DRAFT UPDATED TEST GUIDELINE 488

Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays

INTRODUCTION

1.	The OECD Test Guidelines for the testing of chemicals are periodically reviewed in light of scientific progress, changing regulatory needs and animal welfare considerations. The original Test Guideline 488 was adopted in 2011. In 2013, a revised guideline was adopted that updated: the age range of animals at the start of the treatment; the sections of reproductive tracts to be sampled for sperm collection; and, the time for rodent spermatogonial stem cells to become mature sperm and reach the cauda epididymis/vas deferens. This present version of the Test Guideline (TG) focuses on the optimization of the assay for the analysis of mutations in male germ cells, the integration of this with the analysis of mutations in somatic tissues, and harmonization with recently revised OECD Test Guidelines for genetic toxicology.

2.	A document that provides succinct information on genetic toxicity testing and an overview of the recent changes that were made to the Test Guidelines for genetic toxicology has been developed (1). Additional information on the main changes introduced to these Test Guidelines was also published (2).

3.	OECD Test Guidelines are available for a wide range of *in vitro* mutation assays that are able to detect chromosomal and/or gene mutations. There are Test Guidelines for several *in vivo* genotoxic endpoints (*i.e.,* chromosomal aberrations, micronuclei, unscheduled DNA synthesis, and DNA strand breaks using the comet assay); however, these do not measure gene mutations. Transgenic Rodent (TGR) mutation assays fulfil the need for practical and widely available *in vivo* tests for gene mutations.

4.	Data from the TGR mutation assays have been reviewed extensively (3) (4) (5). They use transgenic rats and mice that contain multiple copies of chromosomally integrated plasmid or phage shuttle vectors. The transgenes contain reporter genes for the detection of various types of mutations induced *in vivo* by test chemicals. The purpose of the TGR assay is to identify substances that cause DNA damage that result in the formation of mutation in the tissue that is being analysed.

5.	Mutations arising in a rodent are scored by recovering the transgene and analysing the phenotype of the reporter gene in a bacterial host deficient for the reporter gene. TGR gene mutation assays measure mutations induced in genetically neutral genes recovered from virtually any tissue of the rodent. These assays, 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, negative/positive selection against mutations).

6.	The weight of evidence suggests that transgenes respond to mutagens in a similar manner to endogenous genes, especially with regard to the detection of base pair substitutions, frameshift mutations, and small deletions and insertions (3).

7.	The International Workshops on Genotoxicity Testing (IWGT) have endorsed the inclusion of TGR gene mutation assays for *in vivo* detection of gene mutations, and have recommended a protocol for their implementation (6) (7). Further analysis supporting the use of this protocol can be found in (8). The present TG

is based on these recommendations for the evaluation of gene mutations in somatic tissues. In this revised TG, more recent recommendations for the evaluation of gene mutations in male germ cells have been endorsed (5).

8. The TGR gene mutation assay design is modelled on a repeat dose toxicity study (TG 407) permitting the option of combining the two assays into one study with an accommodation for performing the necropsy the day after the end of treatment for both studies. In addition, it is possible to integrate additional genotoxicity endpoints into the TGR study, such as assessment of micronuclei, Pig-a mutations or DNA strand breaks via the comet assay. Data are required to indicate that the performance of the repeat dose assay is not adversely affected by using a transgenic rodent strain rather than traditional rodent strains. For the integration of comet analysis in any tissues, consideration should be given to the effects on mutant frequency of continually dosing during the three-day sample period or administering the final dose(s) within the last 3-6 hrs of the three-day sampling period.

9. Definitions of key terms are set out in Annex 1.

INITIAL CONSIDERATIONS

10. TGR gene mutation assays for which sufficient data are available to support their use in this TG are: *lacZ* bacteriophage mouse (MutaMouse); *lacZ* plasmid mouse; *gpt* delta (*gpt* and Spi⁻) mouse and rat; *lacI* bacteriophage mouse and rat (Big Blue®), as performed under standard conditions. In these assays, the mutations are measured in bacterial genes inserted into a lambda vector (*lacI, lacZ, and gpt*). In addition, mutations can be measured in the *cII* gene of the bacteriophage in the Big Blue® and MutaMouse models, and the *red/gam* genes in the gpt delta model under Spi⁻ selection. Methods for the identification of mutants under selective conditions are available (see paragraph 16) and should be used preferentially. Mutagenesis in the TGR models is normally assessed as mutant frequency; if required, however, molecular analysis of the mutations can provide additional information (see Paragraphs 53-54).

11. These rodent *in vivo* gene mutation tests (3) are especially relevant to assessing mutagenic hazard in that the assays' responses are dependent upon *in vivo* metabolism, pharmacokinetics, DNA repair processes, and translesion DNA synthesis, although these may vary among species, among tissues and among the types of DNA damage. An *in vivo* assay for gene mutations is useful for further investigation of a mutagenic effect detected by an *in vitro* system, and for following up results of tests using other *in vivo* endpoints, such as positive tumour results from repeat dose carcinogenicity studies. In addition to being causally associated with the induction of cancer, gene mutation is a relevant endpoint for the prediction of mutation-based non-cancer diseases in somatic tissues (9) (10) as well as diseases transmitted through the germline.

12. Before use of the Test Guideline on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture.

PRINCIPLE OF THE TEST METHOD

13. In the assays described in paragraph 10, the target gene is bacterial or bacteriophage in origin, and the means of recovery from the rodent genomic DNA is by incorporation of the transgene into a λ bacteriophage or plasmid shuttle vector. The procedure involves the extraction of genomic DNA from the rodent tissue of interest,

in vitro processing of the genomic DNA (i.e. packaging of λ vectors, or ligation and electroporation of plasmids to recover the shuttle vector), and subsequent detection of mutations in bacterial hosts under suitable conditions. The assays employ neutral nontranscribed transgenes that are readily recoverable from most tissues.

14.   The basic TGR gene mutation experiment involves treatment of the rodent with a chemical over a period of time. Agents may be administered by any appropriate route, including implantation (*e.g.* medical device testing). The total period during which an animal is dosed is referred to as the administration period. Administration is usually followed by a period of time, prior to humane euthanasia, during which the agent is not administered and during which unrepaired DNA lesions are fixed into stable mutations. In the literature, this period has been variously referred to as the manifestation time, fixation time or expression time; the end of this period is the sampling time (6) (7). After the animal is humanely euthanized, tissues are rapidly frozen and held at or below -70°C until genomic DNA is isolated from the tissue(s) of interest and purified. Moribund animals euthanized during the last week of dosing may have tissue collected, frozen and analysis conducted on a case-by-case basis if needed.

15.   Data for a single tissue per animal from multiple packaging/ligations are usually aggregated, and mutant frequency is generally evaluated using a total of between 10^5 and 10^7 plaque-forming or colony- forming units with a minimum of 125,000 - 300,000 plaques per animal for those TGR models with background mutant frequency in the range of 3 x10^{-5} (Paragraph 52). Models such as *gpt* delta with lower background mutant frequencies require proportionally more colony-forming units to be observed to ensure adequate statistical power. When using positive selection methods, total plaque-forming units are determined with a separate set of non-selective plates.

16.   Positive selection methods have been developed to facilitate the detection of mutations in the *gpt* gene [*gpt* delta mouse and rat, *gpt*⁻ phenotype (11) (12) (13)] and the *lacZ* gene [MutaMouse or *lacZ* plasmid mouse (14) (15) (16) (17)]; whereas, no positive selection methods method are available for *lacI* gene mutations in Big Blue® animals which are detected through a non-selective method that identifies mutants through the generation of coloured (blue) plaques. Positive selection methodology is also in place to detect point mutations arising in the *cII* gene of the λ bacteriophage shuttle vector [Big Blue® mouse or rat (18) and MutaMouse (19)] and deletion mutations in the λ *red* and *gam* bacteriophage genes [Spi⁻ selection in *gpt* delta mouse and rat (13) (20) (21)]. Mutant frequency is calculated by dividing the number of plaques/plasmids containing mutations in the transgene by the total number of plaques/plasmids recovered from the same DNA sample. In TGR gene mutation studies, the mutant frequency is the reported parameter. In addition, a mutation frequency can be determined as the fraction of cells carrying independent mutations; this calculation requires correction for clonal expansion by sequencing the recovered mutants (Paragraphs 53-54).

17.   The mutations scored in the *lacI, lacZ, cII* and *gpt* point mutation assays consist primarily of base pair substitution mutations, frameshift mutations and small insertions/deletions. The relative proportion of these mutation types among spontaneous mutations is similar to that seen in the endogenous *Hprt* gene. Large deletions are detected only with the Spi⁻ selection in the *gpt* delta assay and the *lacZ* plasmid assays (3). Mutations of interest are *in vivo* mutations that arise in the mouse or rat. *In vitro* and *ex vivo* mutations, which may arise during phage/plasmid recovery, replication or repair, are relatively rare, and in some systems can be specifically identified, or excluded by the bacterial host/positive selection system.

DESCRIPTION OF THE METHOD

Preparations

Selection of animal species

18.   Transgenic mouse and rat gene mutation detection models are currently available. Both mouse and rat models are considered equally acceptable. Justification of the model used in the TGR assay should include a consideration of: 1) laboratory proficiency with the model; 2) availability of historical data in the tissues under investigation; as well as, 3) known toxicity differences between the species for the substance under investigation (e.g. when investigating the mechanism of carcinogenesis for a tumour seen only in one rodent species, to correlate with a toxicity study in a specific species, or if metabolism in one rodent species is known to be more representative of human metabolism).

Housing and feeding conditions

19.   The temperature in the experimental animal room ideally should be 22ºC (± 3ºC). Although the relative humidity should be at least 30% and preferably not exceed 70% other than during room cleaning, the goal should be to maintain a relative humidity of 50-60%. Lighting should be artificial, with a daily sequence of 12 hours light, followed by 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. The choice of diet may be influenced by the need to ensure a suitable admixture of a test chemical when administered by this route. Animals should be housed in small groups of the same sex if no aggressive behaviour is expected. Animals may be housed individually if scientifically justified.

Preparation of the animals

20.   Healthy young sexually mature adult animals (8-12 weeks old at start of treatment, although slightly younger or older animals are acceptable with scientific justification) are randomly assigned to the control and treatment groups. The animals are identified uniquely. The animals are acclimated to the laboratory conditions for at least five days. Cages should be arranged in such a way that possible effects due to cage placement are minimized. At the commencement of the study, the weight variation of animals should be minimal and not exceed ± 20% of the mean weight of each sex.

Preparation of doses

21.   Solid test chemicals should be dissolved or suspended in appropriate solvents or vehicles or admixed in diet or drinking water prior to dosing of the animals. Liquid test chemicals may be dosed directly or diluted prior to dosing. For inhalation exposures, test materials can be administered as gas, vapour, or a solid/liquid aerosol, depending on their physicochemical properties. Other routes of exposure should be scientifically justified. Fresh preparations of the test chemical should be employed unless stability data demonstrate the acceptability of storage.

Test Conditions

Solvent/vehicle

22.   The solvent/vehicle should not produce toxic effects at the dose volumes used, and should not be suspected of chemical reaction with the test chemical. It is recommended that wherever possible, the use of an aqueous solvent/vehicle should be considered first. Examples of commonly used compatible solvents/vehicles

include water, physiological saline, methylcellulose solution, carboxymethyl cellulose sodium salt solution, olive oil and corn oil (22). If other than well-known solvents/vehicles are used, their inclusion should be supported with reference data indicating their compatibility. In the absence of historical or published control data showing that no mutations and other deleterious effects are induced by a chosen atypical solvent/vehicle, an initial study should be conducted in order to establish the acceptability of the solvent/vehicle control or both a vehicle and concurrent untreated control groups added to the main TGR study.

Positive Controls

23.   Concurrent positive control animals should normally be used. This may be waived when the testing laboratory has demonstrated proficiency in the conduct of the test and has established a historical control range (see Paragraphs 27-31). When a concurrent positive control group is not used, DNA from previous positive control treated animals may be included with each study to confirm the success of the method. Such DNA from previous experiments should be obtained from the same species and tissues of interest, and properly stored (see Paragraph 49).

24.   When concurrent positive controls are used, it is not necessary to administer them by the same route or duration as the test chemical; however, the positive controls should be known to induce mutations in the tissues of interest for the test chemical. Positive control substances should reliably produce a detectable increase in mutant frequency over the spontaneous level. The doses of the positive control chemicals should be selected so as to produce weak or moderate effects that critically assess the performance and sensitivity of the assay. Examples of positive control substances and some of their target tissues are included in Table 1.

Table 1: Examples of positive control substances and some of their target tissues

Chemical and CAS No.	Characteristics	Mutation Target Tissue	
		Rat	**Mouse**
N-Ethyl-N-nitrosourea [CAS no. 759-73-9]	Direct acting mutagen	Liver, bone marrow, glandular stomach, duodenum, jejenum, nasal epithelium, lung, kidney, spleen, bladder	Liver, bone marrow, forestomach, glandular stomach, duodenum, colon, nasal epithelium, lung, liver, lung, spleen, kidney, ovarian granulosa cells, testicular germ cells, sperm
Ethyl carbamate (urethane) [CAS no. 51-79-6]	Mutagen, requires metabolism but produces only weak effects		Bone marrow, forestomach, small intestine, liver, lung, spleen
2,4-Diaminotoluene [CAS no. 95-80-7]	Mutagen, requires metabolism, also positive in the Spi- assay	Liver	Liver
Benzo[a]pyrene [CAS no. 50-32-8]	Mutagen, requires metabolism	Liver, bone marrow, glandular stomach, duodenum, jejenum,	Liver, bone marrow, breast, colon, forestomach, glandular stomach, duodenum, jejenum, heart, lung, kidney, bladder, testicular germ cells, sperm

		omenta, nasal epithelium, lung, kidney, spleen, bladder,	cells

Negative controls

25.   Negative controls, treated with solvent or vehicle alone, and otherwise treated in the same way as the treatment groups, should be included for every sampling time (however, see paragraph 22 regarding atypical solvents or vehicles).

VERIFICATION OF LABORATORY PROFICIENCY

Proficiency investigations

26.   In order to establish sufficient experience with the conduct of the assay prior to using it for routine testing, the laboratory should have demonstrated the ability to reproduce expected results from published data (3) for both mutant frequencies and transgene recovery from genomic DNA (e.g., packaging efficiency) with a minimum of two positive control substances (including weak responses induced by low doses of positive controls), such as those listed in Table 1 (see paragraph 24), and compatible vehicle/solvent controls (see paragraph 22). Proficiency should be demonstrated in at least two tissues, preferably one slow dividing tissue such as liver, and one rapidly dividing tissue such as bone marrow, glandular stomach or duodenum. These experiments should use doses that give reproducible increases and demonstrate the sensitivity and dynamic range of the test system in the tissue of interest. This requirement is not applicable to laboratories that have experience, i.e. that have a historical database available as defined in paragraphs 27-31.

Historical Control Data

27.   During the course of the proficiency investigations, the laboratory should establish:
     - A historical positive control range and distribution, and
     - A historical negative control range and distribution.

28.   When first acquiring data for a historical negative control distribution, concurrent negative controls should be consistent with published control data, where they exist. As more experimental data are added to the historical control distribution, concurrent negative controls should ideally be within the 95% control limits of that distribution (see below). The laboratory's historical negative control database should be compiled, analysed and regularly updated according to literature recommendations (e.g., 23). This should include consideration of the minimum number of data sets required to establish a robust range, frequency of update and methods to ensure the most recent and/or relevant data are used for assay acceptance and data evaluation (see paragraph 56). Significant deviations from these recommendations should be justified. Laboratories should use quality control methods, such as control charts (e.g. C-charts or X-bar charts (24)), to identify how variable their data are, and to show that the methodology is 'under control' in their laboratory.

29.   Where the laboratory does not complete a sufficient number of experiments to establish a statistically robust negative control distribution (see paragraph 28) during the proficiency investigations (described in

paragraph 26), it is acceptable that the distribution can be built during the first routine tests. This approach should follow the recommendations set out in the literature (23) and the negative control results obtained in these experiments should remain consistent with published negative control data.

30.   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 inconsistencies should result in the establishment of a new historical control database where expert judgement determines that it differs from the previous distribution (see paragraph 28) (23). During the reestablishment, a full negative control database may not be needed to permit the conduct of an actual test, provided that the laboratory can demonstrate that their concurrent negative control values remain either consistent with their previous database or with the corresponding published data.

31.   Negative control data should consist of the mutant frequency for each animal. Concurrent negative controls should ideally be within the 95% control limits of the distribution of the laboratory's historical negative control database. Where concurrent negative control data fall outside the 95% control limits, they may be acceptable for inclusion in the historical control distribution as long as these data are not extreme outliers and there is evidence that the test system is 'under control' (see paragraph 28) and no evidence of technical or human failure.

PROCEDURE

Number and Sex of Animals

32.   The number of animals per group should be predetermined to be sufficient to provide statistical power necessary to detect at least a doubling in mutant frequency. Group sizes will consist of a minimum of five animals; however, if the statistical power is insufficient, the number of animals should be increased as required. In general, the mutation response is similar between male and female animals and, therefore, most studies could be performed in either sex. Where human exposure to chemicals may be sex-specific, as for example with some pharmaceuticals, the test should be performed with the appropriate sex. Data demonstrating important differences between males and females (e.g. differences in systemic toxicity, metabolism, bioavailability etc. including e.g. in a range-finding study) would encourage the use of both sexes. It might be appropriate to use the factorial design in case both sexes are used. Details on how to analyse the data using this design are given in Annex 2. If a TGR study is performed to follow up positive tumour or other toxicology findings, the selection of species and sex will be based on the species and sex of the initial study. In such cases, matching the same strain of mouse or rat may not be possible.

Administration Period

33.   Based on observations that mutations accumulate with each treatment, a repeated-dose regimen is necessary, with daily treatments for a period of 28 days. This is generally considered acceptable both for producing a sufficient accumulation of mutations by weak mutagens, and for providing an exposure time adequate for detecting mutations in slowly proliferating organs. Alternative treatment regimens may be appropriate for some evaluations (e.g., for the evaluation of gene mutations in mature sperm, see paragraphs 42-46) or when testing with potent mutagens), and these alternative dosing schedules should be scientifically justified in the protocol. Treatments should not be shorter than the time required for the complete induction of

all the relevant metabolising enzymes, and shorter treatments may necessitate the use of multiple sampling times that are suitable for organs with different proliferation rates. In any case, all available information (*e.g.* on general toxicity or metabolism and pharmacokinetics) should be used when justifying a protocol, especially when deviating from the above standard recommendations. While it may increase sensitivity, treatment times longer than 8 weeks should be explained clearly and justified, since long treatment times may produce an apparent increase in mutant frequency through clonal expansion (7).

Dose Levels

34.   Dose levels should be based on the results of a dose range-finding study measuring general toxicity that was conducted by the same route of exposure, or on the results of pre-existing sub-acute toxicity studies. Non-transgenic animals of the same rodent strain may be used for determining dose ranges.

35.   In the main test, in order to obtain dose response information, a complete study should include a negative control group (see Paragraph 25) and a minimum of three, appropriately-spaced dose levels of the test substance, except where the limit dose has been used (see Paragraph 39). The highest dose should be the Maximum Tolerated Dose (MTD). The MTD is defined as the highest dose that will be tolerated without evidence of study-limiting toxicity, relative to the duration of the study period, for example, by inducing body weight depression but not death or evidence of pain, suffering or distress necessitating human euthanasia.

36.   Chemicals with specific biological activities at low non-toxic doses (such as hormones and mitogens), and substances that exhibit saturation of toxicokinetic properties, or induce detoxification processes that may lead to a decrease in exposure after long-term administration may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis. The dose levels used should cover a range from the maximum to little or no toxicity.

37.   For germ cell-specific studies, care should be taken during the dose range-finding study to assure that the MTD, as defined above in paragraph 35, does not result in excessive toxicity to the germ cell compartment such that not enough cells are recovered to measure mutations in the bacterial genes. Reduction in testis weight, evaluated either as absolute weight or relative weight to body weight, or relative to organs such as the brain (5) (25) (26), can be considered as a biomarker for transient or permanent male germ cell sterility and is therefore a dose-limiting clinical observation. There are presently insufficient data available to provide specific guidance on a threshold for excessive toxicity based on male gonadal tissue weight loss.

38.   For TGR studies primarily focused on somatic tissues with the option to analyse or retain germ cells for future analysis, dose selection should be made based on the standard MTD dose. However, additional lower doses may be added if data are available to suggest that the MTD may cause excessive testis toxicity.

Limit Test

39.   If dose range-finding experiments, or existing data from related rodent strains, indicate that a treatment regimen of at least the limit dose (see below) produces no observable toxic effects, or if genotoxicity would not be expected based upon data from structurally related substances, then a full study using three dose levels may not be considered necessary. Accordingly, for an administration period of 28 days (i.e. 28 daily treatments), this limit dose is 1000 mg/kg body weight/day.

Administration of Doses

40. The rationale for dose route should be clearly defined and should include a consideration of adequate exposure of the target tissue(s), as well as the anticipated route of human exposure. Routes of exposure that may be justified include dietary, drinking water, topical subcutaneous, intravenous, oral by gavage, inhalation, intratracheal, or implantation). Intraperitoneal injection is generally not recommended since it is not a physiologically relevant route of human exposure, and should only be used with detailed scientific justification. 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 1 mL/100 g of body weight for mice and rats, except in the case of aqueous solutions where a maximum of 2 mL/100g may be used. The use of volumes greater than this should be justified. Except for irritating or corrosive test chemicals, 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 in relation to body weight at all dose levels.

Sampling Time

Somatic Cells

41. The sampling time is a critical variable because it is determined by the period needed for mutations to be fixed. This period is tissue-specific and appears to be related to the turnover time of the cell population (27), with bone marrow and intestine being rapid responders and the liver being much slower. A suitable compromise for the measurement of mutant frequencies in both rapidly and slowly proliferating tissues is 28 consecutive daily treatments (as indicated in Paragraph 33) and sampling three days after the final treatment (i.e., 28+3d); although the maximum mutant frequency may not manifest itself in slowly proliferating tissues under these conditions. If slowly proliferating somatic tissues are of particular importance, then a later sampling time of 28 days following the 28-day administration period (i.e., 28+28d) may be more appropriate (7) (8). In such cases, the later sampling time would replace the 3-day sampling time, and would require scientific justification.

Germ Cells

42. TGR assays are well-suited for the study of gene mutation induction in male germ cells (5) (28) (29) (30), in which the timing and kinetics of spermatogenesis have been well-defined (26). Male germ cells can be collected as either mature sperm (Paragraph 43) or as developing germ cells (Paragraph 44). The low numbers of ova available for analysis, even after super-ovulation, and the fact that there is no DNA synthesis in the oocyte in the adult, preclude the determination of mutation in female germ cells using transgenic assays (31).

43. Mature sperm can be obtained from the cauda epididymis and vas deferens. Collection of sperm from the cauda epididymis is normally sufficient to provide a large enough sample for conducting the gene mutation assay, and therefore, collection of sperm from the vas deferens may not be necessary. The time for the progression of developing germ cells from spermatogonial stem cells to mature sperm reaching the vas deferens/cauda epididymis is ~49 days for the mouse (26) (32) and ~70 days for the rat (26) (33) (34). Thus, in order to sample sperm in the vas deferens/cauda epididymis that developed from spermatogonial stem cells that were exposed to the test substance, a sampling time at a minimum of 49 days (mice), or 70 days (rat), after the end of treatment is required. Sampling of mature sperm from the cauda epididymis at 28+3d does not generate meaningful mutagenicity data and the TGR assay in sperm at this timepoint should not be conducted (5) (35).

44. Cells extruded from seminiferous tubules in the testis after a 28+3d regimen comprise a mixed population

encompassing all stages of developing germ cells (5) (26) (28) (29). Analysis of these cells for gene mutation detection after 28+3d provides some coverage of germ cells exposed across the majority of phases of germ cell development; however, a mutagenic effect may not be detected due to the fact that not all collected germ cells will have received continuous exposure for the full 28 days during the proliferative phase of spermatogenesis (5) (26). Consequently positive results in tubule germ cells after a 28+3d regimen are informative, while negative results after a 28+3d sampling regimen are insufficient to negate the possibility that the chemical is a germ cell mutagen.

45.   Although, as stated above, the time for the progression of developing germ cells from spermatogonial stem cells to mature sperm reaching the cauda epididymis/vas deferens is longer than 28 days in both mice and rats, collection of developing germ cells from seminiferous tubules after a 28+28d sampling regimen is considered appropriate for the assessment of germ cell mutagenicity for both mice and rats. This regimen enables the evaluation of mutations in a cell population that has received most of the 28-day exposure during the proliferative phase of spermatogenesis and has been shown to be acceptable by the analysis of the available data (5) coupled with results from modelling of spermatogenesis (26). Collection of sperm from the cauda epididymis after a 28+28d sampling regimen is, however, unnecessary as it would not provide any further information to that obtained by sampling germ cells from the seminiferous tubules.

46.   As discussed in paragraph 44, a 28+3d sampling regimen would not be appropriate for germ cells (5). However, a 28+28d sampling regimen may be a practical approach to simultaneously assess mutagenicity in germ cells as well as somatic tissues, especially for slowly proliferative tissues (see paragraph 41).

Observations

47.   General clinical observations should be made at least once a day, preferably at the same time(s) each day and considering the peak period of anticipated effects after dosing. The health condition of the animals should be recorded. At least twice daily, all animals should be observed for morbidity and mortality. All animals should be weighed at least once a week, and at sacrifice. Measurements of food consumption should be made at least weekly. If the test chemical is administered via the drinking water, water consumption should be measured at each change of water and at least weekly. Animals exhibiting non-lethal indicators of excess toxicity should be euthanatized prior to completion of the test period (36).

Tissue Collection

48.   The rationale for tissue collection should be defined clearly. Since it is possible to study mutation induction in virtually any tissue, the selection of tissues to be collected should be based upon the reason for conducting the study and any existing genotoxicity, carcinogenicity or toxicity data for the test substance under investigation. Important factors for consideration should include the route of administration (based on likely human exposure route(s)), the predicted tissue distribution, and the possible mechanism of action. In the absence of any background information, several somatic tissues as may be of interest should be collected which should represent rapidly proliferating, slowly proliferating, site of contact tissues and any known target tissues. In addition, developing germ cells from the seminiferous tubules may be collected and stored in case future analysis of germ cell mutagenicity is required. Relevant organ weights should be obtained, and for larger organs, the same area should be collected from all animals.

Storage of Tissues and DNA

49.   Tissues should be quickly frozen and stored at or below −70 ºC and be used for DNA isolation within 5 years. Isolated DNA, stored refrigerated at 4 ºC in appropriate buffer, should be used optimally for mutation analysis within 1 year.

Selection of Tissues for Mutant Analysis

50.   The choice of tissues should be based on considerations such as: 1) the route of administration or site of first contact (*e.g.* glandular stomach or duodenum if administration is oral, lung or nasal epithelium if exposure is through inhalation, or skin if topical application has been used); and 2) ADME parameters observed in general toxicity studies, which indicate tissue distribution, retention or accumulation, or target organs for toxicity. If studies are conducted to follow up carcinogenicity studies, target tissues for carcinogenicity should be investigated. The choice of tissues for analysis should maximize the detection of chemicals that are direct-acting *in vitro* mutagens, rapidly metabolized, highly reactive or poorly absorbed, or those for which the target tissue is determined by route of administration (37).

51.   In the absence of background information and taking into consideration the site of contact due to route of administration, the liver and at least one rapidly dividing tissue (*e.g.* glandular stomach or duodenum, or bone marrow) should be evaluated for mutagenicity. In most cases, the above requirements can be achieved from analyses of two carefully selected tissues, but in some cases, three or more would be needed. If there are reasons to be specifically concerned about germ cell effects, including positive responses in somatic cells, germ cell tissues should be evaluated for mutations as discussed in paragraphs 42-45 above.

Methods of Measurement

52.   Standard laboratory or published methods for the detection of mutants are available for the recommended transgenic models: *lacZ* lambda bacteriophage and plasmid (17); *lacI* mouse (38) (39); *gpt* delta mouse (21); *gpt* delta rat (13); *cII* (18). Modifications should be justified and properly documented. Data from multiple packagings can be aggregated and used to reach an adequate number of plaques or colonies. However, the need for a large number of packaging reactions to reach the appropriate number of plaques may be an indication of poor DNA quality. In such cases, data should be considered cautiously because they may be unreliable. The optimal total number of plaques or colonies per DNA sample is governed by the statistical probability of detecting sufficient numbers of mutants at a given spontaneous mutant frequency. In general, a minimum of 125,000 to 300,000 plaques is required if the spontaneous mutant frequency is in the order of 3×10^{-5} (15; see also Paragraph 15). Tissues and the resulting samples (items) should be processed and analysed using a block design, where equal numbers of items from the vehicle/solvent control group, the positive control group (if used) or positive control DNA (where appropriate), and each treatment group are processed together.

Sequencing of mutants

53.   Clonal amplification of early spontaneously arising mutants may occur in any animal, leading to small to large increases in mutant frequency in individual tissues. Tissues from animals with elevated mutant frequencies outside the historic range and different from other animals in the group may represent such a jackpot or clonal event. As such events are often localized within a tissue; reanalysis of a different portion of the tissue may be one approach. Often extra replacement animals are included in studies to accommodate animals lost due to early death or presence of jackpot animals. Analysis of the extra animals may be appropriate in this case.

54.   While for regulatory applications, DNA sequencing of mutants is not routinely required, particularly where a clear positive or negative result is obtained, sequencing data may be useful when high inter- individual variation is observed. In these cases, sequencing can be used to rule out the possibility of jackpots or clonal events by identifying the proportion of unique mutants from a particular tissue. Sequencing approximately 10 mutants per tissue per animal should be sufficient for simply determining if clonal mutants contribute to the mutant frequency; sequencing as many as 25 mutants may be necessary to correct mutant frequency mathematically for clonality. Sequencing of mutants also may be considered when small increases in mutant frequency (*i.e.* just exceeding the untreated control values) are found. Differences in the mutant spectrum between the mutant colonies from treated and untreated animals may lend support to a mutagenic effect (7). Also, mutation spectra may be useful for developing mechanistic hypotheses. When sequencing is to be included as part of the study protocol, special care should be taken in the design of such studies, in particular with respect to the number of mutants sequenced per sample, to achieve adequate power according to the statistical model used (see Paragraph 58). In this regard, Next Generation Sequencing methods are available *for* both cII (40) and *lacZ* (41) genes, which greatly facilitates the sequencing of large numbers of mutants.

DATA AND REPORTING

Treatment of Results

55.   Individual animal data should be presented in tabular form. The experimental unit is the animal. The report should include the total number of plaque-forming units (pfu) or colony-forming units (cfu), the number of mutants, and the mutant frequency for each tissue from each animal. The report should also include the number packaging/rescue reactions and the number of reactions per DNA sample. While data for each individual reaction should be retained, only the total number of mutants and pfu/cfu need be reported. Data on toxicity and clinical signs as per paragraph 47 should be reported. Any sequencing results should be presented for each mutant analyzed, and resulting mutation frequency calculations for each animal and tissue should be shown.

Statistical Evaluation and Interpretation of Results

Acceptability criteria

56.   The following criteria determine the acceptability of the test:
   a) The concurrent negative control data are considered acceptable for addition to the laboratory historical control database (see paragraphs 27-31).
   b) The concurrent positive controls or scoring controls should induce responses that are compatible with those generated in the historical positive control database and produce a statistically significant increase compared with the concurrent negative control (see paragraphs 24 and 25).
   c) The appropriate number of doses, animals per dose, and plaques have been analysed.
   d) The criteria for the selection of highest dose are consistent with those described in paragraphs 34-37.

Evaluation and Interpretation of Results

57.   Providing that all acceptability criteria above are fulfilled, a test chemical is considered clearly positive if all of the following criteria are met:

a) At least one of the treatment groups exhibits a statistically significant increase in the mutant frequency compared with the concurrent negative control;

b) This increase is dose-related when evaluated with an appropriate trend test (not applicable to the limit test);

c) The result is outside the distribution of the historical negative control data (e.g., Poisson-based 95% control limits).

Positive results in the transgenic rodent mutation test indicate that the test chemical induced gene mutations in the analysed tissue.

58. Statistical tests used should consider the animal as the experimental unit. Appropriate statistical methods can be found in the following references (6) (42) (43) (44) (45). At least three doses plus control should be analysed in order to provide sufficient data for dose-response analysis.

59. Providing that all acceptability criteria are fulfilled, a test chemical is considered clearly negative if all of the following criteria are met in all experimental conditions examined:

a) none of the treatment groups exhibits a statistically significant increase in the mutant frequency compared with the concurrent negative control;

b) if multiple doses were used, there is no dose-related increase, when evaluated by an appropriate trend test;

c) all results are inside the distribution of the historical negative control data (e.g. Poisson-based 95% control limits);

d) tissue exposure to the test substance(s) occurred.

Negative results indicate that, under the test conditions, the test chemical does not induce gene mutations in the tested tissue.

60. Evidence of exposure of the tested tissue to a test substance may be gained from morphological or histopatological data obtained from determinations in the same study or comparable toxicity studies. Alternatively, ADME data, obtained in an independent study using the same route and same species can be used to demonstrate tissue exposure. For germ cell studies, a significant reduction in testis weight can be taken as evidence of exposure (see paragraph 37).

61. There is no requirement for verification of a clear positive or clear negative response.

62. In cases where the response is not clearly negative or positive, or in the case of a positive result at the only dose given in a limit test, and in order to assist in establishing the biological relevance of a result (e.g. a weak or borderline increase), the data should be evaluated by expert judgement and/or further investigations of the existing experiments completed. In some cases, analyzing more plaques, analyzing more animals or performing a repeat experiment using modified experimental conditions could be useful.

63. Sequencing of mutant plaques to determine whether there is a shift in the mutant spectrum induced by the test agent may aid in concluding whether the response is negative or positive. For DNA sequencing analyses, a number of statistical approaches are available to assist in interpreting the results (46) (47) (48) (49).

64. In rare cases, even after further investigations, the data will preclude making a conclusion that the test chemical produces either positive or negative results, and the study will therefore be concluded as equivocal.

Test report

65.   The test report should include the following information:
Test chemical:
-    identification data and CAS n°, if known;
-    source, lot number if available;
-    physical nature and purity;
-    physiochemical properties relevant to the conduct of the study;
-    stability of the testchemical, if known;

Solvent/vehicle:
-    justification for choice of vehicle;
-    solubility and stability of the test chemical in the solvent/vehicle, if known;
-    preparation of dose formulations including dietary, drinking water or inhalation formulations;
-    analytical determinations on formulations (e.g., stability, homogeneity – for non-soluble substances, nominal concentrations);

Test animals:
-    species/strain used and justification for the choice;
-    number, age and sex of animals;
-    source, housing conditions, diet, etc.;
-    individual weight of the animals at the start of the test, including body weight range, mean and standard deviation for each group;

Test conditions:
-    positive and negative (vehicle/solvent) control data;
-    data from the range-finding study;
-    rationale for dose level selection;
-    details of test chemical preparation;
-    details of the administration of the test chemical;
-    rationale for route of administration;
-    methods for measurement of animal toxicity, including, where available, histopathological or haematological analyses and the frequency with which animal observations and body weights were taken;
-    methods for verifying that the test chemical reached the target tissue, or general circulation, if negative results are obtained;
-    actual dose (mg/kg body weight/day) for most dose routes or for diet/drinking water test exposure either parts per million (ppm) or actual dose based onchemical concentration (ppm) and average food or water consumption, if applicable;
-    details of food and water quality, husbandry and environmental enrichment (if any);
-    detailed description of treatment and sampling schedules and justifications for the choices;
-    method of euthanasia;
-    procedures for isolating and preserving tissues;
-    methods for isolation of rodent genomic DNA, rescuing the transgene from genomic DNA, and transferring transgenic DNA to a bacterial host;
-    source and lot numbers of all cells, kits and reagents (where applicable);
-    methods for enumeration of mutants;
-    methods for molecular analysis of mutants and use in correcting for clonality and/or calculating

mutation frequencies, if applicable;

Results:
-       animal condition prior to and throughout the test period, including signs of toxicity;
-       body weights and body weight changes throughout the test period;
-       food and water consumption throughout the test period, if applicable for dosed food or drinking water studies;
-       body and organ weights at sacrifice;
-       for each tissue/animal, the number of mutants, number of plaques or colonies evaluated, number of packagings, mutant frequency;
-       for each tissue/animal group, total number of mutants, mean mutant frequency, standard deviation;
-       dose-response relationship, where possible;
-       for each tissue/animal, the number of independent mutants and mean mutation frequency, where molecular analysis of mutations was performed;
-       concurrent and historical negative control data with ranges, means, standard deviations, range and control limits;
-       concurrent and historical positive control data;
-       analytical determinations, if available (e.g. DNA concentrations used in packaging, DNA sequencing data);
-       statistical analyses and methods applied;

Discussion of the results;

Conclusion.

LITERATURE

(1)   OECD (2016), Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. ENV Publications, Series on Testing and Assessment, No. 234, OCED, Paris.

(2)   Thybaud, V, E. Lorge, D.D. Levy, J. van Benthem, G.R. Douglas, F. Marchetti, M.M. Moore and R. Schoeny (2017), "Main issues addressed in the 2014-2015 revisions to the OECD genetic toxicology test guidelines", *Environ. Mol. Mutagen.*, 58: 284-295.

(3)   OECD (2009), *Detailed Review Paper on Transgenic Rodent Mutation Assays*, Series on Testing and Assessment, N° 103, ENV/JM/MONO(2009)7, OECD, Paris.

(4)   OECD (2011), *Retrospective Performance Assessment of OECD Test Guideline on Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays*, Series on Testing and Assessment, N° 145, OECD, Paris.

(5)   Marchetti, F., M. Aardema, C. Beevers, J. van Benthem, R. Godschalk, C.L. Yauk, B. Young, A. Williams and G.R. Douglas (2018), "Identifying germ cell mutagens using OECD test guideline 488 (transgenic rodent somatic and germ cell mutation assay) and integration with somatic cell testing", Mutation Res., 832-833: 7-18.

(6)   Heddle, J.A*.,* S. Dean, T. Nohmi, M. Boerrigter, D. Casciano, G.R. Douglas, B.W. Glickman, N.J. Gorelick, J.C. Mirsalis, H.-J Martus, T.R. Skopek, V. Thybaud, K.R.Tindall and N. Yajima (2000), "In vivo Transgenic Mutation Assays", *Environ. Mol. Mutagen.*, 35: 253-259.

(7)   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.

(8)   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.

(9)   Erikson, R.P. (2003), "Somatic Gene Mutation and Human Disease other than Cancer", *Mutation Res.*, 543: 125-136.

(10)   Erikson, R.P. (2010), "Somatic Gen Mutation and Human Disease other than Cancer: an Update". Mutation Res., 705: 96-106.

(11)   Nohmi, T., M. Katoh, H. Suzuki, M. Matsui, M. Yamada, M. Watanabe, M. Suzuki, N. Horiya, O. Ueda, T. Shibuya, H. Ikeda and T. Sofuni (1996), "A new Transgenic Mouse Mutagenesis Test System using Spi⁻ and 6-thioguanine Selections", *Environ. Mol. Mutagen.*, 28(4): 465–470.

(12)   Nohmi, T., T. Suzuki and K.I. Masumura (2000), "Recent Advances in the Protocols of Transgenic Mouse Mutation Assays", *Mutation Res.*, 455(1–2): 191–215.

(13)   Toyoda-Hokaiwado, N., T. Inoue, K. Masumura, H. Hayashi, Y. Kawamura, Y. Kurata, M. Takamune, M. Yamada, H. Sanada, T. Umemura, A. Nishikawa and T. Nohmi (2010), "Integration of *in vivo* Genotoxicity and Short-term Carcinogenicity Assays using F344 *gpt* delta Transgenic Rats: *in vivo* Mutagenicity of 2,4-diaminotoluene and 2,6-diaminotoluene Structural Isomers"*, Toxicol. Sci.*, 114(1): 71-78.

(14)   Boerrigter, M.E., M.E. Dollé, H.-J. Martus, J.A. Gossen and J. Vijg (1995), "Plasmid-based Transgenic Mouse Model for Studying *in vivo* Mutations" *Nature*, 377(6550): 657–659.

(15)   Gossen, J.A., W.J. de Leeuw, C.H. Tan, E.C. Zwarthoff, F. Berends, P.H. Lohman, D.L. Knook and J. Vijg

(1989), "Efficient Rescue of Integrated Shuttle Vectors from Transgenic Mice: a Model for Studying Mutations *in vivo*", *Proc. Natl. Acad. Sci. USA,* 86(20): 7971–7975.

(16)   Gossen, J.A. and J. Vijg (1993), "A Selective System for lacZ-Phage using a Galactose-sensitive *E. coli* Host", *Biotechniques*, 14(3): 326, 330.

(17)   Vijg, J. and G.R. Douglas (1996), "Bacteriophage λ and Plasmid *lacZ* Transgenic Mice for studying Mutations *in vivo*" in: G. Pfeifer (ed.), *Technologies for Detection of DNA Damage and Mutations, Part II,* Plenum Press, New York, NY, USA, pp. 391–410.

(18)   Jakubczak, J.L., G. Merlino, J.E. French, W.J. Muller, B. Paul, S. Adhya and S. Garges (1996), "Analysis of Genetic Instability during Mammary Tumor Progression using a novel Selection-based Assay for *in vivo* Mutations in a Bacteriophage λ Transgene Target", *Proc. Natl. Acad. Sci. USA*, 93(17): 9073–9078.

(19)   Swiger, R.R., L. Cosentino, N, Shima, J.H. Bielas, Cruz-Munoz W. and J.A. Heddle (1999), "The cII locus in the MutaMouse system", *Environ. Mol. Mutagen*., 34(2-3): 201-207.

(20)   Nohmi, T., M. Suzuki, K. Masumura, M. Yamada, K. Matsui, O. Ueda, H. Suzuki, M. Katoh, H. Ikeda and T. Sofuni (1999), "Spi⁻ Selection: an Efficient Method to Detect γ-ray-induced Deletions in Transgenic Mice", *Environ. Mol. Mutagen*., 34(1): 9–15.

(21)   Nohmi, T., T. Suzuki and K.I. Masumura (2000), "Recent Advances in the Protocols of Transgenic Mouse Mutation Assays", *Mutation Res.*, 455(1–2): 191–215.

(22)   Gad, S.C., C.D. Cassidy, N. Aubert, Spainhour B. and H. Robbe (2006) "Nonclinical vehicle use in studies by multiple routes in multiple species", Int. J. Toxicol., 25(6): 499-521.

(23)   Hayashi, M., K. Dearfield, P. Kasper, D. Lovell, H.-J. Martus, V. Thybaud (2011), "Compilation and Use of Genetic Toxicity Historical Control Data", *Mutation Res.,* 732(2): 87-90.

(24)   Ryan, T.P. (2000), "Statistical Methods for Quality Improvement", 2nd ed., John Wiley and Sons, New York.

(25)   EPA, (1996) "Guidelines for Reproductive Toxicity Risk Assessment", Federal Register, 61: 56274-56322.

(26)   Marchetti, F., M. Aardema, C. Beevers, J. van Benthem, G.R. Douglas, R. Godschalk, C.L. Yauk, B. Young and A. Williams (2018), "Simulation of mouse and rat spermatogenesis to inform genotoxicity testing using OECD test guideline 488", Mutation Res., 832-833: 19-28.

(27)   White, P.A., G.R. Douglas, D.H. Phillips and V.M. Arlt (2017) "Quantitative relationship between *lacZ* mutant frequency and DNA adduct frequency in Muta™Mouse tissues and cultured cells exposed to 3-nitrobenzanthrone. Mutagenesis, 32(2): 299-312.

(28)   Douglas, G.R., J. Jiao, J.D. Gingerich, J.A. Gossen and L.M. Soper (1995), "Temporal and Molecular Characteristics of Mutations Induced by Ethylnitrosourea in Germ Cells Isolated from Seminiferous Tubules and in Spermatozoa of *lacZ* Transgenic Mice", *Proc. Natl. Acad. Sci. USA*, 92: 7485-7489.

(29)   Douglas, G.R*.,* J.D. Gingerich, L.M. Soper and J. Jiao (1997), "Toward an Understanding of the Use of Transgenic Mice for the Detection of Gene Mutations in Germ Cells", *Mutation Res.*, 388(2-3): 197-212.

(30)   Singer, T.M., I.B. Lambert, A. Williams, G.R. Douglas and C.L. Yauk (2006), "Detection of Induced Male Germline Mutation: Correlations and Comparisons between Traditional Germline Mutation Assays, Transgenic Rodent Assays and Expanded Simple Tandem Repeat Instability Assays", *Mutation. Res*., 598: 164-193.

(31) Yauk, C.L., J.D. Gingerich, L. Soper, A. MacMahon, W.G. Foster and G.R. Douglas (2005), "A *lacZ* Transgenic Mouse Assay for the Detection of Mutations in Follicular Granulosa Cells", *Mutation Res*., 578(1-2): 117-123.

(32) Russell, L.B. (2004), "Effects of male germ-cell stage on the frequency, nature, and spectrum of induced specific-locus mutations in the mouse", Genetica, 122: 25–36.

(33) Clermont, Y. (1972), "Kinetics of spermatogenesis in mammals seminiferous epithelium cycle and spermatogonial renewal". *Physiol. Rev. 52:* 198-236.

(34) Robaire, B., Hinton, B.T., and Oregbin-Crist, M.-C. (2006), "The Epididymis", in Neil, J.D., Pfaff, D.W., Chalis, J.R.G., de Kretser, D.M., Richards, J.S., and P. M, Wassarman (eds.), *Physiology of Reproduction*, Elsevier, the Netherlands, pp. *1071-1148*.

(35) O'Brien J.M., Beal M.A., Yauk, C.L. and F. Marchetti (2016) "Benzo(a)pyrene is mutagenic in mouse spermatogonial stem cells and dividing spermatogonia. *Toxicol. Sci*., 152: 363-371.

(36) 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.

(37) Dean, S.W., T.M. Brooks, B. Burlinson, J. Mirsalis, B. Myhr, L. Recio and V. Thybaud (1999), "Transgenic Mouse Mutation Assay Systems can Play an important Role in Regulatory Mutagenicity Testing *in vivo* for the Detection of Site-of-contact Mutagens", *Mutagenesis*, 14(1): 141–151.

(38) Bielas, J.H. (2002), "A more Efficient Big Blue® Protocol Improves Transgene Rescue and Accuracy in an Adduct and Mutation Measurement", *Mutation Res.*, 518: 107–112.

(39) Kohler, S.W., G.S. Provost, P.L. Kretz, A. Fieck, J.A. Sorge and J.M. Short (1990), "The Use of Transgenic Mice for Short-term, *in vivo* Mutagenicity Testing", *Genet. Anal. Tech. Appl.*, 7(8): 212– 218.

(40) Besaratinia, A., H. Li, J. Yoon, A. Zheng, H. Gao and S. Tommasi (2012) "A high-throughput next-generation sequencing-based method for detecting the mutational fingerprint of carcinogens", *Nucleic Acids Res*., 40(15):e116–6.

(41) Beal, M.A., R. Gagne, A. Williams, Marchetti F. and C.L. Yauk (2015) " Characterizing benzo(a)pyrene-induced lacZ mutation spectrum in transgenic mice using Next Generation Sequencing, *BMC Genomics*, 16: 812.

(42) Carr, G.J. and N.J. Gorelick (1995), "Statistical Design and Analysis of Mutation Studies in Transgenic Mice*", Environ. Mol. Mutagen*, 25(3): 246–255.

(43) Fung, K.Y., G.R. Douglas and D. Krewski (1998), "Statistical Analysis of *lacZ* Mutant Frequency Data from Muta™Mouse Mutagenicity Assays", *Mutagenesis*, 13(3): 249–255.

(44) Piegorsch, W.W., B.H. Margolin, M.D. Shelby, A. Johnson, J.E. French, R.W. Tennant and K.R. Tindall (1995), "Study Design and Sample Sizes for a *lacI* Transgenic Mouse Mutation Assay", *Environ. Mol. Mutagen*., 25(3): 231–245.

(45) Piegorsch, W.W., A.C. Lockhart, G.J. Carr, B.H. Margolin, T. Brooks, G.R. Douglas, U.M. Liegibel, T. Suzuki, V. Thybaud, J.H. van Delft and N.J. Gorelick (1997), "Sources of Variability in Data from a Positive Selection *lacZ* Transgenic Mouse Mutation Assay: an Interlaboratory Study", *Mutation. Res*., 388(2–3): 249–289.

(46) Adams, W.T. and T.R. Skopek (1987), "Statistical Test for the Comparison of Samples from Mutational

Spectra", *J. Mol. Biol*., 194: 391-396.

(47) Carr, G.J. and N.J. Gorelick (1996), "Mutational Spectra in Transgenic Animal Research: Data Analysis and Study Design Based upon the Mutant or Mutation Frequency", *Environ. Mol. Mutagen,* 28: 405–413.

(48) Dunson, D.B. and K.R. Tindall (2000), "Bayesian Analysis of Mutational Spectra", *Genetics*, 156: 1411–1418.

(49) Lewis P.D., B. Manshian, M.N. Routledge, G.B. Scott and P.A. Burns (2008), "Comparison of Induced and Cancer-associated Mutational Spectra using Multivariate Data Analysis", *Carcinogenesis,* 29(4): 772-778.

ANNEX 1

DEFINITIONS

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

Base pair substitution: a type of mutation that causes the replacement of a single DNA nucleotide base with another DNA nucleotide base.

Capsid: the protein shell that surrounds a virus particle.

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

C*olony-forming unit* (cfu): a measure of viable bacterial numbers.

Concatamer: a long continuous biomolecule composed of multiple identical copies linked in series.

Cos site: a 12-nucleotide segment of single-stranded DNA that exists at both ends of the bacteriophage lambda's double-stranded genome.

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

Electroporation: the application of electric pulses to increase the permeability of cell membranes.

Endogenous gene: a gene native to the genome.

Frameshift mutation: a genetic mutation caused by insertions or deletions of a number of nucleotides that is not evenly divisible by three within a DNA sequence that codes for a protein/peptide.

Insertion: the addition of one or more nucleotide base pairs into a DNA sequence.

Jackpot: a large number of mutants that arose through clonal expansion from a single mutation.

Large deletions: deletions in DNA of more than several kilobases (which are effectively detected with the Spi⁻ selection and the lacZ plasmid assays).

Ligation: the covalent linking of two ends of DNA molecules using DNA ligase.

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

Neutral gene: a gene that is not affected by positive or negative selective pressures.

Packaging: the assembly of infective phage particles from a preparation of phage *capsid* and tail proteins and concatenated phage DNA molecules cut from genomic DNA that is cleaved into single copies of the phage genome. Commonly used to package DNA cloned onto a lambda vector (separated by *cos sites*) into infectious lambda particles.

Packaging efficiency: the efficiency with which packaged bacteriophages are recovered in host bacteria.

Plaque forming unit (pfu): a measure of viable bacteriophage numbers.

Point mutation: a general term for a mutation affecting only a small sequence of DNA including small insertions, deletions, and base pair substitutions.

Positive selection: a method that permits only mutants to survive.

Reporter gene: a gene whose mutant gene product is easily detected.

Sampling time: the end of the period of time, prior to sacrifice, during which the agent is not administered and during which unprocessed DNA lesions are fixed into stable mutations.

Shuttle vector: a vector constructed so that it can propagate in two different host species; accordingly, DNA inserted into a shuttle vector can be tested or manipulated in two different cell types or two different organisms.

Transgenic: of, relating to, or being an organism whose genome has been altered by the transfer of a gene or genes from another species.

ANNEX 2

THE FACTORIAL DESIGN FOR IDENTIFYING SEX DIFFERENCES IN THE IN VIVO MICRONUCLEUS ASSAY

The factorial design and its analysis

In this design, a minimum of 5 males and 5 females are tested at each concentration level resulting in a design using a minimum of 40 animals (20 males and 20 females, plus relevant positive controls).

The design, which is one of the simpler factorial designs, is equivalent to a two-way analysis of variance with sex and concentration level as the main effects. The data can be analysed using many standard statistical software packages such as SPSS, SAS, STATA, Genstat as well as using R.

The analysis partitions the variability in the dataset into that between the sexes, that between the concentrations and that related to the interaction between the sexes and the concentrations. Each of the terms is tested against an estimate of the variability between the replicate animals within the groups of animals of the same sex given the same concentration. Full details of the underlying methodology are available in many standard statistical textbooks (see references) and in the 'help' facilities provided with statistical packages.

The analysis proceeds by inspecting the sex x concentration interaction term in the ANOVA table[1]. In the absence of a significant interaction term the combined values across sexes or across concentration levels provide valid statistical tests between the levels based upon the pooled within group variability term of the ANOVA.

The analysis continues by partitioning the estimate of the between concentrations variability into contrasts which provide for a test for linear and quadratic contrasts of the responses across the concentration levels. When there is a significant sex x concentration interaction this term can also be partitioned into linear x sex and quadratic x sex interaction contrasts. These terms provide tests of whether the concentration responses are parallel for the two sexes or whether there is a differential response between the two sexes.

The estimate of the pooled within group variability can be used to provide pair-wise tests of the difference between means. These comparisons could be made between the means for the two sexes and between the means for the different concentration level such as for comparisons with the negative control levels. In those cases where there is a significant interaction comparisons can be made between the means of different concentrations within a sex or between the means of the sexes at the same concentration.

References

There are many statistical textbooks which discuss the theory, design, methodology, analysis and interpretation of factorial designs ranging from the simplest two factor analyses to the more complex forms used in Design of

---

[1] Statisticians who take a modelling approach such as using General Linear Models (GLMs) may approach the analysis in a different but comparable way but will not necessarily derive the traditional anova table, which dates back to algorithmic approaches to calculating the statistics developed in a pre-computer age.

Experiment methodology. The following is a non-exhaustive list. Some books provide worked examples of comparable designs, in some cases with code for running the analyses using various software packages.

Box, G.E.P, Hunter, W.G. and Hunter, J.S. (1978). Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building. New York: John Wiley & Sons.

Box G.E.P. & Draper, N.R. (1987) Empirical model-building and response surfaces. John Wiley & Sons Inc.

Doncaster, C.P. & Davey, A.J.H. (2007) Analysis of Variance and Covariance: How to choose and Construct Models for the Life Sciences. Cambridge University Press.

Mead, R. (1990) The Design of Experiments. Statistical principles for practical application. Cambridge University Press.

Montgomery D.C. (1997) Design and Analysis of Experiments. John Wiley & Sons Inc.

Winer, B.J. (1971) Statistical Principles in Experimental Design. McGraw Hill.

Wu, C.F.J & Hamada, M.S. (2009) Experiments: Planning, Analysis and Optimization. John Wiley & Sons Inc.