obtained from signs of toxicity in the target tissue, from toxicokinetic measurements of the substance or its metabolites in the tissue, or evidence of DNA exposure. For the new TG 489 (comet assay), histopathological changes are considered a relevant measure of tissue toxicity. Changes in clinical chemistry measures, *e.g.* aspartate aminotransferase (AST) and alanine aminotransferase (ALT), can also provide useful information on tissue damage and additional indicators such as caspase activation, terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) stain, Annexin V stain, *etc*. may also be considered. However, there are limited published data where the latter have been used for *in vivo* studies.

Even though a substance may produce toxic effects during range-finding studies, adequate exposure to the target tissue(s) should also be demonstrated in the main test.

4.2.3 Tissue selection, duration of treatment and sampling time

The treatment duration is dependent on the requirements and limitations of each test endpoint, as well as the relationship to the intended, or presumed, exposure of the test substance, if there is a choice of treatment duration in the TG. Appendix A shows the durations of treatment and sampling times for the *in vivo* tests (other regimens can be used if justified scientifically). In selecting exposure duration, it should be noted that gene mutations in transgenic animals can accumulate over time because the genes are "neutral". That is, the mutant cells are at neither a selective advantage nor disadvantage. For the other endpoints, including chromosomal aberrations, micronucleus and DNA damage (Comet), the events are generally either repaired or eliminated through apoptosis, and, therefore, they do not accumulate over time and must measured shortly after administration. For these latter endpoints, even when measured in experiments with chronic exposures, the events that are scored are those resulting only from the recent exposure and not the full duration of exposure.

4.2.3.1 *Chromosomal aberrations and micronuclei (TG 475, TG 474, TG 483)*

The selection of tissues for analysis of somatic chromosomal aberrations or micronuclei is fairly limited. Most historical studies have measured these endpoints in bone marrow or young blood reticulocytes. Methods for measurement of micronucleus induction in other tissues are being developed but are not currently described in these Test Guidelines (Uno 2015a and b).

4.2.3.2 *Transgenic rodent (TGR) gene mutations (TG 488)*

Mutations in transgenic rodents can be studied in any tissue from which sufficient DNA can be extracted. The rationale for selection of tissue(s) to be collected should be defined clearly. It should be based upon the reason for conducting the study together with any existing ADME, genetic toxicity, carcinogenicity or other toxicity data for the test substance under investigation. Important factors for consideration include the route of administration [based on likely human

exposure route(s)], the predicted tissue distribution and absorption, and the role of metabolism and the possible mechanism of action. Site of contact tissues relevant to the route of administration should be considered for sampling. If studies are conducted to follow up carcinogenicity studies, target tissues for carcinogenicity should be included. In the absence of any background information, tissues of potential interest should be collected. Other tissues can be frozen for later analysis, if needed (thus, potentially eliminating the need for an additional animal experiment, in compliance with the 3Rs). Since the induction of gene mutations is dependent on cellular proliferation, a suitable compromise for the measurement of mutant frequencies in both rapidly and slowly proliferating tissues is 28 consecutive daily treatments with sampling 3 days after the final treatment (*i.e.* 28+3 protocol); although the maximum mutant frequency may not manifest itself fully in slowly proliferating tissues under these conditions. It is important to note that TG 488 states if slowly proliferating tissues are of particular importance, then a later sampling time of 28 days following the 28 day administration period may be more appropriate (Heddle *et al.*, 2003; Thybaud *et al.*, 2003). In such cases, the later sampling time would replace the 3 day sampling time, but this would require scientific justification.

TG 488 notes that the 28+3 protocol may not be optimal for detection of mutations in spermatogonial stem cells, but can provide some coverage of cells exposed across the majority of phases of germ cell development, and may be useful for detecting some germ cell mutagens. Therefore, for tests focused on somatic tissues, it is recommended that, where possible, seminiferous tubules and spermatozoa from the cauda epididymis also be collected and stored in liquid nitrogen for potential future use.

Accordingly, studies designed specifically to detect mutagenic effects in male germ cells require additional considerations. In such cases when spermatozoa from the cauda epididymis, and seminiferous tubules from the testes are collected, care should be taken to ensure that the treatment-sampling times are appropriate and allow the detection of effects in all germ cell phases. Currently, TG 488 specifies that (in addition to the 28+3 protocol) a 28+49 regimen should be included to provide the optimal time for collecting spermatozoa from the cauda epididymis that were stem cells at the time of treatment. This requirement doubles the number of animals required. Accordingly, research is now underway to establish a suitable single, compromise sampling time, such as the 28+28 regimen described above for slowly proliferating tissues that would be suitable for both somatic and male germline tissues. It is expected that this research will support a current OECD project directed at updating TG 488 in the near future.

4.2.3.3 *Comet assay (TG 489)*

DNA damage can be studied in most tissues using the comet assay provided that good quality cells or nuclei can be prepared. Proliferation is not required to reveal effects in the comet assay otherwise the discussion of tissue selection in the previous section also applies to the comet assay. However, care should be taken and ADME parameters considered when selecting the sampling time(s) as the DNA damage is rapidly repaired. A sampling time of 2-6 hr after the last treatment for two or more treatments, or at both 2-6 and 16-26 h after a single administration are specified in TG 489. In addition, there is no consensus among experts about the validity of the

use of tissue or cell suspensions that have been frozen rather than analyzed immediately after necropsy (Speit 2015). It should be noted that there is considerably less experience with the regulatory use of this test. Descriptions of current considerations are in an annex to the Test Guideline and can be expected to continue to appear in the published literature (*e.g.* see Speit 2015).

4.2.4 Combination/integration of tests

There is a worldwide interest in reducing the use of experimental animals. In the spirit of the 3Rs principles, the combination of two or more endpoints in a single genetic toxicology study is strongly encouraged whenever possible, and when it can be scientifically justified. Examples of such test combinations are: 1) the *in vivo* bone marrow micronucleus test and liver comet assay (Hamada *et al*., 2001, Madrigal-Bujaidar *et al.,* 2008, Pfuhler *et al.,* 2009, Bowen *et al.,* 2011); 2) integration of genetic toxicology studies into repeated dose toxicity studies (Pfuhler *et al.,* 2009; Rothfuss *et al.*, 2011); and 3) the bone marrow micronucleus test and the transgenic rodent gene mutation assay (Lemieux et al , 2011).

Ideally, it would be best if all the assays being combined had similar treatment and sampling regimens (see Appendix A for treatment and sampling times). There are major considerations concerning the compatibility of test combinations with respect to these factors: 1) the effective length of the administration time; 2) the longevity of the genetic damage; and 3) the sampling time for the assays selected. For example, the micronucleus assay detects only damage that occurs in the 48 to 72 hr prior to tissue sampling if PCEs are examined, so, when combined with an assay using a 28 day sub-chronic administration time, the PCE/micronucleus assay will detect only micronuclei induced in the last few hours of the 28-day treatment. Furthermore, for the micronucleus combined with TGR assay, two sampling times would be needed to meet the sampling requirements of the MN and TGR assays. While this can be accomplished by drawing blood at 48 hr post-treatment for the flow cytometry MN assay and the killing the animals at 72 hr for the TGR assay, it still does not overcome the issue of the difference in the total effective dose delivered for the MN assay *vs.* the transgene mutation assay. A better alignment of doses can be accomplished when all assays in a test combination have the same effective "treatment window" and "endpoint enumeration window", such as would be accomplished with the MN assay and the comet assay; however, there is still a compatibility issue with respect to sampling times. The possibility also exists to combine non-genotoxicity assays, such as the Repeated Dose Oral Toxicity Study (TG 407), with genotoxicity tests (preferably with the same treatment protocol), but compromises with respect to treatment and sampling times will still have to be made, since the oral toxicity test ends on day 28 and genetic toxicity tests require a sampling time after day 28.

4.2.5 Use of one or both sexes

In general, the response of genetic toxicology tests is similar between male and female animals (Hayashi *et al.*, 1994) and, therefore, most studies using TG 474 and TG 475 could be performed in either sex. While TG 488 can be performed using either sex, males are used if germ cell effects are a consideration. Historically, most comet assay data were collected using only males.

There are no data examining sex differences in comet response. Data demonstrating relevant differences between males and females (*e.g.* differences in systemic toxicity, metabolism, bioavailability, bone marrow toxicity, *etc*. observed in a range-finding study) would encourage the use of both sexes. When a genetic toxicology test is incorporated into a test in which both sexes are being exposed, an increased statistical power can be gained with little extra expense by analyzing tissue from both sexes. Where human exposure to chemicals may be sex-specific, as for example with some pharmaceuticals, the test should be performed with the appropriate sex.

4.2.5.1 *Factorial design*

In cases where both sexes are used, it is advantageous to use a factorial design for the study, because the analysis will identify interaction effects between sex and treatment, and, if there are no interaction effects, it will provide greater statistical power. Both TG 474 and 475 provide a detailed description for use and interpretation of factorial designed studies in Annex 2 of these TGs.

4.2.6 Weight/age range

The starting age range of animals (*i.e.* rodents) varies according to the TG; for the *in vivo* MN (TG 474), *in vivo* CA (TG 475), and *in vivo* comet (TG 489) assays, it is 6 to 10 weeks. For the TGR assay and the spermatogonial CA (SCA) assay, it is 8 to12 weeks to facilitate access to sufficient numbers of transgenic animals from relatively small breeding colonies (TGR), and allow time to reach sexual maturity (TGR and SCA). The Dominant Lethal Test (DLT, TG 478) specifies healthy and sexually mature male and female adult animals.

4.3 Issues common to *in vitro* and *in vivo* TGs

4.3.1 Experimental design and statistical analysis considerations

As a part of the TG revision an extensive evaluation was undertaken to analyze how the selection of specific parameters impact the overall ability of the various tests to detect induced genetic damage. In particular, the analysis better defined an appropriate approach to using spontaneous background frequencies both for individual experiment acceptability and data interpretation, and to understand the impact of assay-specific background frequencies on the statistical power of the assay. This analysis was used to develop the new recommendations for the number of cells to be treated for the *in vitro* gene mutations assays and the number of cells to be scored for the cytogenetic tests (both *in vitro* and *in vivo*). A discussion of this analysis can be found in OECD documents (OECD, 2014b).

Recommendations were included in the revised TGs to discourage over-reliance on p-values associated with the statistical significance of differences found by pair wise comparisons. Statistical significance based upon a particular p-value is relevant, but is only one of the criteria used to decide whether to categorize a result as positive or negative. For example, the confidence intervals around the means for the controls and the treated cultures/animals should also be evaluated and compared. One of the goals for the TG revision was to include recommendations that would insure that test results deemed to be positive would be based on biologically relevant responses. It had been proposed that in the revised OECD genetic toxicology guidelines, studies

should be designed to detect a doubling (or 2-fold increase) in the treated group responses over the negative control level. However, subsequent discussions revealed that the sample sizes needed to detect a doubling will depend upon the background level; for example, a doubling from 1% to 2% is a smaller absolute change than one from 3% to 6%. Defining the level of response required to achieve biological relevance, therefore, requires, an appreciation of the nature of the endpoint, consideration of the background (negative control) incidence and whether an absolute or relative difference versus negative control should be considered. These considerations are different for each of the assays and have been taken into account in the new recommendations found in the individual TGs.

4.3.2 Size of samples and statistical power: *in vitro* tests

The TGs were evaluated for and in some cases revised to increase the power of the various assays to detect biologically significant increases. For the *in vitro* gene mutation studies, where the cell is the experimental unit, power calculations showed that designs with relatively small numbers of cells per culture had low power to detect biologically relevant differences. For the cytogenetic tests, in order to reach an acceptable level of statistical power (conventionally 80%) to detect 2 to 3 fold changes would only be achievable if the number of cells scored were increased appreciably in some tests. For revisions to the recommendations for the *in vitro* cytogenetic tests, consideration was given both to both the ideal number of scored cells and to the technical practicalities of actually scoring that number of cells, particularly for the chromosome aberration test.

TG 473 – In vitro mammalian chromosomal aberration test.

The 1997 version of TG 473 indicated that at least 200 well-spread metaphases should be scored and that these could be equally divided among the duplicates (when duplicates were used) or from single cultures. Based on a desire to increase the power of the assay, yet not make the assay too technically impractical, the number of cells to be scored was increased in this revision to at least 300 metaphases to be scored per concentration and control. As before, when replicate cultures are used the 300 cells should be equally divided among the replicates. When single cultures are used per concentration at least 300 well spread metaphases should be scored in the single culture. Scoring 300 cells has the advantage of increasing the statistical power of the test and, in addition, zero values will be rarely observed (expected to be only 5%) (OECD, 2014b). It should be noted that the number of metaphases scored can be reduced when high numbers of cells with chromosome aberrations are observed and the test chemical considered as clearly positive.

TG 487 – In vitro mammalian cell micronucleus test.

Based on the statistical power evaluations, a decision was made not to alter the recommendations for scoring from those made in the 2010 version of TG 487. Therefore, for the *in vitro* micronucleus test, micronucleus frequencies should be analysed in at least 2000 binucleate cells per concentration and control, equally divided among the replicates, if replicates are used. In the case of single cultures per dose at least 2000 binucleate cells per culture should be scored in the single culture. If substantially fewer than 1000 binucleate cells per culture (for duplicate

cultures), or 2000 (for single culture), are available for scoring at each concentration, and if a significant increase in micronuclei is not detected, the test should be repeated using more cells, or at less cytotoxic concentrations, whichever is appropriate. When cytoB is used, a CBPI or an RI should be determined to assess cell proliferation using at least 500 cells per culture

TG476- *In vitro* Mammalian Cell Gene Mutation Tests using the Hprt and xprt genes, and TG490 - *In vitro* mammalian cell gene mutation tests using the thymidine kinase gene.

In the 1997 version of TG476, a general recommendation was made concerning the number of cells that should be used in all of the *in vitro* gene mutation assays. The TG indicated that the minimal number of viable cells surviving treatment and used in each stage of the test should be based on the spontaneous mutant frequency and that number of cells should be at least ten times the inverse of the spontaneous mutant frequency. Furthermore, at least 1 million cells were recommended. The revision to TG 476 and the new TG 490 continue to recommend that the minimum number of cells used for each test (control and treated) culture at each stage in the test should be based on the spontaneous mutant frequency. Emphasis is now, however, placed on assuring that there are a minimum number of spontaneous mutants that are maintained in all phases of the test (treatment, phenotypic expression and mutant selection). The expert workgroup chose to use the recommendation of Arlett *et al.*, (1989) which states that a general guide is to treat and passage sufficient cells in each experimental culture so as to maintain at least 10 but ideally 100 spontaneous mutants.

For the MLA, the recommended acceptable spontaneous mutant frequency is between 35-140 x 10^{-6} (agar version) and 50-170 x 10^{-6} (microwell version). To have at least 10 and ideally 100 spontaneous mutants surviving treatment for each test culture, it is necessary to treat at least 6 x 10^6 cells. Treating this number of cells, and maintaining sufficient cells during expression and cloning for mutant selection, provides for a sufficient number of spontaneous mutants (10 or more) during all phases of the experiment, even for the cultures treated at concentrations that result in 90% cytotoxicity (as measured by an RTG of 10%) (Lloyd and Kidd, 2012; Mei *et al.*, 2014; Schisler *et al.*, 2013).

For the TK6, the spontaneous mutant frequency is generally between 2 and 10 x 10^{-6}. To have at least 10 spontaneous mutants surviving treatment for each culture it is necessary to treat at least 20 x 10^6 cells. Treating this number of cells provides a sufficient number of spontaneous mutants (10 or more) even for the cultures treated at concentrations that cause 90% cytotoxicity during treatment (10% RS). In addition a sufficient number of cells must be cultured during the expression period and plated for mutant selection (Honma and Hayashi 2011).

For the *hpr*t assay, the spontaneous mutant frequency is generally between 5 and 20 x10^{-6}. For a spontaneous mutant frequency of 5 x10^{-6} and to maintain a sufficient number of spontaneous mutants (10 or more), even for the cultures treated at concentrations that cause 90% cytotoxicity during treatment (10% RS), it would be necessary to treat at least 20 x 10^6 cells. In addition a sufficient number of cells (but never less than 2 million) must be cultured during the expression period and plated for mutant selection.

4.3.3 Size of samples and statistical power: *in vivo* tests

Sample sizes were also increased in the *in vivo* tests to increase the power to detect increases. Statistical power increases with the number of cells scored and/or the number of animals per group (OECD, 2014b). The challenge is to select these numbers to best achieve appropriate statistical power while keeping cell numbers within practical limits, and avoiding excessive use of animals. With this goal in mind, most *in vivo* genetic toxicology TGs have been revised to achieve enhanced statistical power.

TG 474 - Mammalian erythrocyte micronucleus test.

The previous version of this TG required the scoring of 2000 or more cells per animal (5 animals per group). Statistical analyses (Kissling *et al.,* 2007; OECD, 2014b) have shown that *in vivo* designs for micronuclei with n = 5 animals have the power to detect 2 to 3-fold effects with 80% power based upon counts of about 4000 cells per animal when the background incidences are relatively high (0.1% and higher). Accordingly, the revised TG 474 now recommends at least 4000 cell per animal. The power increases with higher background control incidences. However, larger sample sizes, either as more animals and/or many more cells, would be needed to have sufficient power to detect a 2-3 fold incidence when the background incidence is lower (*i.e.* <0.05%).

TG 475 - Mammalian bone marrow chromosome aberration test.

For similar statistical reasons, the minimum number of cells has been increased from 100 to 200 cells per animal with 5 animals per group from the previous version of this TG. This sample size is sufficient to detect at least 80% of chemicals which induce a 2-fold increase in aberrant cells over the historical control level of 1.0% and above at the significance level of 0.05 (Adler *et al.*,1998b).

TG 478 - Dominant lethal test.

The original version of TG 478 contained minimal information on the conduct of this test. The revised TG 478 specifies that the number of males per group should be predetermined to be sufficient (in combination with the number of mated females at each mating interval) to provide the statistical power necessary to detect at least a doubling in dominant lethal frequency (*e.g.* about 50 fertilized females per mating; formerly 30-50). A detailed description of the recommended statistical analysis is now provided in the TG.

TG 483 - Mammalian spermatogonial chromosomal aberration test.

For similar statistical reasons, the number of minimum number of cells in TG 483 has also been increased from 100 to 200 cell per animal with 5 animals per group (Adler *et al.,* 1994).

TG 489 - In vivo alkaline comet assay.

This new TG 489 specifies that for each sample (per tissue per animal), at least 150 cells (excluding hedgehogs) should be analysed. Scoring 150 cells per animal in at least 5 animals per dose (less in the concurrent positive control) provides adequate statistical power according to the analysis of Smith *et al*. (2008).

4.3.4 Demonstration of laboratory proficiency and establishing an historical control database

The revised OECD genotoxicity TGs now include a requirement for the demonstration of laboratory proficiency. In consideration of the 3Rs, which place constraints on the use of animals, the recommendations for demonstrating laboratory proficiency are different for *in vitro* tests, for *in vivo* somatic tests, and for *in vivo* germ cell tests. It should be noted that the recommended methods to establish proficiency do not apply to experienced laboratories that have already been able to do so by building historical control databases of both positive and negative controls. Also, as a part of demonstrating proficiency both initially and over time, the new TGs introduce and recommend the concept of using quality control charts to assess the historical control databases (see Section 4.3.4 for more information on control charts).

In order to establish sufficient experience with the test prior to using the test for routine testing the laboratory should have performed a series of experiments using reference substances with different mechanisms of action, showing that the laboratory can discriminate between negative and positive substances, and detect positive substances acting via different mechanisms, and requiring or not metabolic activation. TGs provide recommendations for the substances that could be used for each test.

For *in vitro* and most somatic *in vivo* assays, a selection of positive (at least two *in vivo*) and negative control substances should be investigated under all experimental conditions of the specific test (e.g. short- and long-term treatments for in vitro assays, as applicable) and give responses consistent with the published literature. It should be based on at least 10, but preferably 20, experiments, that demonstrate that the assay conforms to published positive and negative control norms (Hayashi *et al.*, 2011).

For in vivo somatic TGs wherein multiple tissues can be used (*e.g.*, the *in vivo* alkaline comet assay and the transgenic rodent gene mutation assay) proficiency should be demonstrated in each tissue that is being investigated. During the course of these investigations the laboratory establish an historical database of positive and negative control values, as described in Section 4.3.4.

For the TGR and SCA assays and the DLT there is currently no explicit requirement to establish an historical control database. However, competency should be demonstrated by the ability to reproduce expected negative and positive control results from published data when conducting any new study. The positive and negative control literature on the TGR assay has been compiled and is readily available in an OECD Detailed Review Paper (OECD, 2009); however, since such compiled sources are not available for the DLT and SCA assays, summaries of negative control data for these assays are presented herein (Appendices B and C respectively).

The negative control values for percent resorptions in the DLT varies widely depending on the parental strains used from 3.3 [(SECxC57BL)F1 x (C3Hx101)F1] to 14.3 [T-Stock x (CH3x101)F1]. Thus, a recommended range for the negative control value cannot be easily identified. Furthermore, some of the strains shown in Appendix B may not be generally available; therefore, laboratories should choose an available strain with stable negative control variability when planning to perform the DLT.

The negative control values for the percent cells with chromosomal aberrations in the SCA assay also varies among studies (Appendix C). Based on the data in this Table, TG 483 states that the recommended range for negative controls is >0 to ≤1.5 % of cells with chromosomal aberrations.

4.3.5 Concurrent negative and positive controls

In addition to establishing laboratory competence, negative and positive historical control data are important for assessing the acceptability of individual experiments, and the interpretation of test data. In particular, it is necessary to determine whether specific responses fall within or outside the distribution of the negative control. With the 3R principles in mind, the recommendations for positive controls differ for *in vitro* and among various *in vivo* tests.

4.3.5.1 Concurrent negative controls

Negative control groups are important for providing a contemporaneous control group for use in comparisons with the treated groups. This group can also be used to assess, whether the experiment is of acceptable quality by comparison with a set of historical control groups.

Negative controls usually consist of solvent or vehicle treated cells or animals. They should be incorporated into each *in vitro* and *in vivo* test and handled in the same way as the treatment groups, except for not receiving treatment with the test chemical. If an unusual solvent or vehicle, or a common solvent is being used at a greater than normal concentration, then inclusion of untreated controls is required. This allows comparison of the solvent control response with the untreated control and a judgment can be made regarding the acceptability of the solvent for use in the test. In addition, it should be noted that when choosing a solvent or vehicle the decision should be based on obtaining maximum solubility of the test material without interacting with the test chemical and/or test system.

In order to reduce unnecessary animal usage for *in vivo* tests, if consistent inter-animal variability and frequencies of cells with DNA damage are demonstrated by historical negative control data at each sampling time for the testing laboratory, only a single sampling for the negative control may be necessary. Where a single sampling is used for negative controls, it should be the first sampling time used in the study.

4.3.5.2 Concurrent positive controls

The inclusion of concurrent positive controls (reference controls/well-known genotoxic substances) is designed to demonstrate the effectiveness of a particular genetic toxicology test on the day it is performed. Each positive control should be used at a concentration or dose expected to reliably and reproducibly result in a detectable increase over background in order to demonstrate the ability of the test system to efficiently detect DNA damage, gene mutations and/or chromosomal aberrations depending on the test, and in the case of *in vitro* tests, the effectiveness of the exogenous metabolic activation system. Therefore, positive control responses (of both direct-acting substances and substances requiring metabolic activation) should be observed at concentrations or doses that produce weak or moderate effects that will be detected when the test system is optimized, but not so dramatic that positive responses will be seen in sub-optimal test systems, and immediately reveal the identity of the coded samples to the

scorer (i.e for tests using coded samples).

4.3.5.2.1 *In vitro tests*

For each of the *in vitro* genetic toxicology tests, positive control substances should be assayed concurrently with the test substance. Because *in vitro* mammalian cell tests for genetic toxicity are sufficiently standardized, the use of positive controls may be confined to a substance requiring metabolic activation. Provided it is done concurrently with the non-activated test using the same treatment duration, this single positive control response will demonstrate both the activity of the metabolic activation system and the responsiveness of the test system. Longterm treatment should, however, have its own positive control as the treatment duration will differ from the test using metabolic activation. In the case of the *in vitro* micronucleus test, positive controls demonstrating clastogenic and aneugenic activity should be included. For the gene mutation tests using the *tk* locus, positive controls should be selected that induce both large and small colony mutants.

4.3.5.2.2 *In vivo tests*

For *in vivo* tests, a group of animals treated with a positive control substance should normally be included with each test. In order to reduce unnecessary animal usage when performing a transgenic rodent gene mutation, micronucleus, bone marrow chromosomal aberration, or spermatogonial chromosomal aberration tests, this requirement may be waived when the testing laboratory has demonstrated proficiency in the conduct of the test according to the criteria described in the TG for each test. In such cases where a concurrent positive control group is not included, scoring of "reference controls" (fixed and unstained slides, cell suspension samples, or DNA samples from the same species and tissues of interest, and properly stored) must be included in each experiment. These samples can be collected from tests during proficiency testing or from a separate positive control experiment conducted periodically (*e.g.* every 6-18 months), and stored for future use. For the dominant lethal test, concurrent positive controls are required until laboratories have demonstrated proficiency, and then they are not required. Because of insufficient experience with the longevity of alkali labile DNA sites in storage, with the comet assay, concurrent positive controls are always necessary.

Since the purpose of a positive control is primarily to demonstrate that the assay is functioning correctly (and not to validate the route of exposure to the tested compound), it is acceptable that the positive control be administered by a route different from the test substance, using a different treatment schedule, and for sampling to occur only at a single time point.

4.3.5.3 *Historical control distribution and control charts*

Historical control data (both negative and positive) should be compiled separately for each genetic toxicology test type, for each species, strain, tissue, cell type, treatment and sampling time, route of exposure, as well as for each solvent or vehicle within each laboratory. All control data of each individual genetic toxicology test, strain *etc*. during a certain time period (*e.g.* 5 years) or from the last tests performed (*e.g*. the last 10 or 20 tests) should initially be accumulated to create the historical control data set. The laboratory should not only establish the historical negative (untreated, vehicle) and positive control range but also define the distribution

(e.g. Poisson distribution 95% control limits) as this information will be used for data interpretation. This set should be updated regularly. Any changes to the experimental protocol should be considered in terms of their impact on the resulting data remaining consistent with the laboratory's existing historical control database. Only major changes in experimental conditions should result in the establishment of a new historical control database where expert judgement determines that it differs from the previous distribution. Further recommendations on how to build and use the historical data (*i.e.* criteria for inclusion and exclusion of data in historical data and the acceptability criteria for a given experiment) can be found in the literature (Hayashi *et al.*, 2011).

According to the new and revised TGs laboratories should use quality control methods, such as control charts. Control charts are plots of data collected over a period of time with horizontal lines established to define the upper and lower bounds of the range of acceptable values for the particular assay. Control charts are long-established and widely-used methods for quality control laboratories to monitor the variability of samples and to show that their methodology is 'under control' rather than drifting over time. They also provide a visual presentation of the variability within a laboratory, which can help put any possible treatment-related effects into context. There are many different types of control charts. Examples are I-charts for plotting individual values, C-charts for plotting count data and Xbar-charts for plotting the means of groups of individuals such as the individual units in a negative control group (OECD, 2014b)

Most major statistical software packages (*e.g.* SAS, SPSS, Stata, Genstat. Minitab. JMP) have procedures for producing control charts and provide guides for using the procedures. There is also much software specifically designed for quality control methodology in general. The software language R has a package (*qcc*) which can produce control charts. In addition there are a number of textbooks describing the methods (Ryan, 2011, Henderson, 2011, Montgomery, 2005 & Mulllins, 2003). The US National Institute of Standards and Technology (NIST) has a detailed online description and discussion of the methodology (http://www.itl.nist.gov/div898/handbook/pmc/pmc.htm). There are also numerous online discussion groups in areas related to Total Quality Management (TQM), Six-sigma methodology and Statistical Process Control (SPC) that actively discuss issues in the Quality Assurance/Control field.

4.3.5.4 *Data interpretation and criteria for a positive/negative result*

In revising the TGs, the expert workgroup gave extensive consideration to providing more guidance than was given in the previous TGs for interpreting test data. As a result, several new concepts are included in the revised/new TGs. Prior to considering whether a particular experiment is positive or negative, it is important to ascertain whether that experiment is properly conducted. Therefore, the revised TGs clarify the acceptance criteria for each assay. In addition, guidance was developed to provide recommendations as to what defines a biologically relevant positive result. Previous TGs indicated that positive responses should be biologically relevant, but did not provide a means to determine biological relevance. The revised/new TGs include three equal considerations when assessing whether a response is positive or negative. First the test chemical response should be statistically different from the concurrent negative

controls. Second, the response should be concentration/dose related. Finally, a new concept, that utilizes the historical negative control distributions, is introduced to provide for assessing biological relevance.

4.3.5.4.1 *Individual test acceptability criteria*

The revised TGs clarify recommendations for individual assay acceptability as follows.

- The concurrent negative control is considered acceptable for addition to the laboratory historical negative control database, and/or is consistent with published norms (depending on the assay).
- Concurrent positive controls induce responses that are compatible with those generated in the laboratory's historical positive control data base, and produce a statistically significant increase compared with the concurrent negative control.
- For *in vitro assays,* all experimental conditions (based on the recommended treatment times and including the absence and presence of metabolic activation) were tested unless one resulted in clear positive results.
- Adequate number of animals/cells were treated and carried through the experiment or scored (as appropriate for the individual test)
- An adequate number of doses/concentrations covering the appropriate dose/concentration range is analysable.
- The criteria for the selection of top dose/concentration are consistent with those described in the individual TGs.

Apart from the above criteria, MLA-specific acceptability criteria have been defined based on the IWGT MLA expert workgroup's (Moore *et al.*, 2000; 2002; 2003; 2006) data evaluation for several negative control data parameters. Consistent with the general approach to establishing acceptability criteria for the revised genetic toxicology TGs, these recommendations are based on distributions of a very large number of experiments from laboratories proficient in the conduct of the MLA. There are also MLA-specific criteria for positive controls that assure good recovery of both small and large colony mutants. The specific recommendations (*i.e.* acceptable ranges for the main parameters) for the MLA are detailed in TG 490.

4.3.5.4.2 *Criteria for a positive/negative result*

If a genetic toxicity test is performed according to the specific TG and all acceptability criteria are fulfilled (as outlined above), the data can be evaluated as to whether the response is positive or negative. The new TGs recognize it is important that chemicals determined to be positive demonstrate biologically relevant increases that are concentration/dose related. As with the acceptability criteria, the assessment of biological relevance takes the distribution of the negative control data into consideration (*e.g.* Poisson 95% control limits).

For both *in vitro* and *in vivo* assays (with the exception of the MLA—see below) a response is considered a clear positive in a specific test if it meets all the criteria below in at least one experimental condition:

- at least one of the data points exhibits a statistically significant increase compared to the concurrent negative control;
- the increase is concentration- or dose-related at least at one sampling time when evaluated with an appropriate trend test;
- the result is outside the distribution of the historical negative control data (*e.g.* Poisson-based 95% control limits).

A test chemical is considered clearly negative if, in all experimental conditions examined, none of the above criteria for a positive result are met.

Recommendations for the most appropriate statistical methods can be found in the literature (Lovell *et al.*, 1989; Kim *et al.*, 2000).

For the MLA, the IWGT MLA expert workgroup recommendation for determination of a biologically relevant positive result does not rely on statistical significant increases compared to the concurrent negative control but on the use of a predefined induced mutant frequency (*i.e.* increase in MF above concurrent control), designated the Global Evaluation Factor (GEF) which is based on the analysis of the distribution of the negative control MF data from participating laboratories (Moore *et al.*, 2006). For the agar version of the MLA the GEF is 90×10^{-6}, and for the microwell version of the MLA the GEF is 126×10^{-6}.

As outlined above, the revised/new TGs provide criteria for results that are clearly positive or negative. If the response is neither clearly negative nor clearly positive the TGs recommend that expert judgement be applied. Test results that do not meet all the criteria may also be judged to be positive or negative without further experimental data, but they need to be evaluated more closely before any final conclusion is reached. If, after the application of expert judgement, the results remain inconclusive (perhaps as a consequence of some limitation of the test or procedure) they should be clarified by further testing preferably using modification of experimental conditions (*e.g.* other metabolic activation conditions, length of treatment, sampling time, concentration spacing *etc*.)

For all of the tests covered in the genetic toxicology TGs, there is no requirement for verification of a clear positive or negative response.

In rare cases, even after further investigations, the data set will preclude a definitive positive or negative call. Therefore the test chemical response should be concluded to be equivocal (interpreted as equally likely to be positive or negative).

4.3.5.5 *Substances that require specific approaches*

There are some substances, such as nanomaterials, complex mixtures, volatiles, aerosols and gases, that require special modifications of the TGs in order to: 1) properly characterize the test material, 2) adequately expose the cells/animals; 3) conduct an adequate test; and 4) to properly

interpret the test data. Guidance for special modifications are not described in the TG, but can be found in X, Y, and Z.

4.3.5.6 Test batteries, weight of the evidence and interpretation of data

The OECD TGs provide recommendations on how to conduct the various genetic toxicology tests. However, in some cases, regulatory authorities have recommended modifications to specific guidelines appropriate for specific product types, *e.g.* ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use).

The OECD TGs do not make any specific recommendations as to which tests to use in a test battery. Regulatory agencies publish their recommendations and they should be consulted prior to initiating testing. Generally, the recommended genetic toxicology test batteries include tests to detect both gene mutations and structural as well as numerical chromosomal damage in both *in vitro* and *in vivo* tests; however, more recently, in some jurisdictions the emphasis has been on using only *in vitro* (and no, or fewer, *in vivo*) tests.

There are several publications that provide basic information on using genetic toxicology information for regulatory decisions (Dearfield and Moore, 2005; *etc*). In addition there have been expert workgroup discussion concerning appropriate follow-up testing strategies both from chemicals found to be positive *in vitro* tests and/or *in vivo* tests (Dearfield *et al.*, 2011; Thybaud *et al.*, 2007; Thybaud *et al.*, 2011; Tweats *et al.*, 2007; *etc*).

It is important to emphasize that the results from the different assays are not (and should not be) evaluated in isolation. The amount of data available for a weight of evidence evaluation will vary enormously, particularly among different product categories. Data-rich packages prepared for drug or pesticide regulations may permit analyses that would be impossible for substances involving other uses for which less data are available.

5 REFERENCES

- Adler ID.,(1974), Comparative cytogenetic study after treatment of mouse spermatogonia with Mitomycin C, Mutation Res. 23, 369-379.
- Adler ID., (1982), Male germ cell cytogenetics, in “Cytogenetic Assays of Environmental Mutagens”, T.C. Hsu (Ed), Allanheld, Osmun Publishers, Totowa, NJ, USA, pp. 249-276.
- Adler, I.-D., Shelby M. D., Bootman, J., Favor, J., Generoso, W., Pacchierotti, F., Shibuya, T. and Tanaka N., (1994), International Workshop on Standardisation of Genotoxicity Test Procedures. Summary Report of the Working Group on Mammalian Germ Cell Tests. Mutation Res., 312, 313-318.
- Adler, I.D., Kliesch, U., Tiveron, C., and Pacchierotti, F., (1995), Clastogenicity of diepoxybutane in bone marrow cells and male germ cells of mice Mutagenesis 10:535-541

- Adler ID, El-Tarras A., (1989), Clastogenic effects of cis-diamminedichloro platinum. Induction of chromosomal aberrations in somatic and germinal cells of mice, Mutation Res. 211, 131-137.
- Adler, I.D., Filser, J., Gonda, H., and Schriever-Schwemmer, G., (1998a), Dose response study for 1.3-butadiene-induced dominant lethal mutations and heritable translocations in germ cells of male mice, Mutat Res 397:85-92
- Adler I-D., Bootman, J., Favor, J., Hook, G.,, Schriever-Schwemmer, G., Welzl, G.,, Whorton, E., Yoshimura, I., Hayashi, M. (1998b), Recommendations for statistical designs of *in vivo* mutagenicity tests with regard to subsequent statistical analysis, Mutation Res., 417, 19–30.
- Asano N, Katsuma Y, Tamura H, Higashikuni N, Hayashi M, (1998), An automated new technique for scoring the rodent micronucleus assay: computerized image analysis of acridine orange supravitally stained peripheral blood cells. Mutat Res 404, 149-154.
- Azofeifa,J., Fauth,C., Kraus,J., Maierhofer,C., Langer,S., Bolzer,A., Reichman,J., Schuffenhauer,S. and Speicher,M.R., (2002), Induction of chromosomal aberrations by dacarbazine in somatic and germinal cells of mice, Mutagenesis 17:383-389
- Anderson D, Hughes JA, Edwards AJ, Brinkworth MH., (1998), A comparison of male-mediated effects in rats and mice exposed to 1,3-butadiene. Mutat Res. Jan 16;397(1):77-84.
- Anderson, D., Mcgregor, D.B., & Purchase, I.F.H. (1976), Dominant lethal studies with paraquat and diquat in male cd-1 mice, Mutat Res 40:349-358
- Anderson D, Hodge MCE, Purchase IFH. (1976), Vinyl chloride: dominant lethal studies in male cd-1 mice, Mutat Res 40:359-370
- Anderson, D., Hodge, M.C.E. & Purchase, I.F.H. (1977), Dominant lethal studies with the halogenated olefins vinyl chloride and vinylidene dichloride in male CD-i Mice, Environ Health Persp 21:71-78
- Anderson, D., Clapp, M.J.L., Hodge, M.C.E., Weight, T.M. (1981), Irradiated laboratory animal diets dominant lethal studies in the mouse, Mutat Res 80:333-345
- Anderson et al., (1998), A comparison of male-mediated effects in rats and mice exposed to 1,3-butadiene, Mutat Res 397:77-84
- Attia, S,M., (2012), Dominant lethal mutations of topoisomerase II inhibitors etoposide and merbarone in male mice: a mechanistic study, Arch Toxicol 86:725-731
- Attia, S.M., Ahmad, S.F., Okash ,R.M., Bakheet, S.A. (2015), Dominant lethal effects of nocodazole in germ cells of male mice, Food Chem Tox 77:101-104
- Barnett, L.B. & Lewis, S.E., (2003), Chlornaphazine and chlorambucil induce dominant lethal mutations in male mice, Mutat Res 543:145-154

- Barnett, L.B., Lovell, D.P.., Felton, C.F., Gibson, B.J., Cobb, R.R., Sharpe, D.S., Shelby, M.D., & Lewis, S.E., (1992), Ethylene dibromide: negative results with the mouse dominant lethal assay and the electrophoretic specific-locus test, Mutat Res 282:127-133
- Barnett LB, Tyl RW, Shane BS, Shelby MD, Lewis SE, (2002), Transmission of mutations in the lacI transgene to the offspring of ENU-treated Big Blue male mice. Environ. Mol. Mutagen. 40, 251-257.
- Bateman AJ, (1984), The dominant lethal assay in the male mouse. In: Handbook of Mutagenicity Test Procedures, 2nd Edition, Kilbey BJ, Legator M, Nichols W, Ramel C (eds.), Elsevier Scientific, Amsterdam, pp 471-484.
- Bishop, J B.; Kodell, R L.; Whorton, E B.; and Domon, O E. (1983), Dominant lethal test response with IMS and TEM using different combinations of male and female stocks of mice, Mutat Res 12:273-280
- Blaszkowska, J. (2010), Induction of dominant lethal mutations by Ascaris trypsin inhibitor in male mice, Reprod Tox 29:113-119
- Bowen D, Whitwell JH, Lillford L, Henderson D, Kidd D, Mc Garry S, Pearce G, Beevers C, Kirkland DJ, (2011), Evaluation of a multi-endpoint assay in rats, combining the bone-marrow micronucleus test, the comet assay and the flow-cytometric peripheral blood micronucleus test. Mutat. Res. 722, 7-19.
- Brookmire L., J. Chen, R. Gudi and D.D. Levy, (2013), Evaluation of the highest concentrations used in the chromosome aberrations assay. Environ Mol Mutagen. 54; 36-43.
- Cao, X., Mittelstaedt, R.A., Pearce, M.G., Allen, B.C., Soeteman-Hernandez, L.G., Johnson, G.E., Bigger, C.A., and Heflich, R.H. (2014), Quantitative dose-response analysis of ethyl methanesulfonate genotoxicity in adult gpt-delta transgenic mice. *Environ Mol Mutagen*, **55**, 385-399.\
- Cimino, M.C., (2006), Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes, Environ. and Molec. Mutagenesis 47:362-390., *Erratum*, Environ. and Molec. Mutagenesis 47:746 (2006)
- Ciranni R, Adler ID (1991), Clastogenic effects of hydroquinone: induction of chromosomal aberrations in mouse germ cells, Mutation Res. 263, 223-229.
- Ciranni R, Barale R, Adler ID (1991) ,Dose-related clastogenic effects induced by benzene in bone marrow cells and in differentiating spermatogonia of Swiss CD1 mice, Mutagenesis 6, 417-421.
- Claxton L, Stewart-Hook V, Warren S, (2001),Methods for the Spiral *Salmonella* mutagenicity assay 32 including specialised applications. Mutat. Res. 488, 241-257.
- Cliet, I., Melcion, C., Cordier, A., Lack of predictivity of bone marrow micronucleus test versus testis micronucleus test: comparison with four carcinogens, Mutation research, 292 (1993), 105-111.

- Clive, D. and Spector, J.F.S. (1975), Laboratory procedure for assessing specific locus mutations at the TK locus in cultured L5178Y mouse lymphoma cells. Mutat. Res., 31 (1): 17-29.
- COM, (2000) Guidance on a Strategy for Testing of Chemicals for Mutagenicity. Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment (COM). Department of Health, London.
- Corvi R, Albertini S, Hartung T, Hoffmann S, Maurici D, Pfuhler S, van Benthem J, Vanparys P, (2008). ECVAM retrospective validation of the *in vitro* micronucleus test (MNT). Mutagenesis, 271, 271-283.
- J.A. Cox, M.D. Fellows, T. Hashizume, and P.A. White. (2015) The Utility of Metabolic Activation Mixtures Containing Human Hepatic Post-Mitochondrial Supernatant (S9) for In Vitro Genetic Toxicity Assessment. Mutagenesis. Manuscript under review. Manuscript ID Mutage-15-093.
- Dearfield, K. L. and Moore, M. M. (2005), Use of genetic toxicology information for risk assessment. Environ. Mol. Mutagen., 46: 236–245.
- Dearfield KL, Thybaud V, Cimino MC, Custer L, Czich A, Harvey JS, Hester S, Kim JH, Kirkland D, Levy DD, Lorge E, Moore MM, Ouédraogo-Arras G, Schuler M, Suter W, Sweder K, Tarlo K, van Benthem J, van Goethem F, Witt KL. (2011) Follow-up actions from positive results of *in vitro* genetic toxicity testing. Environm. Molec. Mutagen. 52:177-204.
- De Boeck M, van der Leede B, Van Goethem F, De Smedt A, Steemans M, A. L, Vanparys P, (2005), Flow cytometric analysis of micronucleated reticulocytes: Time- and dose-dependent response of known mutagens in mice, using multiple blood sampling. Environm. Molec. Mutagen. 46, 30-42.
- De Flora S, Izzotti A, (2007), Mutagenesis and cardiovascular diseases: molecular mechanisms, risk factors, and protective factors. Mutat. Res; 621, 5–17.
- Dertinger SD, Torous DK, Hayashi M, MacGregor JT, (2011), Flow cytometric scoring of micronucleated erythrocytes: an efficient platform for assessing *in vivo* cytogenetic damage. Mutagenesis 26, 139-145.
- ECHA, (2008), Guidance on information requirements and chemical safety assessment Chapter R.4: Evaluation of available information. Available at: http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r4_en.pdf?vers=20_08_08.
- Dunnick, J. K., Solleveld, H.A., Harris, M. W., Chapin, R., & Lamb, IV, J.C. (1984) Dimethyl methyl phosphonate induction of dominant lethal mutations in male mice, Mutat Res 138:213-218

- Eastmond, D.,A. , A. Hartwig, D. Anderson, W.A. Anwar, M.C. Cimino, I. Dobrev, G.R. Douglas, T. Nohmi, D.H. Phillips, C. Vickers, (2009), Mutagenicity testing for chemical risk assessment: update of the WHO/IPCS harmonized scheme, Mutagenesis 24 341–349.
- Ehling. U.H. (1971), Comparison of radiation- and chemically-induced dominant lethal mutations in male mice, Mutat Res 11:35-44
- Ehling, EH. (1974), Differential spermatogenic response of mice to the induction of mutations by antineoplastic drugs, Mutat Res 26:285-295
- Ehling, E.H.,& Neuhäuser-Klaus (1988), Induction of specific-locus and dominant-lethal mutations by cyclophosphamide and combined cyclophosphamide-radiation treatment, Mutat Res 199:21-30
- Ehling, U.H., & Neuhäuser-Klaus, A. (1989), Induction of specific-locus and dominant lethal mutations in male mice by chlormethine, Mutat Res., 227:81-89
- Ehling, U.H., & Neuhäuser-Klaus, A. (1989), Induction of specific-locus and dominant lethal mutations in male mice by buslfan, Mutat. Res. 249:285-292
- Ehling, U.H., & Neuhäuser-Klaus, A. (1994), Induction of specific-locus and dominant lethal mutations in male mice by trophosphamide, Mutat Res 307:229-236
- Ehling, U.H., & Neuhäuser-Klaus, A. (1995), Induction of specific-locus and dominant lethal mutations in male mice by rt-propyl and isopropyl methanesulfonate, Mutat Res 328:73-82
- Ehling, U.H., Cumming, R.B. & Malling, H.V. (1 (1968), Induction of dominant lethal mutations by alkylating agents in male mice, Mutat Res 5:417-428
- Elhajouji, A., M. Cunha, M. Kirsch-Volders (1998), Spindle poisons can induce polyploidy by mitotic slippage and micronucleate mononucleates in the cytokinesis-block assay, *Mutagenesis*, Vol. 13/2, pp. 193-8.
- Erickson RP, (2010), Somatic gene mutation and human disease other than cancer: an update. Mutat. Res. 705, 96-106.
- Fellows MD, O'Donovan MR, Lorge E. Kirkland D, (2008), Comparison of different methods for an accurate assessment of cytotoxicity in the *in vitro* micronucleus test II: Practical aspects with toxic agents. Mutat. Res. 655, 4-21.
- Fellows M.D., Luker, T., Cooper, A. and O'Donovan, M.R. (2012), Unusual structure-genotoxicity relationship in mouse lymphoma cells observed with a series of kinase inhibitors. Mutat. Res. 746 (1): 21-28.
- Fenech, M. (2007), Cytokinesis-block micronucleus cytome assay, *Nature Protocols*, Vol. 2/5, pp. 1084-1104.
- Fluckiger-Isler S, Baumeister M, Braun K, Gervais V, Hasler-Nguyen N, Reimann R, Van Gompel J, Wunderlich H-G, Engelhardt G, (2004), Assessment of the performance of the

Ames IITM assay: a collaborative study with 19 coded compounds. Mutat. Res. 558, 181-197.

- Fowler P, Smith R, Smith K, Young J, Jeffrey L, Kirkland D, Pfuhler S, Carmichael P, (2012), Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. II. Importance of accurate toxicity measurement. Mutat Res.747, 104-117.
- Frank SA, (2010), Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. Proc. Natl. Acad. Sci. U S A. 107 Suppl 1:1725-1730.
- Generoso, W. M., Preston R. J., & Brewen, J. G. (1975), Ehling & Neuhäuser-Klaus (1989) 6-mercaptopurine, an inducer of cytogenetic and dominant- lethal effects in premeiotic and early meiotic germ cells of male mice, Mutat Res 28:437-447
- Generoso, W.M., Cain, K.T., Hellwig, C.S., & Cacheiro, N.L.A. (1982), Lack of association between induction of dominant-lethal mutations and induction of heritable translocations with benzo[a]pyrene in postmeiotic germ cells of male mice, Mutat Res 94:155-163
- Generoso, W.M., Cain, K.T., Hughes, L.A., Sega, G.A., Braden, P.W., Gosslee, D.G., and Shelby, M.D. (1986), Ethylene oxide dose and dose-rate effects in the mouse dominant-lethal test, Environ Mutagen 8:1-7
- Generoso ,W.M., Witt, Kristine L., Cain, K.T., Hughes, Lori, Cacheiro, N.L.A., Lockhart, Ann-Marie C., Shelby, M.D. (1995) Dominant lethal and heritable translocation tests with chlorambucil and melphalan in male mice, Mutat Res 345:167-180
- Generoso, W.M.. Sega, G.A., Lockhart, A.M., Hughes, L.A., Cain K.T.,Cacheiro, N.L.A., & Shelby, M.D. (1996) Dominant lethal mutations, heritable translocations, and unscheduled DNA synthesis induced in male mouse germ cells by glycidamide, a metabolite of acrylamide, Mutat Res
- Generoso WM, Piegorsch WW, (1993). Dominant lethal tests in male and female mice. Methods. Toxicol. 3, 124-141.
- Gollapudi, B.B., Johnson, G.E., Hernandez, L.G., Pottenger, L.H., Dearfield, K.L., Jeffrey, A.M., Julien, E., Kim, J.H., Lovell, D.P., Macgregor, J.T., Moore, M.M., van Benthem, J., White, P.A., Zeiger, E., and Thybaud, V. (2013), Quantitative approaches for assessing dose-response relationships in genetic toxicology studies. *Environ Mol Mutagen,* **54**, 8-18.
- Goodman and Gilman, (2002), The Pharmacological Basis of Therapeutics, 10th Edition, New York, McGraw-Hill.
- Green S, Auletta A, Fabricant J, Kapp R, Manandhar M, Sheu CJ, Springer J, Whitfield B, (1985), Current status of bioassays in genetic toxicology--the dominant lethal assay. A report of the U.S. Environmental Protection Agency Gene-Tox Program. Mutat. Res. 154, 49-67.

- Guerard, M., Marchand, C., Plappert-Helbig U., *(*2014*),* Influence *of* Experimental Conditions *on* Data Variability *in the* Liver Comet Assay, Environmental *and* Molecular Mutagenesis, Vol. 55/2, pp. 114-21
- Guo, Chih-Hung, Lu, Yi-Fa, & Hsu, Guoo-Shyng Wang (2005), The influence of aluminum exposure on male reproduction and offspring in mice, Environ Tox Pharm 20:135-141
- Hamada S, Sutou S, Morita T, Wakata A, Asanami S, Hosoya S, Ozawa S, Kondo K, Nakajima M, Shimada H, Osawa K, Kondo Y, Asano N, Sato S, Tamura H, Yajima N, Marshall R, Moore C, Blakely D, Schectman L, Weaver J, Torous D, Proudlock R, Ito S, Namiki C, Hayashi M, (2001), 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). Environ. Molec. Mutagen. 37, 93-110.
- Hayashi M, Dearfield K, Kasper P, Lovell D, Martus H, Thybaud V, (2011), Compilation and use of genetic toxicity historical control data. Mutat. Res. 723, 87-90.
- Hayashi M, Tice RR, MacGregor JT, Anderson D, Blakey DH, Kirsh-Volders M, Oleson FB Jr, Pacchierotti F, Romagna F, Shimada H, (1994), *In vivo* rodent erythrocyte micronucleus assay. Mutat Res 312, 293-304.
- Heddle, J.A., H.-J. Martus and G.R. Douglas, (2003), "Treatment and Sampling Protocols for Transgenic Mutation Assays", Environ. Mol. Mutagen., 41: 1-6.
- Henderson, G.R., (2011), *Six Sigma Quality Improvement with Minitab*, 2nd edition. Chichester, John Wiley & Sons Ltd
- Hoeijmakers JH, 2009. DNA damage, aging, and cancer. New Engl. J. Med. 361, 1475-1485.Honma, M., Momose, M., Sakamoto, H., Sofuni, T. and Hayashi, M., (2001), Spindol poisons induce allelic loss in mouse lymphoma cells through mitotic non-disjunction. Mutat. Res., 493 (1-2): 101-114.
- Honma, M., (2011), Cytotoxicity measurement in *in vitro* chromosome aberration test and micronucleus test, *Mutation Research*, Vol. 724, pp. 86-87.
- Hu Q, Zhu S., (1990), Induction of chromosomal aberrations in male mouse germ cells by uranyl fluoride containing enriched uranium, Mutation Res. 244, 209-214.
- Johnson, G.E., Soeteman-Hernandez, L.G., Gollapudi, B.B., Bodger, O.G., Dearfield, K.L., Heflich, R.H., Hixon, J.G., Lovell, D.P., MacGregor, J.T., Pottenger, L.H., Thompson, C.M., Abraham, L., Thybaud, V., Tanir, J.Y., Zeiger, E., van Benthem, J., and White, P.A., (2014), Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. *Environ Mol Mutagen*, **55**, 609-623.
- Kalweit, S. et al., (1999), Chemically induced micronucleus formation in V79 cells – comparison of three different test approaches, *Mutation Research*, Vol. 439/2, pp. 183-190.

- Kim BS, Cho M, Kim HJ, (2000), Statistical analysis of *in vivo* rodent micronucleus assay. Mutat. Res 469, 233-241.
- Kirkland D, Fowler P, (2010), Further analysis of Ames-negative rodent carcinogens that are only genotoxic in mammalian cells *in vitro* at concentrations exceeding 1 mM, including retesting of compounds of concern. Mutagenesis, 25, 539-553.
- Kirkland D, Kasper P, Müller L, Corvi R, Speit G, (2008), Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: A follow-up to an ECVAM workshop, Mutat. Res. 653, 99-108.
- Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F, Elhajouji A, Glatt H, Hastwell P, Hayashi M, Kasper P, Kirchner S, Lynch A, Marzin D, Maurici D, Meunier JR, Müller L, Nohynek G, Parry J, Parry E, Thybaud V, Tice R, van Benthem J, Vanparys P, White P, (2007), How to reduce false positive results when undertaking *in vitro* genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop. Mutat. Res. 628, 31-55.
- Kirkland D, Speit G,. (2008), Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens III. Appropriate follow-up testing *in vivo*. Mutat Res 654, 114-132.
- Kirkland DJ., (1994) Report of the in vitro sub-group, Mutat. Res. 312: 211-215.
- Kirsch-Volders, M. et al., (2004), Corrigendum to "Report from the *in vitro* micronucleus assay working group", *Mutation Research*, 564, 97-100.
- Kirsch-Volders, M. et al., (2011), The in vitro MN assay in 2011: origin and fate, biological significance, protocols, high throughput methodologies and toxicological relevance, *Archives of Toxicology*, Vol. 85/8, pp. 873-99.
- Kissling, G.E., Dertinger, S.D., Hayashi, M. & MacGregor, J.T., (2007), Sensitivity of the erythrocyte micronucleus assay: dependence on number of cells scored and inter- animal variability. Mutation Research 634 235–240
- Kligerman, A.D., Bishop, J.B., Erexson, G.L., Price, H.C., O'Connor, R.W., Morgan, D.L., & Zeiger, E., (1994), Cytogenetic and Germ Cell Effects of Phosphine Inhalation by Rodents: II. Subacute Exposures to Rats and Mice, EMM 24:301-306
- Lang, R., & Adler, I-D., (1977), Heritable translocation test and dominant--lethal assay in male mice with methyl methanesulfonate, Mutat Res 48:75-88
- Lambert IB, Singer TM, Boucher SE, Douglas GR, (2005). Detailed review of transgenic rodent mutation assays. Mutat Res. 590, 1-280.
- Lemieux CL, Douglas GR, Gingerich J, Phonethepswath S, Torous DK, Dertinger, SD, Phillips DH, Arlt VM, White PA., (2011), Simultaneous measurement of benzo[a]pyrene-induced Pig-a and lacZ mutations, micronuclei and DNA adducts in Muta™ Mouse. Environ Mol Mutagen. Dec;52(9):756-65.

- Lloyd, M. and Kidd, D., (2012), The Mouse Lymphoma Assay. Springer Protocols: Methods in Molecular Biology 817, Genetic Toxicology Principles and Methods, ed. Parry and Parry, Humana Press. ISBN 978-1-61779-420-9, 35-54.
- Lorenzo, Y., Costa, S., Collins, A.R., Azqueta, A. (2012), DNA repair and cytotoxicity*:* Hedgehogs *are* not dead. Mutagenesis 28(4*):*427-32.
- Lorge, E., Hayashi, M., Albertini, S. and Kirkland, D., (2008), Comparison of different methods for an accurate assessment of cytotoxicity in the *in vitro* micronucleus test. I. Theoretical aspects, *Mutation Res.*, 655, 1-3.
- Lovell DP, Anderson D, Albanese R, Amphlett GE, Claire G, Ferguson R, Richold M, Papworth DG Savage JRK., (1989), Statistical Analysis of *In vivo* Cytogenetic Assays In: D.J. Kirkland (Ed.) Statistical Evaluation of Mutagenicity Test Data. UKEMS Sub-Committee on Guidelines for Mutagenicity Testing, Report, Part III. Cambridge University Press, Cambridge, UK, pp. 184-232.
- Lovell, D.P., Anderson, D., & Jemkinson, P.C., (1987), The use of a battery of strains of mice in a factorial design to study the induction of dominant lethal mutations, Mutat Res 187:37-44
- Luippold HE, Gooch PC, Brewen JC., (1978), The production of chromosome aberrations in various mammalian cells by triethylenemelamine, Genetics, 88, 317-326.
- MacGregor, J.T., Frötschl, R., White, P.A., Crump, K.S., Eastmond, D.A., Fukushima, S., Guérard, M., Hayashi, M., Soeteman-Hernández, L.G., Johnson, G.E., Kasamatsu, T., Levy, D.D., Morita, T., Müller, L., Schoeny, R., Schuler, M.J., and Thybaud, V., (2015), IWGT report on quantitative approaches to genotoxicity risk assessment II. Use of point-of-departure (PoD) metrics in defining acceptable exposure limits and assessing human risk *Mutat Res*., 783:66–78.
- MacGregor, J.T., Frötschl, R., White, P.A., Crump, K.S., Eastmond, D.A., Fukushima, S., Guérard, M., Hayashi, M., Soeteman-Hernández, L.G., Kasamatsu, T., Levy, D.D., Morita, T., Müller, L., Schoeny, R., Schuler, M.J., Thybaud, V., and Johnson, G.E., (2015), IWGT report on quantitative approaches to genotoxicity risk assessment I. Methods and metrics for defining exposure–response relationships and points of departure (PoDs) *Mutat Res*., 783:55–65.
- Madrigal-Bujaidar E, Madrigal-Santillan E, Alvarez-Gonzalez I, Baez R, Marquez P., (2008), Micronuclei Induced by Imipramine and Desipramine in Mice: A Subchronic Study. Bas. Clinic. Pharmacol. Toxicol. 103, 569-573.
- Manjanatha MG, Shelton SD, Haber L, Gollapudi B, MacGregor JA, Rajendran N, Moore MM.(2015), Evaluation of cII mutations in lung of male Big Blue mice exposed by inhalation to vanadium pentoxide for up to 8 weeks, Mutat Res Genet Toxicol Environ Mutagen., 789-790:46-52.

- Marchetti F, Bishop J, Cosentino L, Moore II D, Wyrobek AJ., (2004), Paternally transmitted chromosomal aberrations in mouse zygotes determine their embryonic fate. Biology of Reproduction 70:616-624

- Mei, N.,Guo, X. and Moore, M.M. (2014), Methods for Using the Mouse Lymphoma Assay to Screen for Chemical Mutagenicity and Photo-Mutagenicity, in Eds:Yan Z and Caldwell Optimization in Drug Discover: In Vitro Methods, 2nd Edition, GW; Humana Press, Totowa, NJ.Miltenburger HG, Träger H, Adler ID, Schmaltz A, Rathenberg R, Schwegler H, Pawlowitzki IH, Nautsch C, Müller D.,(1978), The effect of isoniazid (INH) on Chinese hamster and mouse spermatogonia, Hum Genet. 42, 44-49.

- Montgomery, D., (2005),. *Introduction to Statistical Quality Control*. Hoboken, New Jersey: John Wiley & Sons, Inc

- Moore, M.M., and D. Clive., (1982), The quantitation of $TK^{-/-}$ and HGPRT- mutants of L5178Y/$TK^{+/-}$ mouse lymphoma cells at varying times post treatment, Environmental and Molecular Mutagenesis 4:499-519.

- Moore, M.M., M. Honma, J. Clements, T. Awogi, G. Bolcsfoldi, J. Cole, B. Gollapudi, K. Harrington-Brock, A. Mitchell, W. Muster, B. Myhr, M. O'Donovan, M.-C. Ouldelhkim, R. San, H. Shimada, and L.F. Stankowski, Jr., (2000), The mouse lymphoma thymidine kinase locus (tk) gene mutation assay: International Workshop on Genotoxicity Test Procedures (IWGTP) Workgroup Report, Environmental and Molecular Mutagenesis 35:185-190.

- 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, and L.F. Stankowski, Jr., (2002), Mouse Lymphoma Thymidine Kinase Locus Gene Mutation Assay: Follow-Up International Workshop on Genotoxicity Test Procedures, New Orleans, Louisiana, April 2000, Environmental and Molecular Mutagenesis 40:292-299.

- Moore MM, Honma M, Clements J, Bolcsfoldi G, Cifone M, Delongchamp R, Fellows M, Gollapudi B, Jenkinson P, Kirby P, Kirchner S, Muster W, Myhr B, O'Donovan M, Oliver J, Omori T, Ouldelhkim MC, Pant K, Preston R, Riach C, San R, Stankowski LF Jr, Thakur A, Wakuri S, Yoshimura I., (2003), Mouse lymphoma thymidine kinase gene mutation assay: International Workshop on Genotoxicity tests Workgroup report – Plymouth, UK 2002. Mutat. Res. 540, 127-140.

- 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. Thakur, F. Van Goethem, S. Wakuri, and I. Yoshimura. (2006), Mouse Lymphoma Thymidine Kinase Gene Mutation Assay: Follow-up Meeting of the International Workshop on Genotoxicity Tests – Aberdeen, Scotland, 2003 – Assay acceptance criteria, positive controls, and data evaluation, Environmental and Molecular Mutagenesis 47:1-5.

- Moore, M.M., R.H. Heflich, L. Haber, B.C. Allen, A. M. Shipp, and R.L. Kodell, (2008), Analysis of *in vivo* mutation data can inform cancer risk assessment, Regulatory Toxicology and Pharmacology 51:151-161.
- Morita T, Honma M, Morikawa K, (2012), Effect of reducing the top concentration used in the *in vitro* chromosomal aberration test in CHL cells on the evaluation of industrial chemical genotoxicity. Mutat. Res. 741, 32-56.
- Mullins, E., (2003), *Statistics for the Quality Control Chemistry Laboratory*. The Royal Society of Chemistry, Cambridge.
- OECD, (1987). Introduction to the OECD Guidelines on Genetic Toxicology Testing and Guidance on the Selection and Application of Assays." In the Third Addendum to the OECD Guidelines for Testing of Chemicals. Paris, France: OECD Publications, March 1, 1987. [latest edition: 10th ed. Oct. 1998]
- OECD, (2000), Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation, Series on Testing and Assessment, N°19, ENV/JM/MONO(2000)7, OECD, Paris.
- OECD, (2009), Detailed Review Paper on Transgenic Rodent Mutation Assays, Series on Testing and Assessment, N° 103, ENV/JM/MONO 7, OECD, Paris.
- OECD, (2011), OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects, Test No. 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays, 2011, http://www.oecd-ilibrary.org/environment/test-no-488-transgenic-rodent-somatic-and-germ-cell-gene-mutation-assays_9789264122819-en;jsessionid=k3voawso6gsv.delta
- OECD, (2014a), Reports of the JaCVAM initiative international pre-validation and validation studies of the *in vivo* rodent alkaline comet assay for the detection of genotoxic carcinogens, Series on Testing and Assessment, Nos. 195 and 196, OECD Publishing, Paris.
- OECD, (2014b), Report On Statistical Issues Related To OECD Test Guidelines (TGs) On Genotoxicity, Series on Testing and Assessment No. 198, ENV/JM/MONO(2014)12, OECD Publishing, Paris.
- OECD, (2014c), Document Supporting The WNT Decision To Implement Revised Criteria For The Selection Of The Top Concentration in the in vitro Mammalian Cell Assays on Genotoxicity (Test Guidelines 473, 476 and 487), ENV/JM/TG/RD(2014)20, OECD Publishing, Paris.
- Palo AK, Pandit RS, Choudhury RC., (2011), Vinblastine-induced cytogenotoxicity in spermatogonia and its transmission in the germline cells of Swiss mice, J. Environ. Pathol. Toxicol. Oncol. 30, 113-121.
- Palo AK, Sahoo D, Choudhury RC., (2009,) Cytosine arabinoside-induced cytogenotoxicity in bone marrow and spermatogonial cells of mice and its potential transmission through the male germline, Mutation Res. 673, 29-36.

- Palo AK, Sahu P, Choudhury RC., (2005), Etoposide-induced cytogenotoxicity in mouse spermatogonia and its potential transmission, J Appl Toxicol 25, 94-100.
- Parsons BL, Manjanatha MG, Myers MB, McKim KL, Shelton SD, Wang Y, Gollapudi BB, Moore NP, Haber LT, Moore MM., (2013), Temporal changes in K-ras mutant fraction in lung tissue of big blue B6C3F_1 mice exposed to ethylene oxide, Toxicol Sci. 2013 Nov;136(1):26-38.
- Ogawa I, Furukawa S, Abe M, Tanaka Y, Hayashi S, Usuda K., (2009), Multi-endpoint genotoxic assay using L5178Y (Tk(+/-) -3.7.2c) cells, J Toxicol Sci. 2009 Oct;34(5):547-53.
- Parry JM, Parry E, Phrakonkham P, Corvi R., (2010), Analysis of published data for top concentration considerations in mammalian cell genotoxicity testing. Mutagenesis, 25, 531-538.
- Parton JW, Hoffman WP, Garriott ML, (1996), Validation of an automated image analysis micronucleus scoring system. Mutat Res 370, 65-73.
- Perry PE, Thomson EJ., (1984), The methodology of sister chromatid exchanges. In: Handbook of Mutagenicity Test Procedures, 2nd Edition, Kilbey BJ, Legator M, Nichols W, Ramel C (eds.), Elsevier Scientific, Amsterdam, pp 485-530.
- Pfuhler S, Kirkland D, Hayashi M, Vanparys P, Carmichael P, Dertinger S, Eastmond D, Elhajouji A, Krul C, Rothfuss A, Schoening G, Smith A, Speit G, Thomas C, van Benthem J, Corvi R., (2009), Reduction of use of animals in regulatory genotoxicity testing: Identification and implementation opportunities- Report from an ECVAM workshop. Mutat. Res. 680, 31-42.
- Pfuhler, S. et al., (2011), In vitro genotoxicity test approaches with better predictivity: Summary of an IWGT workshop, *Mutation Research*, Vol. 723/2, pp. 101-107.
- Rao,U.N., Aravindakshan,M.,& Chauhan,P.S., (1994), Mutat Res 311:69-76
- Rathenberg R., (1975), Cytogenetic effects of cyclophosphamide on mouse spermatogonia, Humangenetik 29, 135-140.
- Rothfuss A, Honma M, Czich A, Aardema MJ, Burlinson B, Galloway S, Hamada S, Kirkland D, Heflich RH, Howe J, Nakajima M, O'Donovan M, Plappert-Helbig U, Preistley C, Reicio L, Schuler M, Uno Y, Martus H., (2011), Improvement of *in vivo* genotoxicity assessment: combination of acute tests and integration into standard toxicity testing. Mutat Res. 723, 108-120.
- Ryan TP., (2000), Statistical methods for quality improvement. 2nd ed. New York: John Wiley and Sons.
- Ryan, T. P., (2011), Statistical Methods for Quality Improvement. 3rd ed. John Wiley and Sons, New York

- Schisler, M.R., M.M. Moore, and B.B. Gollapudi, (2013), In Vitro Mouse Lymphoma (L5178Y TK+/--3.7.2C) Forward Mutation Assay, in A. Dhawan and M. Bajpayee (Eds.) Protocols in Genotoxicity Assessment, Springer Protocols, Humana Press. p. 27-50.
- Scott D, Galloway SM, Marshall RR, Ishidate M, Brusick Jr D, Ashby J, Myhr BC., (1991), Genotoxicity under extreme culture conditions. A report from ICPEMC Task Group 9. Mutat. Res. 257, 147–204.
- Shelby, M.D., Cain, K.T., Hughes, L.A., Braden, P.W., & Generoso, W.M., (1991,) Mouse dominant lethal and bone marrow micronucleus studies on methyl vinyl sulfone and divinyl sulfone, Mutat Res 250:431-437
- Shelby, M.D., Gutierrez-Espeleta, G.A., Generoso, W.M., & McFee. A.F., (1986), Dominant lethal effects of acrylamide in male mice, Mutat Res 173:35-40
- Slatter MA and Gennery AR., (2010), Primary immunodeficiencies associated with DNA-repair disorders. Expert Rev. Mol. Med. 12:e9.
- Smith, C.C, et al., (2008), Recommendations for design of the rat Comet assay, Mutagenesis, Vol. 23/3, pp. 233-40.
- Günter Speit, Hajime Kojima, Brian Burlinson, Andrew Collins, Peter Kasper, Ulla Plappert, Yoshifumi Uno, Marie Vasquez, Carol Beevers, Marlies de De Boeck, Patricia Escobar, Sachiko Kitamoto, Kamala Pant, Stefan Pfuhler, Jin Tanaka, Dan D. Levy, (2015) Critical issues with the in vivo comet assay: a report of the comet assay working group in the 6th International Workshop on Genotoxicity Testing (IWGT). Mutation Research 783: 6-12
- Styles JA, Clark H, Festing MF, Rew DA., (2001), Automation of mouse micronucleus genotoxicity assay by laser scanning cytometry. Cytometry 44, 153-155.
- Sudman, P.D., Rutledge, J.C., Bishop, J.B., Generoso, W.M., (1992), Bleomycin: female-specific dominant lethal effects in mice, Mutat Res 296:143-156
- Tates AD, Natarajan AT., (1976), A correlative study on the genetic damage induced by chemical mutagens in bone marrow and spermatogonia of mice. 1. CNU-ethanol, Mutation Res. 37, 267-278.
- Thybaud, V., S. Dean, T. Nohmi, J. de Boer, G.R. Douglas, B.W. Glickman, N.J. Gorelick, J.A. Heddle, R.H. Heflich, I. Lambert, H.-J. Martus, J.C. Mirsalis, T. Suzuki and N. Yajima, (2003), "*In vivo* Transgenic Mutation Assays", Mutation Res., 540: 141-151.
- Thybaud V, Macgregor JT, Müller L, Crebelli R, Dearfield K, Douglas G, Farmer PB, Gocke E, Hayashi M, Lovell DP, Lutz WK, Marzin D, Moore M, Nohmi T, Phillips DH, Van Benthem J., (2011), Strategies in case of positive *in vivo* results in genotoxicity testing, Mutat Res., 723(2):121-128.
- Tice RR, Andrews PW, Singh NP,. (1990), The single cell gel assay: a sensitive technique for evaluating intercellular differences in DNA damage and repair. Basic Life Sci. 53, 291-301.

- Torous D, Dertinger S, Hall N, Tometsko C., (2000), Enumeration of micronucleated reticulocytes in rat 1 peripheral blood: a flow cytometric study. Mutat. Res. 465, 91-99.
- Tweats DJ, Blakey D, Heflich RH, Jacobs A, Jacobsen SD, Morita T, Nohmi T, O'Donovan MR, Sasaki YF, Sofuni T, Tice R., (2007a), Report of the IWGT working group on strategies and interpretation of regulatory *in vivo* tests I. Increases in micronucleated bone marrow cells in rodents that do not indicate genotoxic hazards. Mutat. Res. 627, 78-91.
- Tweats DJ, Blakey D, Heflich RH, Jacobs A, Jacobsen SD, Morita T, Nohmi T, O'Donovan MR, Sasaki YF, Sofuni T, Tice R., (2007b) Report of the IWGT working group on strategies and interpretation of regulatory *in vivo* tests II. Identification of *in vivo*-only positive compounds in the bone marrow micronucleus test. Mutat. Res. 627, 92-105.
- UNEP, (2000), Stockholm Convention on Persistent Organic Pollutants, United Nations Environment Programme (UNEP). Available at: [http:/www.pops.int/]
- van Buul, PPW, Goudzwaard JH., (1980), Bleomycin-induced structural chromosomal aberrations in spermatogonia and bone-marrow cells of mice, Mutation Res. 69, 319-324.
- Wang, J., J.R. Sawyer, L. Chen, T. Chen, M. Honma, N. Mei, and M.M. Moore., (2009), The mouse lymphoma assay detects recombination, deletion, and aneuploidy, Toxicological Sciences 109:96-105.
- Witt, K.L., Hughes, L.A., Burka, L.T. McFee, A. F. Mathews, J.M. Black, SL. & Bishop, J. B., (2003), Mouse bone marrow micronucleus test results do not predict the germ cell mutagenicity of n-hydroxymethylacrylamide in the mouse dominant lethal assay, EMM 41:111-120
- Yauk CL, Aardema, M.J, van Benthem, J., Bishop, J.B., Dearfield, K.L., DeMarini, D.M., Dubrova, Y.E., Honma, M., Lupski, J.R., Marchetti, F., Meistrich, M.L., Pacchierotti, F., Stewart, J., Waters, M.D., Douglas, G.R. (2015) Approaches for Identifying Germ Cell Mutagens: Report of the 2013 IWGT Workshop on Germ Cell Assays. Mutat Res Genet Toxicol Environ Mutagen., 783:36-54.
- Yoshifumi Uno, Takeshi Morita, Mirjam Luijten, Carol Beevers, Shuichi Hamada, Satoru Itoh, Wakako Ohyama, Hironao Takasawa (2015a) Recommended protocols for the liver micronucleus test: Report of the IWGT working group, Mutation Research 783: 13-18
- Yoshifumi Uno, Takeshi Morita, Mirjam Luijten, Carol Beevers, Shuichi Hamada, Satoru Itoh, Wakako Ohyama, Hironao Takasawa (2015b). Micronucleus test in rodent tissues other than liver or erythrocytes: Report of the IWGT working group, Mutation Research 783: 19-22
- Zhang H, Zheng RL, Wang RY, Wei ZQ, Li WJ, Gao QX, Chen WQ, Wang ZH, Han GW, Liang JP., (1998), Chromosomal aberrations induced by $^{12}C^{6}$+ heavy ion irradiation in spermatogonia and spermatocytes of mice, Mutation Res. 398, 27-31.
- Zhang H, Zhao W, Wang Y, Li N, Wu Z, Liu Y, Chen J, Cai Y., (2008), Induction of cytogenetic adaptive response in spermatogonia and spermatocytes by pre-exposure of mouse testis to low dose $^{12}C^{6}$+ ions, Mutation Res. 653, 109-112.

1802

6 APPENDIX A. TREATMENT AND SAMPLING TIMES FOR *IN VIVO* GENETIC TOXICOLOGY TGS IN MICE (THE APPLICABLE TGS SHOULD BE CONSULTED FOR MORE DETAILED INFORMATION).

Test	Treatment	Sampling	
TG 474 (mammalian erythrocyte micronucleus):	Single	Bone Marrow: at least 2x, 24 - 48 hr after treatment	Peripheral Blood: at least 2x, 36 - 72 hr after treatment
	2 daily	BM: once 18 - 24 hr after treatment	PB: at leas once 36 - 48 hr after treatment
	3 or more daily	BM: 24 hr after treatment	PB: 40 hr after treatment
TG 475 (mammalian bone marrow chromosome aberration):	Single	2x: first; 1.5 cell cycle lengths after treatment; second, 24 hr later	
TG 478 (Rodent dominant lethal):	1-5 daily	8 (mouse) or 10 (rat) weekly matings following last treatment	
	28 daily	4 (mouse) weakly matings following last treatment	
TG 483 (Mammalian spermatogonial chromosome aberration):	Single	Highest dose: 2x, 24 and 48 hr after treatment Other doses: 1x, 24 hr after treatment	
	Extended regimens can be used (*e.g.* 28 daily)	Same as for single treatment	
TG 488 (transgenic rodent somatic and germ cell gene mutation):	28 daily	Somatic tissues: 3 days following treatment: however, for slowly dividing tissues longer sampling times (e.g. 28 days) may be used. Germ Cells: seminiferous tubule cells, 3 days; sperm: 49 days.	
TG 489 (mammalian comet)	Single	2-6 hr and 26 hr after treatment	
	2 or more daily	2-6 hr after treatment	

7 APPENDIX B. COMPILATION OF PUBLISHED NEGATIVE (VEHICLE) CONTROL DATA FOR THE DOMINANT LETHAL TEST

Female Strain	Male Strain	No. of Females	Total implantations	Total Resorptions	% Resorptions	St Dev	Reference
(101xC3H)F1	(101xC3H)F1	331	2441	189	7.7	3.0	Generoso et al (1982)
(101xC3H)F1	(101xC3H)F1	294	3366	267	7.9	1.3	Ehling & Neuhauser-Klaus (1989)
(101xC3H)F1	(101xC3H)F1	540	5830	581	10.0	1.0	Ehling & Neuhauser-Klaus (1988)
(101xC3H)F1	(101xC3H)F1	169	1266	117	9.3	2.4	Ehling et al (1968)
(101xC3H)F1	(101xC3H)F1	304	2123	148	7.0	3.3	Generoso et al (1975)
(101xC3H)F1	(101xC3H)F1	506	5375	490	9.1	1.1	Ehling (1971)
		2582	25175	2206	8.8[1]	1.1	Ehling (1971)
(102xC3H)F1	(102xC3H)F1	116	1221	98	8.0	2.8	Adler et al (1998)
(102xC3H)F1	(102xC3H)F1	494	5467	466	8.5	1.6	Adler et al (2002)
(102xC3H)F1	(102xC3H)F1	349	3965	359	9.0	1.7	Ehling & Neuhauser-Klaus (1995)
(102xC3H)F1	(102xC3H)F1	353	4059	350	8.6	1.2	Ehling & Neuhauser-Klaus (1995)
(102xC3H)F1	(102xC3H)F1	229	2665	257	9.6	2.2	Adler et al (1995)
(102xC3H)F1	(102xC3H)F1	341	3921	403	10.3	1.9	Ehling & Neuhauser-Klaus (1991)
(102xC3H)F1	(102xC3H)F1	589	6528	525	8.0	1.3	Ehling & Neuhauser-Klaus (1991)
		2920	32764	2851	8.7[1]	0.8	Ehling & Neuhauser-Klaus (1991)
(C3Hx101)F1	(C3Hx101)F1	90	704	44	6.3	1.5	Shelby et al (1986)
(C3Hx101)F1	(101xC3H)F1	50	407	34	8.5	2.2	Generoso et al (1982)
		140	1111	78	7.0[1]	1.6	Generoso et al (1982)
(C3HxC57BL)F1	(101xC3H)F1	67	713	47	6.6	0.9	Generoso et al (1982) 163
BALB/c	BALB/c	24	181	54	29.8	10.1	Lovell et al (1987)
BALB/c	BALB/c	60	562	25	4.4	1.2	Blaszkowska (2010)
B6CF1	Various	129	1296	71	5.5	4.3	Bishop et al (1983)
B6C3F1	Various	128	1224	53	4.3	1.2	Bishop et al (1983)
B6C3F1	(101xC3H)F1	388	4340	183	4.2	2.0	Witt et al (2003)
B6C3F1	(101xC3H)F1	91	957	46	4.6	0.1	Witt et al (2003)
B6C3F1	(101xC3H)F1	168	1855	89	4.8	0.6	Sudman et al (1992)
B6C3F1	B6C3F1	290	2846	134	3.1	1.7	Kligerman et al (1994)
		1065	11222	505	4.5[1]	0.7	Kligerman et al (1994)
CBB6F1	CBB6F1	45	461	45	9.8	4.6	Lovell et al (1987)
CBA/Ca	CBA/Ca	24	198	23	11.6	1.1	Lovell et al (1987)
C57BL/6J	DBA/2J	129	1115	67	6.0	3.1	Barnett et al (1992)

C57BL/6J	DBA/2J	199	1832	118	6.4	2.6	Barnett & Lewis (2003)
C57BL/6J	C57BL/6J	42	329	52	15.8[1]		Rao et al (1994)
		370	3276	237	7.2[1]	5.5	Rao et al (1994)
CD-1	CD-1	46	572	37	6.5		Anderson et al (1998)
CD-1	B6C3F1	178	1983	131	6.6	3.1	Dunnick et al (1984)
CD-1	CD-1				3.6	0.7	Guo et al (2005) Environ
CD-1	CD-1	447	5217	299	5.7	1.7	Anderson et al (1976)
CD-1	CD-1	323	4035	289	7.2	0.7	Anderson et al (1976)
CD-1	CD-1	702	8575	523	6.1	0.9	Anderson et al (1977)
							Anderson et al (1981)
		1696	20382	1279	6.3[1]	1.3	Anderson et al (1981)
(SECxC57BL)F1	(C3Hx101)F1	39	359	8	2.2		Shelby et al (1986)
(SECxC57BL)F1	(C3Hx101)F1	733	7098	260	3.7	2.3	Generoso et al (1995)
(SECxC57BL)F1	(C3Hx101)F1	615	5883	198	3.4	1.3	Generoso et al (1988)
(SECxC57BL)F1	(C3Hx101)F1	386	4021	115	2.9	1.1	Generoso et al (1996)
(SECxC57BL)F1	(C3Hx101)F1	288	2969	91	3.1	1.3	Shelby et (1991)
(SECxC57BL)F1	(101xC3H)F1	50	514	20	4.0	1.5	Generoso et al (1982)
(SECxC57BL)F1	(101xC3H)F1	200	2111	55	2.6	1.7	Sudman et al (1992)
		2311	22955	747	3.3[1]	0.6	Sudman et al (1992)
NMRI	(102xC3H)F1	103	1535	93	6.1	1.3	Adler et al (1998)
NMRI	NMRI	137	1692	83	4.9	0.5	Lang & Adler (1977)
		240	3227	176	5.5[1]	0.8	Lang & Adler (1977)
Swiss Albino	Swiss Albino	243	2693	282	10.5	1.3	Attia (2012) Arch
Swiss Albino	Swiss Albino	243	2672	274	10.3	1.0	Attia (2012) Arch
Swiss Albino	Swiss Albino	322	3541	357	10.1	1.3	Attia et al (2015)
		808	8906	913	10.1[1]	0.2	Attia et al (2015)
Swiss	Swiss	275	2804	164	5.8	1.2	Rao et al (1994)
Swiss	C57BL	71	722	32	4.4	1.7	Rao et al (1994)
Swiss	CBA	76	710	26	3.7	2.5	Rao et al (1994)
		422	4236	222	3.7[1]	1.1	Rao et al (1994)
T-Stock	(CH3x101)F1	755	6851	1125	16.4	3.1	Shelby et al (1986)
T-Stock	(CH3x101)F1	822	7713	935	12.1	2.6	Generoso et al (1995)
T-Stock	(CH3x101)F1	323	3116	472	15.2	2.6	Shelby et al (1991)
		1900	17680	2532	14.3[1]	2.2	Shelby et al (1991)

1810 • [1] weighted mean

1811

8 APPENDIX C. COMPILATION OF PUBLISHED NEGATIVE (VEHICLE) CONTROL DATA FOR THE MOUSE SPERMATOGONIAL CHROMOSOMAL ABERRATION TEST.

No. of mice	Strain	No. of cells	No of aberration/cell x 100					No. of aberrations/ cell x 100 (excluding gaps)	% aberrant cells (excluding gaps)	Ref.
			Gaps	Chromatid type		Chromosome type				
				Breaks	Exchanges	Breaks	Exchanges			
24	(101x C3H)F1	1600	0.56	0.13	0	0	0	0.13	0.13	1 Adler, 1982
28	(101x C3H)F1	1400	0.79	0.14	0	0	0	0.14	0.14	1 Adler, 1982
10	(101x C3H)F1	20,000	0.63	0.14	0.005	0	0	0.15	0.15	1 Adler, 1982
4	(101x C3H)F1	400	0	0	0	0	0	0	0	1 Adler, 1982
35	(101 x C3H)F1	1750	0.97	0.11	0	0	0	0.11	0.11	2 Adler, 1974
6	CD1	700	NR	0.54	0	0	0	0.54	NR	3 Luippold, 1978
1	CBA	300	NR	0.33	0	0	0	0.33	0.33	4 Tates and Natarajan, 1976
2	Swiss	250	NR	0	0	0	0	0	0	5 vanBuul and Goudzwaard, 1980
6	(101 x C3H)F1	600	6.0	0.5	0	0	0	0.50	0.5	6 Adler and El-Tarras
6	(102 x C3H)F1	600	5.0	0.83	0	0	0	0.83	0.83	7 Ciranno and Adler, 1991
20	Balb/c	2000	0.05	0.05	0	0	0	0.05	0.05	8 Hu and Zu, 1990
5	Kun-Ming	250	0	0.4	0	0	0	0.40	0.4	9 Zhang *et al.*, 1998
7	Kun-Ming	350	0.29	0.29	0	0	0	0.29	0.29	10 Zhang *et al.*, 2008
6	Swiss	274	3.65	1.46	0	0	0	1.46	1.46	11,12,13 * Palo *et al.*, 2011; Palo *et al.*, 2009, Palo *et al.*, 2005
5	Swiss	1000	4.9	1.20	0	0	0	1.20	1.20	14 Ciranno *et al.*, 1991
6	NMRI	300	0.33	0.33	0	0	0	0.33	0.33	15 Rathenberg, 1975
8	(101 x C3H)F1	800	0	0	0	0	0	0	0	16 Mltenburger, 1978
6	A-AJAX	560	0.90	0	0	0	0	0	0	16 Mltenburger, 1978
32	NMRI	3200	0.34	0.13	0	0	0	0.13	0.13	16 Mltenburger, 1978

* These three papers report the same control data. NR: Not reported

9 APPENDIX D. DEFINITIONS

Administration period: the total period during which an animal is dosed.

Aneugen: any substance or process that, by interacting with the components of the mitotic and meiotic cell division cycle, leads to aneuploidy in cells or organisms.

Aneuploidy: any deviation from the normal diploid (or haploid) number of chromosomes by a single chromosome or more than one, but not by entire set(s) of chromosomes (polyploidy).

Apoptosis: programmed cell death triggered by DNA damage and characterized by a series of steps leading to the disintegration of cells into membrane-bound particles that are then eliminated by phagocytosis or by shedding. It is a reliable marker for DNA damage (genotoxicity), but only when it can be differentiated from other types of necrosis.

Base pair substitution: A gene mutation characterized by the substitution of one base pair for another in the DNA.

Cell proliferation: the increase in cell number as a result of mitotic cell division. Reduction in cell proliferation is generally considered cytotoxicity, a key parameter in genotoxicity assays.

Centromere: the DNA region of a chromosome where both chromatids are held together and on which both kinetochores are attached side-to-side.

Chromatid break: structural chromosomal damage consisting of a discontinuity of a single chromatid in which there is a clear misalignment of one of the chromatids.

Chromatid gap: non-staining region (achromatic lesion) of a single chromatid in which there is minimal misalignment of the chromatid.

Chromatid-type aberration: structural chromosome damage expressed as breakage of single chromatids or breakage and reunion between chromatids.

Chromosome-type aberration: structural chromosome damage expressed as breakage, or breakage and reunion, of both chromatids at an identical site.

Chromosome diversity: diversity of chromosome shapes (*e.g.* metacentrique, acrocentriques, *etc.*) and sizes.

Clastogen: a substance which causes structural chromosomal aberrations in populations of cells or organisms.

Clonal expansion: the production by cell division of many cells from a single (mutant) cell.

Cloning efficiency: The percentage of cells plated in a mammalian cell assay that are able to grow into a colony that can be counted.

Comet: The shape that nucleoids adopt after submitted to one electrophoretic field: the head is the nucleus and the tail is constituted by the DNA migrating out of the nucleus in the electric field. The shape resembles a comet.

Concentrations: The final amount of the test chemical in culture medium.

Critical variable/parameter: A protocol variable for which a small change can have a large impact on the conclusion of the assay. Critical variables can be tissue-specific. Critical variables should not be altered, especially within a test, without consideration of how the alteration will

alter an assay response, for example as indicated by the magnitude and variability in positive and negative controls. The test report should list alterations of critical variables made during the test or compared to the standard protocol for the laboratory and provide a justification for each alteration.

Cytokinesis: the process of cell division immediately following mitosis to form two daughter cells, each containing a single nucleus.

Cytokinesis-block proliferation index (CBPI): A measure of cell proliferation consisting of the proportion of second-division cells in the treated population relative to the untreated control).

Cytotoxicity: Cytotoxicity is defined for each specific test (see individual TGs).

Deletion: a gene mutation in which one or more (sequential) nucleotides is lost by the genome.

Dominant lethal mutation: a mutation occurring in a germ cell, or is fixed after fertilization, that causes embryonic or foetal death.

Endoreduplication: a process in which after an S period of DNA replication, the nucleus does not go into mitosis but starts another S period. The result is chromosomes with 4, 8, 16, ...chromatids.

Erythroblast: An early stage of erythrocyte development, immediately preceding the immature erythrocyte, where the cell still contains a nucleus.

Fertility rate: the number of mated pregnant female over the number of mated females.

Forward mutation: a gene mutation from the parental type to the mutant form which gives rise to an alteration of the activity or the function of the encoded protein.

Frameshift mutation: A gene mutation characterized by the addition or deletion of single or multiple base pairs in the DNA molecule.

Gap: an achromatic lesion smaller than the width of one chromatid, and with minimum misalignment of the chromatids. It is not considered a reliable marker of structural chormomal damage because it can be observed after non-genotoxic treatments.

Genotoxic: a general term encompassing all types of DNA or chromosomal damage, including DNA breaks, adducts, rearrangements, mutations, chromosome aberrations, and aneuploidy. Not all types of genotoxic effects result in mutations or stable (transmissible) chromosomal damage.

Insertion: A gene mutation characterized by the addition of one or more nucleotide base pairs into a DNA sequence.

Interphase cells: cells not in the mitotic stage.

Kinetochore: a protein-containing structure that assembles at the centromere of a chromosome to which spindle fibers associate during cell division, allowing orderly movement of daughter chromosomes to the poles of the daughter cells.

Large deletions: deletions in DNA of more than several kilobases. Gene mutation assays vary in their ability to detect large deletions.

Mating interval: the time between the end of exposure and mating of treated males. By controlling this interval, chemical effects on different germ cell types can be assessed. Effects

originating in testicular sperm, condensed spermatids, round spermatids, pachytene spermatocytes, early spermatocytes, dividing spermatogonia, and stem cell spermatogonia are detected at various times after mating.

Micronuclei: Small fragments of nuclear chromosomes, separate from and additional to the main nuclei of cells, produced during telophase of mitosis (meiosis) by lagging chromosome fragments or whole chromosomes..

Mitogen: a chemical substance that stimulates a cell to commence cell division, triggering mitosis (*i.e.* cell division).

Mitotic index: A measure of the proliferation status of a cell population consisting of the ratio between the number of cells in mitosis and the total number of cells in a population, which.

Mitosis: division of the cell nucleus usually divided into prophase, prometaphase, metaphase, anaphase and telophase.

Mitotic recombination: During mitosis, recombination between homologous chromatids possibly resulting in the induction of DNA double strand breaks or in a loss of heterozygosity.

Mutagen: a chemical that induces genetic events that alter the DNA and/or chromosomal structure and that are passed to subsequent generations through clonal expansion.

Mutant frequency (MF): the number of mutant colonies observed divided by the number of cells plated in selective medium, corrected for cloning efficiency (or viability) at the time of selection.

Mutation frequency: The frequency of independently generated mutations. Generally calculated as the number of observed independent mutations divided by the number of cells that are evaluated for the presence of mutations. In the context of the TGs it is used for the transgenic mutation assays in which mutants are sequenced and the mutant frequency is corrected based on the number of mutants found to be siblings (from clonal expansion).

Non-disjunction: A chromosomal aberration characterized by failure of paired chromatids to disjoin and properly segregate to the developing daughter cells, resulting in daughter cells with abnormal numbers of chromosomes.

Normochromatic or mature erythrocyte: A fully matured erythrocyte that has lost the residual RNA that remains after enucleation and/or has lost other short-lived cell markers that characteristically disappear after enucleation following the final erythroblast division.

Numerical aberration: A chromosomal aberration consisting of a change in the number of chromosomes from the normal number characteristic of the animals utilised (aneuploidy).

Phenotypic expression time: The time after treatment during which the genetic alteration is fixed within the genome and any preexisting gene products are depleted to the point that the phenotypic trait is altered and, therefore, can be enumerated using a selective drug or procedure.

Polychromatic or immature erythrocyte: A newly formed erythrocyte in an intermediate stage of development. It stains with both the blue and red components of classical blood stains such as Wright's Giemsa because of the presence of residual RNA in the newly-formed cell. Such newly formed cells are approximately the same as reticulocytes, which are visualised using a vital stain that causes the residual RNA to clump into a reticulum. Other methods, including monochromatic staining of RNA with fluorescent dyes or labeling of short-lived surface markers such as CD71 with fluorescent antibodies, are now often used to identify the immature

erythrocyte. Polychromatic erythrocytes, reticulocytes, and CD71-positive erythrocytes are all immature erythrocytes, though each has a somewhat different developmental distribution.

Polyploidy: A numerical chromosomal aberration consisting of a change in the number of the entire set of chromosomes, as opposed to a numerical change in part of the chromosome set (*cf.* aneuploidy).

Postimplantation loss: the ratio of dead implant in the treated group compared to the ratio of dead to total implants in the control group.

Preimplantation loss: the difference between the number of implants and the number of corpora lutea. It can also be estimated by comparing the total implants per female in treated and control groups.

Reticulocyte: A newly formed erythrocyte stained with a vital stain that causes residual cellular RNA to clump into a characteristic reticulum. Reticulocytes and polychromatic erythrocytes have a similar cellular age distribution.

Relative cell counts (RCC): measure of cell proliferation consisting of a simple ratio of cells at the beginning and end of treatment. Revised TGs discourage using this as a measure of cytotoxicity.

Relative increase in cell count (RICC): A measure of cell proliferation based on the doubling frequency of the cells.

Relative population doubling (RPD): A measure of cell proliferation based on the doubling frequency of the cells .

Relative survival (RS): RS is used as the measure of treatment-related cytotoxicity. RS is cloning efficiency (CE) of cells plated immediately after treatment adjusted by any loss of cells during treatment as compared with cloning efficiency in negative controls (assigned a survival of 100%).

Replication index (RI): A measure of cell proliferation consisting of the proportion of cell division cycles completed in a treated culture, relative to the untreated control, during the exposure period and recovery.

S9 liver fractions: supernatant of liver homogenate after 9000g centrifugation, *i.e.*, raw liver extract.

S9 mix: mix of the liver S9 fraction and cofactors necessary for cytochrome p450 metabolic enzyme activity.

Solvent control: General term to define the negative control cultures receiving the solvent alone used to dissolve the test substance.

Structural chromosomal aberration: a change in chromosome structure detectable by microscopic examination of the metaphase stage of cell division, observed as deletions and fragments, intrachanges or interchanges.

Untreated control: cultures that receive no treatment (*i.e.* neither test chemical nor solvent) but are processed concurrently and in the same way as the cultures receiving the test chemical.