Section 4
Health effects

Test Guideline No. 470

Mammalian Erythrocyte Pig-a Gene Mutation Assay

30 June 2022

OECD Guidelines for the Testing of Chemicals

OECD
BETTER POLICIES FOR BETTER LIVES

Adopted:
30 June 2022

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

Mammalian Erythrocyte Pig-a Gene Mutation Assay

INTRODUCTION

1. The *in vivo* erythrocyte *Pig-a* gene mutation assay (hereafter called the *Pig-a* assay) uses an endogenous mammalian gene, the phosphatidylinositol glycan class A gene (*Pig-a*), as a reporter of somatic-cell gene mutation. *In vivo* gene mutation tests, such as the *Pig-a* assay, are especially relevant for assessing mutagenicity because physiological factors, such as absorption of the test chemical from the site of exposure, distribution of the test chemical throughout the test system *via* systemic circulation, and *in vivo* metabolism and DNA repair processes, all contribute to the mutagenic responses. For these reasons, the *in vivo Pig-a* assay can be used to further investigate genotoxic potential detected by *in vitro* systems.

2. The *Pig-a* assay can be performed with commonly used strains of rats or mice (1) (see paragraph 28) and the test can be conducted without euthanizing the animals. These properties facilitate integration of the *Pig-a* assay into many *in vivo* rodent testing protocols (see paragraphs 7-8 and Annex 2) (2) (3).

3. This Test Guideline (TG) is based on the recommendations included in a Detailed Review Paper (DRP), which describes the scientific principles, technical conduct, performance, and context of use of the *Pig-a* assay (3). In addition, the DRP compares the strengths and weaknesses of the *Pig-a* assay in relation to other *in vivo* genetic toxicology tests (3). The DRP was based on the review by the International Workshops on Genotoxicity Testing of several *Pig-a* protocols (2). The main components and procedures for the *Pig-a* assay are described below. Further guidance for conducting the test, including detailed laboratory protocols that are compliant with the recommendations in this TG, can be found in these and other publications (4) (5) (6) (7).

4. An independent peer review (8) of the Retrospective Performance Assessment conducted on the *in vivo* erythrocyte *Pig-a* gene mutation assay (3) (9) concluded that the test was adequate for identifying *in vivo* mutagens with a high degree of accuracy. The review indicated that the test was sufficiently well validated for the development of a TG used to support regulatory decisions.

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

© OECD, (2022)

You are free to use this material subject to the terms and conditions available at http://www.oecd.org/termsandconditions/.

INITIAL CONSIDERATIONS AND LIMITATIONS

6. *Pig-a* gene mutations have been observed to accumulate with repeat dosing of mutagens, and increased frequencies of mutant erythrocytes generally persist for months in peripheral blood (10) (11) (12). These observations suggest that the kinetics by which *Pig-a* mutant cells appear and disappear from blood circulation are similar to those of *Pig-a* wild-type cells, indicating that the mutations measured in the *Pig*-a assay act as if they have a relatively neutral phenotype, *i.e.*, one that does not confer a cell growth advantage or disadvantage. Although different dosing regimens have been employed for the test, a protocol employing 28 consecutive days of dosing (or longer, see paragraph 43) takes advantage of mutant accumulation, provides sufficient time for mutant expression, and facilitates integrating the test into standard repeat-dose toxicity studies. The 28-day dosing protocol also has displayed slightly greater accuracy for detecting weak *in vivo* mutagens than shorter-term (*i.e.,* less than 14 consecutive days) dosing protocols (*e.g.*, concordance of 93% and 89% using two sets of criteria *vs*. 88% using data from short-term treatments only: see (3) for full discussion). Finally, a 28-day dosing protocol detects some mutagens at lower (per day) dose levels compared to shorter dosing schedules (13).

7. 3Rs principles (*i.e.*, replacement, reduction, and refinement) and experimental data support integrating the *Pig-a* assay with other test(s) whenever possible. However, there are several important factors that should be considered in order to do so effectively. The recommended procedure for conducting the *Pig-a* assay involves 28 consecutive days of dosing followed by testing for *Pig-a* gene mutation (paragraph 40). Thus, the test readily integrates into study designs that use such repeat-dosing schedules, *e.g.*, the 28-day subacute general toxicology studies described in OECD TG 407 (14) and TG 412 (15), or the transgenic rodent gene mutation test described in OECD TG 488 (16). The *Pig-a* assay also can be combined with tests with extended exposure durations, such as the subchronic toxicity studies described in OECD TG 408 (17) and TG 413 (18). Combination of *Pig-a* assays with tests having short-term treatment and sampling schedules is discussed in Annex 2.

8. Under the conditions outlined in paragraph 7 and Annex 2, the *Pig-a* and erythrocyte micronucleus tests can be integrated into single experiments, since only small volumes of peripheral blood are required for each endpoint. Thus, both clastogenicity/aneugenicity (micronucleus test) and gene mutation (*Pig-a* assay) can be evaluated in the same study (19). The *Pig-a* assay also complements the *in vivo* comet test (so long as mutant manifestation time is taken into account), since the latter detects a different type of endpoint (DNA damage in the form of strand breaks) than the *Pig-a* assay. The comet test also shows low sensitivity for detecting test substances that induce mutation in the bone marrow compartment (20), while the *Pig-a* assay specifically measures mutations induced in bone marrow erythroid cells (paragraph 10).

9. The *Pig-a* gene is relatively large (1302 bp of protein-coding sequence in 5 exons). Mutations that produce the *Pig-a* mutant phenotype (see paragraph 15) have been detected throughout the coding region and flanking splice-site sequences, without any major hot-spots for mutation (21) (22) (23) (24). Most of the mutations detected by the test have been base-pair substitutions and frameshifts, including mutations resulting in exon deletions, without any known bias as to the nature of the mutations detected.

10. The erythrocyte *Pig-a* assay detects mutations induced in erythroid precursor cells, which in adult rodents are primarily found in the bone marrow. Because the bone marrow is a well-perfused tissue, many substances that induce mutations in other tissues, induce mutation in the bone marrow as well. The assay is sensitive to both test chemicals that are direct-acting mutagens and test chemicals that require metabolic activation to induce mutation (*i.e.*, promutagens). However, the *Pig-a* assay is not applicable for assessing the mutagenicity of test chemicals or their metabolites that do not reach the bone marrow (see paragraph 71). In general, the *Pig-a* assay described in this TG cannot investigate genotoxicity at the site of first contact.

11. Analyses of *Pig-a* gene mutation spectra have been conducted by DNA sequence analysis of *Pig-a* mutant phenotype erythroid precursor cells from the bone marrow of rodents dosed with known

© OECD, (2022)

mutagens. Findings from these studies indicate that *Pig-a* mutant phenotype cells contain *Pig-a* mutations while *Pig-a* wild-type cells do not (21) (22) (23) (24).

12. Although this TG is applicable only to rodent blood erythrocyte-based *Pig-a* assays, *Pig-a* mutations can be detected *in vitro*, in cultured mammalian cell lines, and in various animal species, including humans (3) (25) (26). Analogous measurements in the peripheral blood of other mammalian species may be feasible and eventually valuable, provided adequate sensitivity and specificity to detect mutagenic and non-mutagenic substances have been demonstrated.

13. Most peer-reviewed *Pig-a* assay data are from rats. Mouse data collected to date (1) (25) indicate this species also can be used for the assay (3). However, several characteristics of mouse-based tests, including the low circulating blood volume in mice relative to rats, should be taken into consideration when designing studies in mice (*e.g.*, see paragraph 56).

14. The most appropriate *in vivo* methods for testing manufactured nanomaterials may be tests that can measure effects at the site of contact (27), making the *Pig-a* assay of lesser relevance for testing these types of substances. The *Pig-a* assay, however, may be appropriate for testing the mutagenicity of manufactured nanomaterials provided that bone marrow or systemic availability of the manufactured nanomaterials can be demonstrated (*e.g.,* following intravenous treatments) (28) (29).

PRINCIPLE OF THE TEST METHOD

15. The protein encoded by the *Pig-a* gene is necessary for the biosynthesis of glycosylphosphatidylinositol (GPI) cell membrane anchors (30). GPI anchors link approximately 150 unique proteins to the exoplasmic surface of mammalian cytoplasmic membranes. Of the approximately 30 genes involved in GPI synthesis, *Pig-a* is the only gene located on the X-chromosome. As there is only one X chromosome in males and one functional X chromosome in females, a single mutation has the potential to inactivate the *Pig-a* gene product and result in cells deficient in GPI anchors and their associated GPI-anchored proteins. Thus, male and female animals are equally sensitive to *Pig-a* mutant induction (3). Other genes involved in GPI biosynthesis only can be inactivated by mutations in both alleles, making them less likely targets for GPI inactivation and their mutation less likely to cause GPI anchor deficiency, which is the mutant phenotype measured by the test (31).

16. Microliter volumes of peripheral blood contain millions of red blood cells (RBC) [*e.g.*, (32)], a few percent of which are immature erythrocytes, also known as reticulocytes (RET). RET contain residual amounts of RNA and express surface markers, such as CD71, that are typical of erythroid precursor cells, whereas mature RBC do not. For the purpose of this guideline, RBC are used to signify all circulating erythrocytes, irrespective of age, while RET refers to the most immature fraction of circulating erythrocytes. RET can be discriminated from mature RBC based on their residual RNA and/or presence of the CD71 surface marker using appropriate fluorescent dyes or antibodies.

17. Peripheral blood RBC and RET can be analysed by flow cytometry for the *Pig-a* mutant phenotype (*i.e.*, loss of GPI anchors) by labeling with fluorochrome-conjugated antibodies against GPI-anchored proteins (33) (34) (35). Thus, *Pig-a* mutant cells will be non-fluorescent with regard to GPI-anchored epitope(s), and *Pig-a* wild-type cells will fluoresce. In this way, millions of labeled RBC and RET can be tested for the *Pig-a* mutant phenotype using flow cytometry. As noted in paragraphs 23-25 and 60, methods that concentrate RET and/or mutant cells have been developed to make evaluating million(s) of cells per sample practical.

18. Both the RET and total RBC populations should be evaluated for the mutant phenotype. Increased frequencies of mutant reticulocytes (MUT RET) appear sooner in peripheral blood circulation compared to

© OECD, (2022)

mutant erythrocytes (MUT RBC). Even so, it is important to enumerate MUT RBC in addition to MUT RET because MUT RBC represent an important confirmatory, lagging indicator of mutation. The demonstration of mutant manifestation in both cohorts of erythrocytes represents the strongest evidence of test chemical-induced mutagenic activity (see paragraph 66), and the dosing/blood collection schedules described herein (see paragraphs 40-45) were designed to provide sufficient time for both MUT RET and MUT RBC to be increased by exposure to mutagenic substances. Indeed, statistically significant increases in only one population, MUT RET or MUT RBC, is a rare occurrence when rodents are exposed to known mutagenic substances and when a 28-day repeat-dose protocol is utilised. However, since mutations are assayed at only one or a few sampling times, it is conceivable that positive MUT RET, but not MUT RBC responses could occur for a late arising RBC mutant frequency or that negative responses are detected stochastically in one or the other erythrocyte cohort. Such results benefit from expert evaluation for establishing the mutagenicity of the test chemical.

19. In addition to collecting data on *Pig-a* mutant induction, the fraction of RET among total RBC can be used as a measure of toxicity to the erythroid cell compartment. Percent RET (%RET) data can be useful for establishing that the bone marrow has been exposed to biologically significant levels of the test chemical or its metabolites ((2) (7); see also paragraphs 49 and 55). When blood samples are collected within approximately 3 days following cessation of dosing, bone marrow toxicity often manifests as a reduction in %RET relative to %RET in concurrent vehicle/solvent control animals. However, it is important to note that when blood samples are collected later than approximately 3 days following cessation of dosing, bone marrow toxicity can at times be expressed as elevated %RET relative to concurrent vehicle/solvent control animals (36). Elevated %RET following dosing is colloquially referred to as a ‘rebound effect’ or ‘stress erythropoiesis’, and usually represents a compensatory response to a period of bone marrow toxicity, especially involving reduced erythropoiesis. Finally, dose-related changes in %RET may occur if the test chemical causes haemolysis (37). Whatever the cause, dose-related changes in %RET indicate systemic exposure at biologically significant levels.

VERIFICATION OF LABORATORY PROFICIENCY

Training and proficiency investigations

20. Before laboratories conduct the *Pig-a* assay for regulatory safety assessments, it is important for staff to gain experience with blood collection, blood labeling and flow cytometric procedures. To minimise the number of animals required for the preliminary training exercises, reconstruction experiments, also known as spiking experiments, can be conducted as described in Annex 3 (38).

21. Establishing laboratory proficiency requires reproducing the expected results from published data (1). This should be demonstrated for both MUT RBC and MUT RET frequencies using a minimum of two mutagens (including weak responses induced by low doses of mutagens; see Table 1) and with compatible vehicle/solvent controls (see paragraph 31). These experiments should use doses of mutagens that give reproducible and dose-related increases in mutant frequency and demonstrate the sensitivity and dynamic range of the test system using the methods employed in the laboratory. Note that by collecting and analysing pre-dosing blood samples (see also paragraph 46) along with vehicle/solvent control samples from rodents used in these proficiency experiments, laboratories can begin building up their historical negative control data base (described in more detail in paragraph 22 and Annex 4, section 4.1).

© OECD, (2022)

Table I. Recommended Mutagens for Training and Proficiency Testing.

Chemical[a]	Cas No.	Vehicle	Route	Highest Suggested Rat Dose[b] (mg/kg/day)	Highest Suggested Mouse Dose[b] (mg/kg/day)	Blood Collection Day[c]
N-Ethyl-*N*-Nitrosourea, also known as ENU[d]	759-73-9	Phosphate-buffered saline, pH 6.0	Oral gavage	20	40	15-30
Benzo[*a*]pyrene, also known as B[a]P[e]	50-32-8	Sesame oil or Corn oil	Oral gavage	125	125	15-30
7,12-Dimethylbenz[a]-anthracene, also known as DMBA[e]	57-97-6	Sesame oil or Corn oil	Oral gavage	30	75	15-30

[a]Other mutagens can be used if justified.
[b]Generally well-tolerated, highly mutagenic dose level assuming 3 consecutive days of administration (Days 1-3) (11) (39).
[c]Blood collection day describes a range of recommended days for collecting peripheral blood for the *Pig-a* assay with the days relative to the start of dosing (*i.e.*, Day 1).
[d]Potent direct-acting mutagen; pH 6 phosphate-buffered saline should be used to dissolve ENU in order to inhibit hydrolysis; the ENU dosing solution(s) should be prepared fresh each day of use; commercial supplies of ENU have large amounts of stabilisers and this needs to be accounted for when preparing dosing solution(s).
[e]Requires metabolic activation.

Historical negative control data

22. During the course of the proficiency investigations, each laboratory should establish historical negative control ranges and distributions for MUT RET and MUT RBC frequencies. Negative control data can be obtained from two sources. First, negative control data can be derived from animals tested as negative controls in conjunction with animals dosed with test chemicals (hereafter referred to as concurrent negative controls). These animals typically will be dosed with the solvent/vehicle alone (paragraph 31). Secondly, for laboratories that conduct pre-dosing *Pig-a* assays (paragraph 46), negative control data can be obtained from these untreated, naive animals (referred to as pre-dose control data). Pre-dose animals have been shown to have *Pig-a* mutant frequencies equivalent to solvent/vehicle concurrent control animals when tests are conducted soon after a 28-day repeat-dose treatment protocol, as recommended in this TG (3). Whatever the source(s) of negative control data, animals that exhibit exceptionally high mutant cell frequencies (*e.g.*, identified by an outlier test) should not be used to construct negative control distributions, as they may have experienced a jackpot-type mutational event (as defined in paragraph 46 and Annex 1). Negative control data and distributions are important for determining the minimum number of cells that should be evaluated for the mutant phenotype (paragraph 60), assessing test acceptability (paragraph 64), and for evaluating and interpreting test chemical-related effects (paragraph 66). More detailed recommendations for developing historical negative control datasets are given in Annex 4, section 4.1.

© OECD, (2022)

DESCRIPTION OF THE METHOD

Intellectual property considerations

23. Some of the specific techniques used for assessing the mutagenic DNA-damaging potential of an agent and for evaluating the effects of an agent that may modify mutagenic DNA damage based on GPI-anchor-deficient erythrocytes are patented by Litron Laboratories in the U.S., Canada, and Europe. Note that this patent protection does not cover the principle of the assay or the general design of scoring GPI negative cells. Laboratories implementing the test method can obtain kits for cell analysis from Litron (commercially known as MutaFlow® kits) or ship blood samples to a proficient laboratory for testing with a Litron kit. MutaFlow kits come with a limited-use license, and a signature is not necessary. Alternatively, laboratories that do not purchase a kit from Litron or an authorised kit distributor must sign a license agreement with Litron Laboratories to perform the techniques covered by the patents in the U.S., Canada, and Europe.

24. The PIGRET method including the immunomagnetic separation for reticulocyte enrichment is patented by Teijin Limited in Japan only, and requires signing a license with Teijin to use this technique in Japan.

25. This TG describes the requirements and important characteristics of erythrocyte-based analysis of *Pig-a* mutation. In regions where patents apply, protocols using alternative techniques, that do not infringe upon the patents held by Litron Laboratories or Teijin Limited, may be conducted freely using the guidance provided by this TG. The use of Litron or Teijin patented techniques are not necessary to conduct a valid test as long as the acceptability criteria (paragraph 64) are fulfilled.

Preparations

Selection of animal species

26. Common laboratory strains of healthy young adult rats and mice are normally used (*i.e.*, 6-12 weeks old at the start of dosing). Slightly younger or older animals are also acceptable (*e.g.*, 4 weeks of age at the start of treatment). However, such deviations should be justified. Justification could include facilitating integration with other tests or to avoid killing animals that have been bred but not used in a procedure. Provision should be made for any alterations to the historical negative control data base such deviations in age may cause. Strains of rats and mice such as Han Wistar, Sprague-Dawley and Fisher 344 rats and CD-1 mice have been used extensively (1). The choice of rodent species and strain should be based on (i) the animals used in other toxicity studies (to be able to correlate data and to facilitate integration), (ii) animals that developed tumours in a carcinogenicity study (when investigating the mode of action for carcinogenesis), or (iii) animals with the most relevant metabolism for humans, if known.

Animal housing and feeding conditions

27. The temperature in the experimental animal room ideally should be 22°C (±3°C). The relative humidity ideally should be 50-60%, being at least 30% and preferably not exceeding 70% other than during room cleaning. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with *ad libitum* drinking water. The choice of diet may be influenced by the need to ensure a suitable mixture of a test chemical when administered by the diet. Animals should be housed in small groups (usually no more than two for rats, five for mice) of the same sex if no aggressive behaviour is expected. Cages should conform with animal welfare standards [*e.g.*, (40)]. Animals may be housed individually only if scientifically justified. Based on Animal Care and Use

© OECD, (2022)

Committee (or equivalent) recommendations, solid floors should be used wherever possible and appropriate environmental enrichment should be provided.

Preparation of animals

28. Animals are randomly assigned to solvent/vehicle control and test chemical dosing groups. The animals are identified uniquely and acclimated to the laboratory conditions for a least five days (or as prescribed by the applicable Animal Care and Use Committee or equivalent) before the start of dosing. The least invasive method of uniquely identifying animals should be used. Appropriate methods include ringing, tagging, micro-chipping and biometric identification. Toe and ear clipping are not scientifically justified for these tests. Cages should be arranged in such a way that possible effects due to cage placement are minimised. At the commencement of the study, the weight variation for each sex should not exceed ±20% of the mean weight.

Preparation of doses

29. Solid test chemicals should be dissolved or suspended in appropriate solvents or vehicles (*e.g.*, water, corn oil; see (3) for examples) or mixed in diet or drinking water prior to dosing of the animals. Liquid test chemicals can be dosed directly, if necessary; however, dilution with an appropriate vehicle should be used to generate different doses and maintain a consistent dose volume among the treatment groups. For inhalation exposures, test chemicals can be administered as a gas, vapour, or solid/liquid aerosol, depending on their physicochemical properties (15) (18).

30. Fresh preparations of the test chemical should be employed unless stability data demonstrate the acceptability of the storage conditions used.

Test conditions

Solvent/vehicle

31. The solvent/vehicle should not produce toxic effects at the dose volumes used and should not be suspected of chemical reactions with the test chemicals. If other than well-known solvents or vehicles are used, their inclusion should be supported with reference data indicating their non-reactivity with the test chemical, and their compatibility in terms of test animals, route of administration, and endpoint. It is recommended that wherever possible, the use of aqueous solvents/vehicles should be considered first.

© OECD, (2022)

Controls

Positive controls and flow cytometry standards

32. Laboratories will have established their proficiency for conducting the *Pig-a* assay as described in paragraphs 20-21. It is only in this context that the inclusion of concurrent positive control animals are not required for each test. Rather than relying on mutagen-exposed animals, biological flow cytometry standards are included in each analysis (paragraph 33) and represent a useful type of analytical control that is 3Rs-friendly and focuses quality assurance measures on important factors that demonstrate the reliability of a study (2) (7) (11). Specifically, the use of blood-based flow cytometry standards ensures that i) blood collection and cell staining/labeling were performed properly, ii) the reagents used to stain/label cells were effective, and iii) the flow cytometer adequately differentiated mutant phenotype cells from wild-type cells. Finally, in order to assure continued proficiency in detecting increases in mutant frequency, laboratories should occasionally (at least once per year) perform additional tests with mutagen-treated animals similar to the proficiency studies described in paragraph 21.

33. Given their importance, it is necessary that biological flow cytometry standards be made and used each time *Pig-a* mutant analyses are conducted. For this purpose, a "mutant mimic" (*i.e.*, a blood sample prepared to have the properties of mutant erythrocytes when tested) or comparable sample (see Annex 5) is prepared that demonstrates the light scatter and fluorescence characteristics of MUT RET and MUT RBC versus wild-type RET and wild-type RBC. Mutant mimics can be created by processing extra blood from a solvent/vehicle concurrent control animal; a portion of the blood is prepared normally for evaluating *Pig-a* mutant frequencies, but another portion is prepared by omitting the fluorescent antibody(s) for GPI-anchored proteins from the fluorescent antibody labeling protocol (11) (38). To make the standards, these two blood samples are then mixed in approximately equal proportion. Since standards created from mutant mimics (or the alternative described in Annex 5) are valuable for guiding instrumentation settings and software/data analysis parameters, as well as serving as a control on the adequacy on the analytical procedure (paragraph 32), they should be generated for every study, and used each day blood samples are analysed. Once established, instrument settings generally should remain unchanged for all samples processed for mutant analysis throughout that particular data acquisition session.

Negative controls

34. Concurrent negative control animals should be included at every sampling time and otherwise handled in the same way as animals in the dosing groups, except for not being dosed with the test chemical. If a solvent/vehicle is used in administering the test chemical, the control group should receive the same solvent/vehicle. The MUT RET and MUT RBC frequencies from concurrent negative control animals of a given study should be consistent with the historical negative control distribution. There are several valid approaches for calculating lower and upper bound limits that can be used to facilitate such comparisons, and they are described in Annex 4, section 4.2.

35. The requirement for analysing blood from solvent/vehicle concurrent negative-control animals at every sampling time also applies to studies that include the collection and analysis of pre-dose blood samples (paragraph 46). That is, pre-dose blood samples are not a substitute for concurrent solvent/vehicle controls, irrespective of the statistical approach(es) utilised.

© OECD, (2022)

PROCEDURE

Number and sex of animals

36. It is technically feasible to conduct the *Pig-a* gene mutation assay using blood from animals of either sex (3). Where human exposure to the test chemical may be sex specific, as for example with some pharmaceuticals, the test should be performed with the appropriate sex.

37. When performing a single-sex study, group sizes at study initiation (and during establishment of laboratory proficiency) should be established with the aim of providing a minimum of 6 analysable animals [for justification of group size, see statistical power calculation in (41)]. Thus, a study conducted with one sex according to the parameters established in paragraph 47, with three dose groups and a concurrent negative control, would require 24 animals at the start of dosing. The number of animals will be higher when pre-dose blood samples are used to eliminate outliers as described in paragraph 46, or when additional animals are included to accommodate unforeseen events that reduce the number of animals. Expert judgement should be used in interpreting the data in situations where the number of animals in any group is less than 6 (*e.g.*, due to excluding an animal with a jackpot mutation response).

38. When data demonstrate relevant differences for the test chemical between males and females (*e.g.,* differences in systemic toxicity, metabolism, and bioavailability, including in a range-finding study), it is preferable to perform a study using both sexes.

39. For studies that treat the two sexes with different dose levels, the number of animals required per sex is the same as indicated in paragraph 37, at least 6 males and 6 females per group. For studies that treat sexes with the same dose levels, it is often possible to reduce the number of animals per group by taking advantage of a factorial design (see Annex 4 section 2.1). The factorial statistical design is capable of maintaining adequate statistical power while using fewer animals of each sex.

Dosing schedule and sampling times

40. A 28-day repeat-dose protocol is recommended for conducting the *Pig-a* assay. Of the dosing schedules published and analysed to date, 28-day repeat-dose schedules offer the most compelling evidence that a negative test result for mutation is accurate [*e.g.*, has the highest percent negative predictivity (3)]. Alternate dosing schedules can be utilised, if scientifically justified (see paragraph 45). Test chemicals may be administered as a single dose per day or as a split dose, *i.e.*, two or more exposures on the same day, to facilitate administering a large volume.

41. When animals are exposed to a test chemical for 28 consecutive days, blood should be collected at least once, within hours to day(s) of cessation of exposure (*e.g.*, Day[1] 28-31). Data collected from testing diverse genotoxicants suggest that sampling on Day 28-31 gives sufficient time for manifestation of MUT RET and MUT RBC responses and that the exact timing of sample collection is not critical. This schedule has the advantage of facilitating integration of the *Pig-a* assay with commonly utilised general toxicology studies and other genetic toxicology tests (paragraphs 7-8). Note that an extra administration of the test chemical on Day 29, followed by sample collection several hours later, is permissible for facilitating tissue harvest for conducting the *in vivo* comet test.

42. Optionally, and when logistically feasible, there may be merit to conducting *Pig-a* analyses on blood samples collected from a 28-day repeat-dose protocol at an additional, later time point. For instance,

[1] Here and elsewhere, study day is relative to start of dosing, which is defined as Day 1.

© OECD, (2022)

some toxicology experiments include satellite “recovery” or “withdrawal” groups to evaluate whether toxic effects are diminished or increased upon discontinuation of dosing. Such blood samples are typically collected between 2 to 4 weeks after cessation of dosing, and represent an opportunity to evaluate erythrocytes for the *Pig-a* mutant phenotype after additional time for mutant manifestation.

43. When the dosing period exceeds that of a 28-day repeat-dose study (*e.g.*, 90-day subchronic studies), blood samples should be collected at least once, after approximately 28 days of dosing (*e.g.*, at Day 28-31). As explained in paragraph 42, there may be merit to collecting blood later in the experimental protocol, (*e.g.*, within a few days of cessation of dosing), but recognizing the possibility that clonal expansions of erythroid precursor cells may disproportionately affect mutant frequencies at extended sampling times [*e.g.*, (42)]. In addition, sampling of animals at much later times (*e.g.*, sampling animals that are ≥8 months of age) may require developing an historical negative control data base for older animals (see Annex 4, section 4.1).

44. Dosing schedules that involve fewer than 28-consecutive days of exposure (*e.g.*, dosing on one or three consecutive days) should be scientifically justified as they exhibit lower sensitivity than repeat-dose, 28-day treatment protocols for detecting mutagens (paragraph 6). If a short-term treatment protocol is used, blood collection and analyses should occur at least two times: at approximately Day 15, and again at approximately Day 30. Test sensitivity for detecting mutagens is anticipated to be higher for short-term dosing schedules that involve two post-dosing blood collection/analysis times as opposed to one (13).

45. Other dosing and sampling schedules may be used when scientifically justified, especially in instances when they facilitate integration with other toxicology studies, or combination with other genotoxicity tests. For instance, protocols that do not involve daily dosing but do use several administrations of test chemical per week for several weeks have been shown to retain high *Pig-a* assay sensitivity while facilitating combination with other genotoxicity tests (43). As above, such studies need to take into consideration the different manifestation times of MUT RET and MUT RBC responses and the manifestation times for any other genotoxicity tests being conducted as part of the study (see discussion in Annex 2, section 2 regarding combining *Pig-a* assays with short-term treatment schedules with other short-term genetic toxicology assays).

46. Whatever the dosing schedule, it may be advantageous (but not required) to perform *Pig-a* analyses before the first administration of the test chemical. Pre-dosing samples are taken within one week prior to test chemical dosing (if adult mice are used, it is recommended that pre-dose sampling be conducted ~7 days prior to dosing to allow the haemopoietic system to re-establish homeostasis). Mutation ‘jackpots’, resulting in mutant frequencies that are unusually high and outside the historical negative control distribution (see Annex 4, section 4.2 for approaches for describing the distribution), may occur due to mutations in germ cells or in early erythroid progenitor cells. It may be useful to identify animals having such mutant frequencies so that they can be excluded from the main study and replaced with animals having mutant frequencies within the historical negative control distribution.

Dose groups and dose levels

47. A minimum of three test chemical dosing groups and a negative control group generally should be used. Additional dose groups may be necessary for a detailed analysis of dose-response (see paragraph 48); the use of fewer dose groups should be justified. Except for dosing with the test chemical, animals in the concurrent negative control group should be handled in an identical manner to animals in the test chemical groups. If a solvent/vehicle is used in administering the test chemical, the concurrent control group and all dosed groups should receive an equal volume of the vehicle or solvent.

48. Any existing toxicity and toxicokinetic data for the test chemical should be taken into consideration in setting dose levels. If a preliminary range-finding study is performed because there are insufficient

© OECD, (2022)

suitable data already available to guide dose selection, it should be performed in the same laboratory, using the same strain, source, sex, and treatment route to be used in the main study. The highest dose level should be chosen with the aim of inducing toxic effects but not death or severe suffering [see (44)]. This definition is taken from TG 407 (14) and is used here to facilitate integration of the *Pig-a* endpoint with other repeat-dose studies, thereby maximising the toxicological information acquired. When limited by other factors, such as palatability for dietary or drinking water administration (and when dosing cannot be done by gavage) or explosiveness for test chemicals administered by inhalation, the highest dose level will be the maximum feasible dose. Thereafter, a descending sequence of dose levels should be selected with a view to demonstrating any dose-related response. Two- to four-fold intervals are frequently used for setting the descending dose levels when genotoxicity endpoints are evaluated. With most test chemicals, the dose levels used should cover a range from the maximum to little or no toxicity. When additional toxicity endpoints are integrated into the study, other dose spacing may be considered. The addition of a fourth test group is often preferable to using excessively large spacing between doses.

49. The highest dose also may be defined as a dose that produces toxicity to the bone marrow compartment (see paragraph 19). When this evidence is provided by a reduction in mean %RET (relative to the mean concurrent solvent/vehicle control, measured *e.g.* with blood sampled 1-3 days after a 28-day dosing schedule), the reduction should be statistically significant, but ideally it should not exceed 80%. This is because new erythrocytes need to be generated in order to manifest a chemical-induced effect in this test.

50. For test chemicals that do not produce toxicity, or that are not limited by other factors (see paragraph 48), the highest dose for dosing protocols of $\geq$ 14 consecutive days is 1000 mg/kg/day, while the maximum for protocols involving shorter-term dosing is 2000 mg/kg/day.

Limit test

51. As noted above, studies normally should employ three test chemical dosing groups and a solvent/vehicle concurrent control group. However, if from assessment of other data, no observable toxic effects (including no depression of bone marrow proliferation or other evidence of target tissue cytotoxicity) would be expected at the highest test dose of 1000 mg/kg body weight/day in a study of $\geq$ 14 days of dosing or 2000 mg/kg/day in studies employing <14 days of dosing (paragraph 50), and if genotoxicity would not be expected based upon *in vitro* genotoxicity studies or data from structurally related substances, then this one dose level can be used. In the case of inhalation exposures, the limit test concentrations are 20 mg/L, 5 mg/L or 20,000 ppm for vapours, dusts/mists (aerosols) and gases, respectively. Limit test doses may not apply in cases where human exposure indicates the need to use a higher dose level (15). Also, when integrating or combining the *Pig-a* assay with other test(s), the applicability of using a single limit dose to these other test(s) should be ensured before conducting a limit test. For instance, the single limit dose provision does not include TG 489 for which the full 3-dose regimen applies (45).

Administration of doses

52. The primary concern in choosing a route of administration is ensuring exposure of the target tissue, bone marrow; however, the anticipated route of human exposure also should be considered. Therefore, routes of exposure such as oral (by gavage), dietary, drinking water, topical, subcutaneous, intravenous, inhalation, intratracheal, or implantation may be chosen as justified. Intraperitoneal injection is generally not recommended since it is not a usual route of human exposure and only should be used with specific scientific justification. If the test chemical is mixed in diet or drinking water, especially in the case of single

© OECD, (2022)

or short-term dosing, care should be taken to ensure that the delay between food and water consumption and blood sampling provide time for mutant manifestation (see paragraphs 40-45).

53. 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 normally exceed 1 mL/100 g body weight except in the case of aqueous solutions where a maximum of 2 mL/100 g may be used. The use of volumes greater than this should be justified and all volumes should be consistent with applicable Animal Care and Use Committee (or equivalent) guidelines. Except for irritating or corrosive test chemicals, which will normally produce exacerbated effects at higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure administration of a constant volume in relation to body weight at all dose levels.

Observations

54. General clinical observations of the test animals should be made, and clinical signs recorded at least once each day, preferably at the same time(s) and considering the peak period of anticipated effects after dosing (44). At least twice each day all animals should be observed for morbidity and mortality. All animals should be weighed at least once each week, and when they are euthanised. Quantitative measurements of food consumption should be made at least weekly when test chemicals are administered in the diet; general observations of food consumption should be made to ascertain animal health. If the test chemical is administered *via* the drinking water, water consumption should be measured at each change of water and at least weekly. Animal Care and Use Committee (or equivalent) guidelines should be observed; but in general, animals exhibiting non-lethal indicators of excess toxicity should be euthanised prior to completion of the test period.

Target tissue exposure

55. A blood sample should be taken at appropriate times(s) in order to permit measuring the plasma levels of the test chemical and/or its metabolite(s) for the purposes of demonstrating that exposure of the bone marrow occurred and where other exposure data do not exist (*e.g.*, significant perturbations to %RET, see paragraphs 19 and 49; see paragraph 71 for further discussion). The blood sample should be large enough to detect relevant levels of the test chemical and its metabolites without perturbing the biological processes being measured in the animals. The bone marrow is well perfused by the peripheral blood circulation, receiving up to 15% of the cardiac output, and measurement of plasma levels of the test chemical and its metabolites is considered qualitative evidence for exposure of bone marrow (46) (47) (48) (49).

Blood collection for conducting the Pig-a assay

56. Applying animal welfare standards, small volumes of peripheral blood (less than about 300 µl) are obtained either using a method that permits survival of the test animal, such as bleeding from the tail vein, jugular vein, or other appropriate blood vessel, or by cardiac puncture or sampling from a large blood vessel when the animal is humanely killed. The size of the animal and animal welfare standards may limit the options for the method used. As flow cytometric analysis requires single cell suspensions, care should be taken to avoid blood coagulation; this is normally accomplished using an anticoagulant, such as heparin and/or EDTA. Samples should be collected and processed blinded as to the treatment or by employing a blocked design (Annex 4, section1).

© OECD, (2022)

Storage of blood

57. In the presence of anticoagulant, blood samples can be stored for up to 5 days before they are processed for flow cytometric analysis as long as they are maintained cold, but not frozen (*e.g.*, in a 2-8°C refrigerator) (2). Furthermore, in the presence of anticoagulant, blood samples can be shipped to an analytical test site provided they are maintained cold, not frozen, throughout transportation and any subsequent storage, and as long as they are further processed and analysed within approximately 5 days of collection (2). Whatever storage and/or shipment method is utilised, each laboratory should develop their own data to support it.

58. Procedures have been described for freezing and later thawing blood samples for subsequent processing and flow cytometric analysis of *Pig-a* mutations (50). These procedures can be useful for delaying analysis for reasons that include but are not limited to instrument failure, deferring the decision to acquire *Pig-a* data, and storing blood from mutagen-treated animals for use as analytical standards (Annex 5). Furthermore, frozen blood samples can be transported from the animal experimental site to an analytical site as long as they are maintained frozen throughout transportation (*e.g.*, on dry ice) (50). Whatever freezing and thawing method is employed, it is important to demonstrate minimal lysis of RBCs, and that the freezing and thawing process and the length of storage have minimal impact on MUT RET, MUT RBC, and %RET frequencies.

Methods of measurement

59. Flow cytometric analysis is used for determining %RET, MUT RET and MUT RBC frequencies in peripheral blood samples.

60. Zero (0×10^{-6}) mutant cell frequency readings should be kept to an occasional rather than a common occurrence. Proficient laboratories have shown that for commonly used rodent models, mean baseline MUT RET and MUT RBC frequencies typically range from $1\text{-}3 \times 10^{-6}$. With this in mind, it is generally necessary for most laboratories to evaluate a minimum of $1\text{-}3 \times 10^{6}$ RBC and $1\text{-}3 \times 10^{6}$ RET per animal per time point. Note that given the rarity of RET in peripheral blood circulation, it is not practical to evaluate $\geq 1 \times 10^{6}$ RET directly from blood samples. In order to overcome this problem, immunomagnetic separation procedures were developed (51) (52) to increase the number of RET (and in some cases the number of RBC) investigated for mutant cell frequency measurements. These immunomagnetic separation techniques, or a validated alternative, can be used to facilitate conducting the *Pig-a* assay.

61. As described in paragraph 60, each laboratory should set the number of cells analysed to minimise the occurrence of zero mutant cell frequency readings. Even so, it is possible that for a particular study, more than an occasional zero mutant cell frequency reading will be observed. In order to avoid invalidating such a study, the number of cells evaluated per animal should be increased by processing left-over blood samples that have been stored properly (paragraphs 57-58).

62. The preferred antibodies for scoring the incidence of MUT RET and MUT RBC are anti-CD59 for rats and anti-CD24 for mice, which target the GPI-anchored proteins CD59 and CD24, respectively. Other GPI-anchored proteins exist on the surface of wild-type RBCs (*e.g.*, CD55) that potentially could be used in the test, and it also is possible to use combinations of antibodies for GPI-anchored proteins to distinguish wild-type and mutant phenotype cells (*e.g.*, anti-CD59 plus anti-CD55). Alternative antibodies or combinations of antibodies for identifying *Pig-a* mutant and wild-type erythrocytes can be used if they generate known *Pig-a* responses (1) and upon demonstration of proficiency by the laboratory (see paragraph 21).

© OECD, (2022)

DATA AND REPORTING

Treatment of results

63. Individual animal data, as well as group means and a measure of within-group variation (*e.g.*, standard deviation or standard error), should be presented in tabular form. The *number* of RET, RBC, MUT RET, and MUT RBC scored should be listed separately for each animal and time point analysed, and they should be accompanied by RET, MUT RET, and MUT RBC *frequencies* for each individual animal. Whereas RET frequency is usually expressed as a percentage, it is preferable to express MUT RET frequency as number of MUT RET per 1 million total RET, and MUT RBC frequency as the number of MUT RBC per 1 million total RBC. Data on animal toxicity and clinical signs also should be reported.

Statistical evaluation and interpretation of results

Acceptability criteria

64. For a valid test, data analysis is conducted only when the study fulfills all acceptability criteria in terms of animal number, dosing, sampling times, mutant analysis, etc. described in paragraphs 36-62. The following criteria determine the acceptability of the test:

- The concurrent negative control frequencies are consistent with the distribution described by the laboratory historical negative control data base. It is important that the historical negative control data base was constructed appropriately (for instance as recommended in paragraph 22 and Annex 4, section 4.1), and the historical negative control mutant frequencies and their distribution are consistent with literature values (1) (3) (7).
- When they are used, concurrent positive controls should induce a statistically significant increase compared to the concurrent negative control (see paragraphs 34-35).
- Mutant mimic controls or historical samples from animals treated with a positive control substance were used for setting flow cytometer gates for experimental sample evaluations as described in paragraph 33 and Annex 5.
- The number of dose levels, route of exposure, and the criteria for the selection of the highest dose are consistent with those described in paragraphs 47-52.
- The appropriate numbers of MUT RBC and MUT RET have been analysed (paragraph 60).

Evaluation and interpretation of results

65. Data from *Pig-a* experiments are generally evaluated with the intention of determining whether the frequencies of MUT RET and MUT RBC in test chemical dosing groups are increased relative to those of the concurrent solvent/vehicle control group. Such analyses should be accomplished by appropriate statistical methods, using the animal as the experimental unit. The statistical method(s) employed should be indicated.

66. When evaluating the responses, all data should be taken into consideration and, in all cases, expert judgement applied.

- Provided that the acceptability criteria in paragraph 64 are fulfilled, a test chemical is considered clearly positive if all of the following criteria are met:

© OECD, (2022)

a) At least one of the treatment groups exhibits a statistically significant increase in both MUT RET and MUT RBC frequency compared with the concurrent solvent/vehicle control;
b) The mutant frequency responses are dose-related, for example when evaluated with an appropriate trend test (Annex 4, section 3) (not applicable to the limit test);
c) Both the MUT RET and MUT RBC frequencies of any of the test chemical dose groups (at any post-treatment time point) exceeds the upper bound limit of the historical negative control data distribution.

Positive results indicate that, under the test conditions, the test chemical induces *Pig-a* gene mutations in the erythropoietic cells of the test species.

- Provided that the 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) No MUT RET nor MUT RBC frequency, for any test chemical dose group, exhibits a statistically significant increase compared with the concurrent negative control;
b) None of the mutant frequency responses are dose-related, for example when evaluated by an appropriate trend test (not applicable to the limit test);
c) Neither the MUT RET nor MUT RBC frequency in any test chemical dose group exceeds the upper bound limit of the historical negative control data distribution.
d) Bone marrow exposure to the test chemical(s) and/or its metabolites occurred (see paragraph 71).

Negative results indicate that, under the test conditions, the test chemical does not induce Pig-a gene mutations in the erythropoietic cells of the test species.

67. Regarding the statistical approaches described in paragraph 66 above, it is important to recognise there is no single correct method of conducting a statistical analysis. A practical approach is to suggest a particular set of statistical analyses as *an example* of the sort of evaluations that can be carried out. This is the context that three types of analyses (A-C) described above were presented. See Annex 4, sections 1-3 for information that testing laboratories and regulatory reviewers may find useful when considering whether a particular statistical analysis is appropriate and whether the data interpretation is scientifically sound. Whatever statistical approach is utilised, it is important that it is described in advance of conducting a study.

68. If some, but not all of the Criteria are met for a clearly negative or clearly positive result, then expert judgment will be applied in an effort to interpret the results as either positive or negative. The application of expert judgement is consistent with recommendations made by an OECD genotoxicity working group that met in Ottawa, Canada in 2013, who stated "…data should be interpreted based both on statistics and biological relevance" (53).

69. In some cases (*e.g.*, when additional, properly stored cells are available), analysing more cells (see paragraph 61) can be useful in resolving findings that are neither clearly positive nor clearly negative. If the application of expert judgement and the analysis of additional cells are unable to resolve a response as either positive or negative, a repeat experiment using the same or modified experimental conditions (*e.g.*, different strain, mouse instead of rat, different route of administration, different mutant manifestation time) might be needed.

© OECD, (2022)

70. Even after applying expert judgement and possibly analysing additional data, there may be instances when test results cannot be interpreted as either positive or negative; in such instances, the test is considered equivocal (Annex 4, section 6).

71. A negative result is only considered reliable when evidence is provided that the bone marrow (and/or general circulation) has been exposed to the test chemical and/or its metabolites (Criterion D in paragraph 66). With toxic substances, direct evidence of bone marrow exposure is provided by a statistically significant change in the fraction of RET among total RBC (%RET; see paragraph 19); in the absence of such evidence, measurement of the plasma or blood levels of the test chemical and its metabolites (*e.g.*, using the blood sample described in paragraph 55) can be useful for documenting exposure. A discussion of this topic is provided by the European Food Safety Agency (54). Alternatively, ADME (absorption, distribution, metabolism, and excretion) data, obtained in an independent study, using the same dosing route and same species (or a different species, with justification), can be used to demonstrate bone marrow exposure. In general, plasma levels calculated from such studies can be used for this purpose. In some instances, modeling software programs can assist in bridging information on blood levels and tissue concentrations and help in extrapolating from one route to another or from one species to a different species (55). Finally, a weight of evidence approach can be used to demonstrate systemic and/or bone marrow exposure to the test chemical, *e.g.* for coloured test substances, a matching urine colour could be an indicator of systemic exposure to the test substance.

72. There is no requirement for further verification of a clearly positive or clearly negative response (see paragraph 66).

Test report

73. Consistent with Good Laboratory Practice, the test report should include the following information:

Evidence for laboratory proficiency

Test chemical:

- Source, lot number, and expiration/retest date, if available;
- Physical nature and purity;
- Physicochemical properties relevant to the study;
- Stability of the test chemical, if known;
- For mono-constituent substances: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, chemical identity of impurities as appropriate and practically feasible;
- For multi-constituent substances, substances of unknown or variable composition, complex reaction products or biological materials (UVBCs) and mixtures: characterization of chemical identity as far as possible (see above), quantitative occurrence and relevant physicochemical properties of the constituents;

Test chemical preparation:

- Justification for choice of solvent/vehicle;
- Solubility and stability of the test substance in the solvent/vehicle, if known;
- Preparation of dietary, drinking water or inhalation formulation, if appropriate;
- Analytical determinations on formulations (*e.g.*, stability, homogeneity, nominal concentrations), when conducted;

Test animals:

© OECD, (2022)

- Species and strain and justification for choice;
- Number, age and sex of animals;
- Cage type, housing conditions, diet, enrichment, etc.;
- Method used to uniquely identify animals;
- Procedure used to form dosing groups;
- Individual weight of animals at the start of the test, including body weight range, mean and standard deviation for each dosing group;

Test conditions (reference can be made to established/published methods in lieu of detailed descriptions):

- Protocol for dose range-finding study, if conducted;
- Rationale for dose level selection;
- Details of test chemical preparation;
- Details on the administration of the test chemical;
- Rationale for route and duration of administration (if other than recommended);
- Methods for analytical determinations of dose, when used;
- Details of food and water composition and quality;
- Methods for the measurement of animal toxicity, including %RET and, where available, histopathological or clinical pathological analysis, and frequency with which animal observations and body weights were obtained;
- Methods for the bioanalytical/toxicokinetic analysis (if conducted);
- Details and rationale for the peripheral blood sampling protocol, including pre-dose sampling (if conducted), and sampling within (if conducted) and after the dosing period, including recovery sampling (if conducted);
- Methods used for blood collection and storage, if used;
- Method used for humane killing, if part of the study;
- Methods for verifying that the test chemical reached the bone marrow, or general circulation, if negative results are obtained;
- Methods used for immunomagnetic separation of cells;
- Materials used for quantification of mutant cells, including source and lot of antibody(s) and immunomagnetic separation reagents;
- Methods used for calculation of mutant frequencies; rationale used for the number of RET and RBC interrogated for mutation (paragraph 60);
- Methods used for a positive control, if included in the analysis;
- Criteria for acceptability of the study;

Results:

- Data from dose range-finding study, if conducted;
- Animal condition prior to and throughout the test period, including body weight observations and signs of toxicity for individual animals;
- For dietary or drinking water exposures, dose (mg/kg body weight/day) calculated from diet/drinking water test chemical concentrations and individual animal consumption;
- Body weight, organ weight, clinical pathology and histopathology for individual animals, if available, at the conclusion of the study;

© OECD, (2022)

- Instrument calibration approach, including descriptions of: gating strategy, how the MUT RET and MUT RBC scoring regions were constructed (*e.g.*, use of mutant mimics, etc.), and graphical output from the flow cytometer demonstrating the fluorescent resolution achieved between mutant erythrocytes (or mutant mimics) and wild-type erythrocytes;
- Pre-dosing data for each animal, if collected, including number of cells and cell equivalents evaluated, the number of mutant cells observed (MUT RET and MUT RBC), the frequency of mutant cells (reported as number per million), and %RET. In addition, group means should be reported and a measure of within group variation (standard deviation or standard error) should be reported for each proposed dosing group;
- If pre-dosing data were used to eliminate any animals from the main study, provide the rationale;
- From the main study, data from each negative (solvent/vehicle) concurrent control animal and each dosed animal at each sampling and analysis point, including number of cells evaluated, the number of mutant cells observed (MUT RET and MUT RBC), the frequency of mutant cells (reported as number per million) and %RET. In addition, group means and a measure of within group variation should be reported for each sampling and analysis point. In addition, graphical flow cytometric output from each sample analysis should be available, upon request;
- Concurrent positive control data, if any, at each analysis time, including number of cells evaluated, the number of mutant cells observed (MUT RET and MUT RBC), the frequency of mutant cells (reported as number per million), and %RET. In addition, group means and a measure of within group variation should be reported for each analysis point. Include any laboratory historical control data or literature values for comparison;
- Historical negative control data with number of rodent blood specimens clearly specified, statistical description of the distribution including type of calculation used to generate lower and/or upper bound limits used to aid data interpretation and justification for its use (see Annex 4, section 4.2), as well as the time period covered and evidence that the assay is under control during the period covered;
- Statistical analyses performed to evaluate the effect of the test chemical on MUT RET, MUT RBC and %RET frequency;
- Data supporting that bone marrow exposure occurred for negative findings (see paragraph 71);
- Criteria for supporting a positive or negative response that are met;

Discussion of the Results.

- Explain how scientific judgement was used in evaluating the results of the test.

Conclusions.

© OECD, (2022)

LITERATURE

1. Shemansky JM, LP McDaniel, C Klimas, SD Dertinger, VN Dobrovolsky, T Kimoto, K Horibata, JE Polli, RH Heflich (2019) *Pig-a* gene mutation database. Environ Mol Mutagenesis 60:759-762.
2. Gollapudi BB, AM Lynch, RH Heflich, SD Dertinger, VN Dobrovolsky, R Froetschl, K Horibata, MO Kenyon, T Kimoto, DP Lovell, LF Stankowski Jr, PA White, KL Witt, JY Tanir (2015) The *in vivo Pig-a* assay: A report of the International Workshop on Genotoxicity Testing (IWGT) workgroup. Mutat Res 783:23-35.
3. OECD (Organisation for Economic Cooperation and Development) (2020) The in vivo erythrocyte *Pig-a* gene mutation assay - Part 1: Detailed Review Paper and Retrospective Performance Assessment. Available as Publication 315 at: http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm
4. Chikura S, T Kimoto, S Itoh, H Sanada, S Muto, K Horibata (2019) Standard protocol for the total red blood cell *Pig-a* assay used in the interlaboratory trial organized by the Mammalian Mutagenicity Study Group of the Japanese Environmental Mutagen Society. Genes Environ 41:5. Available at: https://doi.org/10.1186/s41021-019-0121-z.
5. Bemis JC, NE Hall, SD Dertinger (2019) Erythrocyte-based *Pig-a* gene mutation assay. In: A Dhawan, M Bajpayee (eds), Genotoxicity Assessment: Methods and Protocols, Methods in Molecular Biology, vol 2031, Humana, NY, NY. pp 29-57. Available at: https://doi.org/10.1007/978-1-4939-9646-9_2.
6. Chikura S, T Kimoto, S Itoh, H Sanada, S Muto, K Horibata (2021) Standard protocol for the PIGRET assay, a high-throughput reticulocyte *Pig-a* assay with an immunomagnetic separation, used in the interlaboratory trial organized by the Mammalian Mutagenicity Study Group of the Japanese Environmental Mutagen and Genome Society. Genes Environ 43:10. Available at: https://doi.org/10.1186/s41021-021-00181-7.
7. Dertinger SD, JA Bhalli, DJ Roberts, LF. Stankowski Jr, BB Gollapudi, DP Lovell, L Recio, T Kimoto, D Miura, RH Heflich (2021) Recommendations for conducting the rodent erythrocyte *Pig-a* assay: A report from the HESI GTTC *Pig-a* Workgroup. Environ Mol Mutagenesis 62:227-237.
8. OECD (Organisation for Economic Cooperation and Development) (2020) Report of the peer review of the validation status of the in vivo erythrocyte Pig-a gene mutation assay. Available as Publication 319 at: http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm
9. OECD (Organisation for Economic Cooperation and Development) (2020) The in vivo erythrocyte Pig-a gene mutation assay - Part 2 - Validation report. Available as Publication 316 at: http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm
10. Miura D, VN Dobrovolsky, T Kimoto, Y Kasahara, RH Heflich (2009) Accumulation and persistence of *Pig-A* mutant peripheral red blood cells following treatment of rats with single and split doses of *N*-ethyl-*N*-nitrosourea. Mutat Res 677:86-92.
11. Phonethepswath S, D Franklin, DK Torous, SM Bryce, JC Bemis, S Raja, S Avlasevich, P Weller, O Hyrien, J Palis, JT MacGregor, SD Dertinger (2010) *Pig-a* mutation: Kinetics in rat erythrocytes following exposure to five prototypical mutagens. Toxicol Sci 114:59-70.
12. Dertinger SD, SL Avlasevich, DK Torous, JC Bemis, S Phonethepswath, C Labash, K Carlson, J Mereness, J Cottom, J Palis, JT MacGregor (2014) Persistence of cisplatin-induced mutagenicity in hematopoietic stem cells: Implications for secondary cancer risk following chemotherapy. Toxicol Sci 140:307-314.
13. Elhajouji A, TT Hove, O O'Connell, H Martus, SD Dertinger (2020) *Pig-a* gene mutation assay study

© OECD, (2022)

design: Critical assessment of 3- versus 28-day repeat-dose treatment schedules. Mutagenesis 35:349-358. Available at: https://doi.org/10.1093/mutage/geaa014.

14. OECD (Organisation for Economic Cooperation and Development) (2008) OECD Guidelines for the testing of chemicals: Repeated dose 28-day oral toxicity study in rodents. Test Guideline 407. Available at: https://www.oecd-ilibrary.org/docserver/9789264070684-en.pdf?expires=1619269423&id=id&accname=guest&checksum=BFC32AEBFBFAF108ABF2818FE30770AF.

15. OECD (Organisation for Economic Cooperation and Development) (2018) OECD Guidelines on the testing of chemicals: 28-day (subacute) inhalation toxicity study. Test Guideline 412. Available at: https://www.oecd-ilibrary.org/docserver/9789264070783-en.pdf?expires=1591459924&id=id&accname=guest&checksum=5CF6E4990007E8850A376A91B0C20687.

16. OECD (Organisation for Economic Cooperation and Development) (2020) OECD Guideline for testing of chemicals: Transgenic rodent somatic and germ cell gene mutation assays. Test Guideline 488. Available at: https://www.oecd-ilibrary.org/docserver/9789264203907-en.pdf?expires=1608906827&id=id&accname=guest&checksum=F82C7AA8D9B443320C9BD6DF8E541269.

17. OECD (Organisation for Economic Cooperation and Development) (2018) OECD Guideline for the testing of chemicals: Repeated dose 90-day oral toxicity study in rodents. Test Guideline 408. Available at: https://www.oecd-ilibrary.org/docserver/9789264070707-en.pdf?expires=1619554311&id=id&accname=guest&checksum=E4654CBAD5DF751E5E9D8D54540D6AD0.

18. OECD (Organisation for Economic Cooperation and Development) (2018) OECD Guidelines for the testing of chemicals: 90-day (subchronic) inhalation toxicity study. Test Guideline 413. Available at: https://www.oecd-ilibrary.org/docserver/9789264070806-en.pdf?expires=1591459983&id=id&accname=guest&checksum=81BC208072B006A03D81A6E5DE579A66.

19. Dertinger SD, S Phonethepswath, D Franklin, P Weller, DK Torous, SM Bryce, S Avlasevich, JC Bemis, O Hyrien, J Palis, JT MacGregor (2010) Integration of mutation and chromosome damage endpoints into 28-day repeat dose toxicology studies. Toxicol Sci 115:401-411.

20. Kirkland D, DD Levy, MJ LeBaron, MJ Aardema, C Beevers, J Bhalli, GR Douglas, PA Escobar, CS Farabaugh, M Guerard, GE Johnson, R Kulkarni, F Le Curieux, AS Long, J Lott, DP Lovell, M Luijten, F Marchetti, JJ Nicolette, S Pfuhler, DJ Roberts, LF Stankowski Jr, V Thybaud, SK Weiner, A Williams, KL Witt, R Young (2019) A comparison of transgenic rodent mutation and *in vivo* comet assay responses for 91 chemicals. Mutat Res 839:21-35.

21. Kimoto T, K Suzuki, Xm Kobayashi, VN Dobrovolsky, RH Heflich, D Miura, Y Kasahara (2011) Manifestation of *Pig-a* mutant bone marrow erythroids and peripheral blood erythrocytes in mice treated with *N*-ethyl-*N*-nitrosourea: Direct sequencing of *Pig-a* cDNA from bone marrow cells negative for GPI-anchored protein expression. Mutat Res 723:36-42.

22. Revollo JR, MG Pearce, A Dad, DM Petibone, TW Robison, D Roberts, VN Dobrovolsky (2018) Analysis of mutation in the rat *Pig-a* assay: I) Studies with bone marrow erythroid cells. Environ Mol Mutagenesis 59:722-732.

23. Revollo JR, A Dad, MG Pearce, RA Mittelstaedt, TW Robison, VN Dobrovolsky (2019) *Pig-a* mutations in bone marrow erythroblasts of rats treated with 7,12-dimethyl-benz[*a*]anthracene. Mutat Res 848:503106.

© OECD, (2022)

24. Revollo JR, A Dad, MG Pearce, RA Mittelstaedt, A Casildo, RG Lapidus, TW Robison, VN Dobrovolsky (2020) CD59-deficient bone marrow erythroid cells from rats treated with procarbazine and propyl-nitrosourea have mutations in the *Pig-a* gene. Environ Mol Mutagenesis 61:797-806.

25. Olsen A-K, SD Dertinger. CT Krüger, DM Eide, C Instanes, G Brunborg, A Hartwig, A Graupner (2017) The *Pig-a* gene mutation assay in mice and human cells: A review. Basic Clin Pharmacol Toxicol 121:78-92. Available at: https://doi.org/10.1111/bcpt.12806.

26. Bemis JC, RH Heflich (2019) *In vitro* mammalian cell mutation assays based on the *Pig-a* gene: A report of the 7th International Workshop on Genotoxicity Testing (IWGT) workgroup. Mutat Res 847:403028. Available at: https://10.1016/j.mrgentox.2019.03.001 847:403028.

27. OECD (Organisation for Economic Cooperation and Development) (2009) Preliminary review of OECD Test Guidelines for their applicability to manufactured nanomaterials. Available at: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2009)21.

28. OECD (Organisation for Economic Cooperation and Development) (2014) Genotoxicity of manufactured nanomaterials: Report of the OECD Expert Meeting. Series on the Safety of Manufactured Nanomaterials, No. 43. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)34&doclanguage=en.

29. Elespuru R, S Pfuhler, MJ Aardema, T Chen, SH Doak, A Doherty, CS Farabaugh, J Kenny, M Manjanatha, B Mahadevan, MM Moore, G Ouédraogo, LF Stankowski Jr, JY Tanir (2018) Genotoxicity assessment of nanomaterials: Recommendations on best practices, assays, and methods. Toxicol Sci 164:391-416.

30. Kinoshita T (2014) Biosynthesis and deficiencies of glycosylphosphatidylinositol. Proc Jpn Acad, Ser B 90:130-143.

31. Dad A, VN Dobrovolsky, RH Heflich, JR Revollo (2020) Molecular analysis of GPI-anchor biosynthesis pathway genes in rat strains used for the *Pig-a* gene mutation assay. Mutat Res 858-860:503256.

32. Giknis MLA, CB Clifford (2006) Clinical laboratory parameters for Crl:CD(SD) rats. Charles River Laboratories. pp 1-18. Available at: https://www.criver.com/sites/default/files/resources/rm_rm_r_clinical_parameters_cd_rat_06.pdf.

33. Miura D, VN Dobrovolsky, Y Kasahara, Y Katsuura, RH Heflich (2008) Development of an in vivo gene mutation assay using the endogenous *Pig-A* gene: I. Flow cytometric detection of CD59-negative peripheral red blood cells and CD48-negative spleen T-cells from the rat. Environ Mol Mutagenesis 49:614-621.

34. Bryce SM, JC Bemis, SD Dertinger (2008) In vivo mutation assay based on the endogenous *Pig-a* locus. Environ Mol Mutagenesis 49:256-264.

35. Phonethepswath S, SM Bryce, JC Bemis, SD Dertinger (2008) Erythrocyte-based *Pig-a* gene mutation assay: Demonstration of cross-species potential. Mutat Res 657:122-126.

36. Dertinger SD, S Phonethepswath, SL Avlasevich, DK Torous, J Mereness, SM Bryce, JC Bemis, S Bell, P Weller, JT MacGregor (2012) Efficient monitoring of *in vivo Pig-a* gene mutation and chromosomal damage: Summary of 7 published studies and results from 11 new reference compounds. Toxicol Sci 130:328-348.

37. Kenyon MO, SL Coffing, JI Ackerman, WC Gunther, SD Dertinger, K Criswell, KL Dobo (2015) Compensatory erythropoiesis has no impact on the outcome of the *in vivo Pig-a* mutation assay in rats following treatment with the haemolytic agent 2-butoxyethanol. Mutagenesis 30:325-334.

© OECD, (2022)

38. Raschke M, B-W Igl, J Kenny, J Collins, SD Dertinger, C Labash, JA Bhalli, CCM Tebbe, KM McNeil, A Sutter (2016) In vivo *Pig-a* gene mutation assay: Guidance for 3Rs-friendly implementation. Environ Mol Mutagenesis 57:678-686.

39. Labash C, SL Avlasevich, K Carlson, A Berg, DK Torous, SM Bryce, JC Bemis, JT MacGregor, SD Dertinger (2016) Mouse *Pig-a* and micronucleus assays respond to *N*-ethyl-*N*-nitrosourea, benzo[*a*]pyrene, and ethyl carbamate, but not pyrene or methyl carbamate. Environ Mol Mutagenesis 57:28-40. Available at: https://doi. org/10.2903/j.efsa.2017.5113.

40. EU (European Union) (2010) Directive 2010/63/EU of the European Parliament and the Council of 22 September 2010 on the protection of animals used for scientific purposes. Available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:en:PDF.

41. Dertinger SD, S Phonethepswath, P Weller, S Avlasevich, DK Torous, JA Mereness, SM Bryce, JC Bemis, S Bell, S Portugal, M Aylott, JT MacGregor (2011b) Interlaboratory *Pig-a* gene mutation assay trial: Studies of 1,3-propane sultone with immunomagnetic enrichment of mutant erythrocytes. Environ Mol Mutagenesis 52:748-755.

42. Mittelstaedt RA, VN Dobrovolsky, JR Revollo, MG Pearce, Y Wang, A Dad, PB McKinzie, H Rosenfeldt, B Yucesoy, R Yeager, S-C Hu, Y Tang, S Min, H-K Kang, D-J Yang, M Basavarajappa, RH Heflich (2019) Evaluation of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) mutagenicity using *in vitro* and *in vivo Pig-a* assays. Mutat Res 837:65-72.

43. Dertinger SD, SL Avlasevich, DK Torous, P Singh, S Khanal, C Kirby, A Drake, JT MacGregor, JC Bemis (2019) 3Rs friendly study designs facilitate rat liver and blood micronucleus assays and *Pig-a* gene mutation assessment: Proof-of-concept with 13 reference chemicals. Environ Mol Mutagenesis 60:704-739.

44. OECD (Organisation for Economic Cooperation and Development) (2000) Guidance document on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluation. Available at: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2000)7&doclanguage=en.

45. OECD (Organisation for Economic Cooperation and Development) (2016b) OECD Guideline for the testing of chemicals: *In vivo* mammalian alkaline comet assay. Test Guideline 489. Available at: https://www.oecd-ilibrary.org/docserver/9789264264885-en.pdf?expires=1608823507&id=id&accname=guest&checksum=A9E2E3FB6FC0EB098C9099D235C61839.

46. US Federal Register (1996) International Conference on Harmonisation; Guidance on specific aspects of regulatory genotoxicity tests for pharmaceuticals. Vol 61 (no. 80):18198-18202. Department of Health and Human Services, Food and Drug administration. Available at: https://www.govinfo.gov/content/pkg/FR-1996-04-24/html/96-10021.htm.

47. Marenzana M, TR Arnett (2013) The key role of the blood supply to bone. Bone Research 3:203-215.

48. Grüneboom A, I Hawwari, D Weidner, S Culemann et al. (2019) A network of trans-cortical capillaries as mainstay for blood circulation in long bones. Nature Metabolism 1:236-250.

49. Kirkland D, Y Uno, M Luijten, C Beevers, J van Benthem, B Burlinson, S Dertinger, GR Douglas, S Hamada, K Horibata, DP Lovell, M Manjanatha, H-J Martus, N Mei, T Morita, W Ohyama, A Williams (2019) *In vivo* genotoxicity testing strategies: Report from the 7th International Workshop on Genotoxicity Testing (IWGT). Mutat Res 847:403035. Available at: https://10.1016/j.mrgentox.2019.03.008.

50. Avlasevich SL, DK Torous, JC Bemis, JA Bhalli, CC Tebbe, J Noteboom, D Thomas, DJ Roberts, M

© OECD, (2022)

Barragato, B Schneider, J Prattico, M Richardson, BB Gollapudi, SD Dertinger (2019) Suitability of long-term frozen rat blood samples for the interrogation of *Pig-a* gene mutation by flow cytometry. Environ Mol Mutagenesis 60:47-55.

51. Dertinger SD, SM Bryce, S Phonethepswath, SL Avlasevich (2011) When pigs fly: Immunomagnetic separation facilitates rapid determination of *Pig-a* mutant frequency by flow cytometric analysis. Mutat Res 721:163-170.

52. Kimoto T, S Chikura, K Suzuki, Xm Kobayashi, Y Itano, K Horibata, M Honma, VN Dobrovolsky, RH Heflich, D Miura, Y Kasahara (2011) Further development of the rat *Pig-a* mutation assay: Measuring rat *Pig-a* mutant bone marrow erythroids and a high throughput assay for mutant peripheral blood reticulocytes. Environ Mol Mutagenesis 52:774-783.

53. OECD (Organisation for Economic Cooperation and Development). (2016) Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. Series on Testing and Assessment, No. 238 – 2nd edition. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV-JM-MONO(2016)33/rev1&doclanguage=en.

54. EFSA (European Food Safety Authority) Scientific Committee, A Hardy, D Benford, T Halldorsson, M Jeger, HK Knutsen, S More, H Naegeli, H Noteborn, C Ockleford, A Ricci, G Rychen, V Silano, R Solecki, D Turck, M Younes, G Aquilina, R Crebelli, R Gürtler, KI Hirsch-Ernst, P Mosesso, E Nielsen, J van Benthem, M Carfi, N Georgiadis, D Maurici, J Parra Morte, J Schlatter (2017) Scientific Opinion: Clarification of some aspects related to genotoxicity assessment. EFSA Journal 15:5113. Available at: https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2017.5113.

55. Shin HK, YM Kang, KT No (2017) Predicting ADME Properties of Chemicals. In: Leszczynski J, A Kaczmarek-Kedziera, T Puzyn, MG Papadopoulos, H Reis, MK Shukla (eds). Handbook of Computational Chemistry. Springer, Cham. Available at: https://doi.org/10.1007/978-3-319-27282-5_59.

56. OECD (Organisation for Economic Cooperation and Development) (2016) OECD Guideline for the testing of chemicals: Mammalian erythrocyte micronucleus test. Test Guideline 474. Available at: https://www.oecd-ilibrary.org/docserver/9789264264762-en.pdf?expires=1608922977&id=id&accname=guest&checksum=11DD1C964576589D2EB86B428141ECF6.

57. Bretz F, LA Hothorn (2003) Statistical analysis of monotone or non-monotone dose-response data from in vitro toxicological assays. ATLA-Altern Lab Animal 31(Suppl 1):81-96.

58. Labash C, SL Avlasevich, K Carlson, DK Torous, A Berg, JC Bemis, JT MacGregor, SD Dertinger (2015) Comparison of male versus female responses in the *Pig-a* mutation assay. Mutagenesis 30:349-357.

59. Ryan TP (2011) Statistical Methods for Quality Improvement, 3rd Ed., John Wiley and Sons, NY, NY. pp 1-704.

60. Hayashi M, K Dearfield, P Kasper, D Lovell, H-J Martus, V Thybaud (2011) Compilation and use of genetic toxicity historical control data. Mutat Res 723:87-90.

61. Vardeman SB (1992) What about the other intervals? The American Statistician 46:193-197.

62. InfinityQS International (2014) A practical guide to selecting the right control chart. pp 1-21. Available at: http://www.infinityqs.com/sites/infinityqs.com/files/files/PDFs/InfinityQS_Practical_Guide_to_Selecting_the_Right_Control_Chart_Oct2013.pdf.

© OECD, (2022)

63. Kluxen FM, K Weber, C Strupp, SM Jensen, LA Hothorn, J-C Garcin, T Hofmann (2021) Using historical control data in bioassays for regulatory toxicology. Reg Toxicol Pharmacol 125:105024.

64. Igl B-W, SD Dertinger, VN Dobrovolsky, M Raschke, A Sutter, R Vonk (2018) A statistical approach for analyzing data from the *in vivo Pig-a* gene mutation assay. Mutat Res 831:33-44.

65. Nelson LS (1984) The Shewhart control chart—tests for special causes. Jrnl Qual Tech 16:238-239.

© OECD, (2022)

Annex 1 - Definitions and acronyms

3Rs = Replacement, reduction, and refinement; these principles provide a framework for performing more humane animal research

ADME = absorption, distribution, metabolism, and excretion

Clonal expansion = over-representation of the progeny of a cell in a population, usually due to the accelerated growth of that cell and its progeny relative to other cells in the population

Confidence interval (CI) = a range of values that is likely to include a population value with a certain degree of confidence

Control limit = horizontal line(s) drawn on a statistical control chart; these are investigator-defined and use-case-dependent values; in the field of Quality Control, these values are typically the sample mean ± 3 standard deviations, but some other multiples of the standard deviation can be useful *e.g.*, 2x or 1.96x

Erythrocytes = red blood cells, irrespective of age; for the purposes of this document, this term is interchangeable with RBC

Erythroid mutagen = a chemical capable of altering the DNA sequence of erythroid cells

Flow cytometry = the determination of the characteristics of a population of cells using a flow cytometer

GPI = glycosylphosphatidylinositol, a type of cell membrane anchor that links approximately 150 unique proteins to the exoplasmic surface of mammalian cytoplasmic membranes; *Pig-a* is among the genes necessary to form GPI anchors

Jackpot mutation = a germ cell mutation or more likely a mutation in a bone marrow erythroid precursor cell that results in a mutant frequency that is unusually high and outside the historical negative control distribution

Mutant mimic = erythrocytes that are processed in such a manner that they exhibit the same fluorescence characteristics as *bona fide Pig-a* mutant erythrocytes; these cells are often generated by processing extra blood from a vehicle/solvent control animal and omitting the fluorescent GPI-anchored antibody(s) from the fluorescent antibody labeling protocol; mutant mimics are valuable for guiding instrumentation settings and software/data analysis parameters

MUT RBC = mutant red blood cells; for the purposes of this document, the term describes circulating erythrocytes, irrespective of age, that exhibit the *Pig-a* mutant phenotype (*i.e.*, loss of GPI anchors)

MUT RET = mutant reticulocytes; for the purposes of this document, the term describes the most immature fraction of circulating erythrocytes, identified by their residual RNA and/or presence of the CD71 surface marker, that exhibit the *Pig-a* mutant phenotype (*i.e.*, loss of GPI anchors)

Negative control = in most instances solvent- or vehicle-dosed groups of animals used to identify significant substance-related effects in the *Pig-a* assay. In some instances, *e.g.*, establishing a negative control data base, pre-dose data may substitute for data from solvent- or vehicle-treated animals

Neutral phenotype = a term used to describe mutant cells that are not highly selected for or against *in vivo*, *e.g.*, cells that grow at the same rate as cells from the predominant, wild-type cell population

***Pig-a* gene** = phosphatidylinositol glycan class A gene

Prediction interval = an estimate of an interval in which a future observation will fall, with a certain probability, given what has already been observed

© OECD, (2022)

Promutagen = a chemical requiring structural alteration, usually by enzymatic metabolism, to induce mutations

Quantile = divisions of a probability distribution or frequency distribution into equal, ordered subgroups, for example quartiles or percentiles

RBC = red blood cells; for the purposes of this document, it encompasses all circulating erythrocytes, irrespective of age

RET = reticulocytes (in some contexts called polychromatic erythrocytes), an immature erythrocyte developed from bone marrow erythroblasts that is identified by its residual RNA and/or presence of the CD71 surface marker

Residual distribution = residuals are estimates of experimental error obtained by subtracting the observed responses from the predicted responses; the predicted response is calculated from the chosen model, after all the unknown model parameters have been estimated from the experimental data; examining residual distribution is important because it provides information about model assumptions and appropriateness

Tolerance interval = statistical interval within which, with some confidence level, a specified proportion of a sampled population falls

UVBC = substances of unknown or variable composition, biological materials or complex reaction products

Variance homogeneity = this is an assumption of most parametric tests including the independent samples t-test and ANOVA, and states that all comparison groups have the same variance; the independent samples t-test and ANOVA utilise the *t* and *F* statistics respectively, which are generally robust to violations of the assumption as long as group sizes are equal

© OECD, (2022)

Annex 2 - Combination of micronucleus and comet tests with *Pig-a* assays

1. Combination of the micronucleus and comet assays with Pig-a assays employing a 28-day dosing protocol

Micronucleus and comet tests often are conducted hours to day(s) following short-term dosing protocols (*e.g.*, two or three consecutive days of exposure). However, TG 474 for the *in vivo* micronucleus test (56) and TG 489 for the *in vivo* alkaline comet test (45) also describe integration of these genotoxicity tests into longer, repeat-dosing protocols. Thus these two genotoxicity tests can be combined with *Pig-a* assays that use the recommended 28-day repeat-dosing protocol.

Blood sampling for blood-based assays, such as the *in vivo* micronucleus assay, can be performed early in the *Pig-a* repeat-dose treatment schedule (*e.g.*, on Day 4). In addition, the blood-based micronucleus assay and the comet assay on multiple tissues can be conducted after the 28-day treatment protocol is complete, often at Day 29, with the addition of a dose on Day 29 prior to tissue harvest for conducting the comet assay.

2. Combination of micronucleus and comet assays with Pig-a assays employing a short-term dosing protocol

While it is technically possible to combine the *Pig-a* assay with the micronucleus and comet assay while using a single, short-term treatment protocol for the tests, two important considerations should be addressed. First, when acute dosing schedules are used to generate *Pig-a* data, such studies should be accompanied by scientific justification for measuring the *Pig-a* endpoint as they generally result in the animals being tested with higher daily doses, but a lower total dose of test chemical than with a 28-day repeat-dosing protocol. Such treatments generally result in lower *Pig-a* mutant frequencies than generated by longer-term treatment schedules. Second, the *Pig-a* assay requires an extended period for mutant manifestation - usually a minimum of 2 to 4 weeks (3), which is much longer than the duration of a short-term treatment comet or micronucleus test.

The following are three possible designs for combining micronucleus and comet assays with *Pig-a* assays employing short-term treatments:

1. Use a single set of animals and a single short-term treatment (*e.g.*, 2 or 3 consecutive days of treatment). Using this protocol, combining the *Pig-a* assay with micronucleus and/or comet tests limits the comet and micronucleus tests to blood-based analyses that can be accomplished without euthanizing the animal. Thus, blood-based micronucleus and comet tests are conducted following the short-term treatment, and the animals are held for an additional period (*e.g.*, a total of 28 days) for manifestation of *Pig-a* mutants. Limiting the analysis to blood is problematic for the comet test, which has displayed reduced accuracy for detecting DNA damage in blood (20). Thus, using this protocol, the *Pig-a* assay combines much more readily with short-term blood-based micronucleus tests than it does with short-term comet tests.
2. Incorporate a second round of short-term treatment(s) during the final days of the *Pig-a* mutant manifestation period. The comet and micronucleus tests, as well as the *Pig-a* assay, then can be performed after this second round of short-term exposure(s). Following this protocol, the

© OECD, (2022)

Pig-a assay would mainly be measuring mutations induced by the first round of treatment(s), with the micronucleus and comet tests measuring genotoxicity induced by the second round of treatment(s). Note that if desired, a blood-based micronucleus test also can be performed following the first round of short-term treatment(s).

3. A less 3Rs-friendly alternative consists of a single short-term treatment, with half the animals euthanised shortly thereafter to perform the comet and/or micronucleus tests as described in TG474 and TG 489. In order to perform the *Pig-a* assay, the remaining animals are maintained until later blood collection times, *e.g.*, Day 15 and Day 30 as described in paragraph 44. This ensures that animals are available that have an adequate time for manifestation of the *Pig-a* mutant phenotype.

© OECD, (2022)

Annex 3 - Animal-friendly training exercises for learning techniques used in the *Pig-a* assay

A 3Rs-friendly strategy for learning animal-treatment, blood-collection, blood-processing, and flow-cytometric analysis techniques used in the *Pig-a* assay has been described by (38). For this strategy, a single rodent is exposed to a known, potent mutagenic chemical, *e.g.*, one listed in Table 1. After an appropriate phenotypic expression time that allows elevated MUT RET and MUT RBC frequencies to appear in the peripheral blood circulation, blood from the exposed rodent can be collected along with blood from a sex- and age-matched untreated or vehicle/solvent control animal. A series of mixtures of the two blood samples is made that contains various amounts of blood from the mutagen-exposed animal (*i.e.*, spiked samples). After determining MUT RET and MUT RBC frequencies for the rodent dosed with a mutagen and the negative control rodent, expected (intermediate) frequencies can be calculated for the spiked samples based on the proportion of blood from the rodent dosed with a mutagen added to negative control blood. By conducting reconstruction experiments on 2 or more separate occasions, and with several replicates per spiked sample, staff can be trained on important elements of the test and the proficiency of staff members can be established by demonstrating good agreement between observed and expected mutant cell frequencies. Successful completion of spiking experiments represents a useful, 3Rs-friendly gateway to the laboratory proficiency investigations described in paragraph 21.

© OECD, (2022)

Annex 4 - Statistical analysis and data interpretation

1. Recommendations for evaluating *Pig-a* assay data

There is no single correct method of conducting a statistical analysis, and statisticians can differ in their preferred methodology. A practical approach is to suggest a particular set of statistical analyses as an example of the sort of evaluations that can be carried out. This is the context that three types of analyses described in paragraph 66 of this Test Guideline were presented. It should be made clear that this is not a prescribed method and may not be suitable for all sets of data. It would be acceptable to use an alternative method, especially if the suggested method is not considered suitable. In these cases, the investigators should specify their methods before a study is initiated (i.e., written study or validation plan), and should be prepared to justify their approach using sound statistical arguments.

Note that there is an important fundamental principle that the statistical methods described below should all take into consideration. Specifically, these and other statistical tests assume that the experimental design is based on random sampling, or at least a blocking approach. Randomization and/or blocking represent important methods that help mitigate the influence of factors that may have subtle, or not so subtle, effects on experimental results, but are not the main factors being studied. An illustrative example is blood samples or some other tissue(s) derived from vehicle control animals processed and analysed entirely by one laboratory technician, while another technician is responsible for all these steps for the high dose treatment group. This represents an experimental design asymmetry that should be addressed either through randomization or blocking.

2. Pairwise comparisons

One set of statistical tests is pairwise comparisons of MUT RET/RBC and RET frequencies in the concurrent vehicle/solvent (negative) control group with those measured in the test chemical dose groups. Parametric analyses that use analysis of variance (ANOVA) with an appropriate multiple comparisons test are commonly used, but other methodologies are equally acceptable. Generally, these types of parametric tests should be performed only when assumptions about normality and homogeneity of variance are valid. If normality and/or unequal variance is identified, an appropriate data transformation such as a logarithmic (log_{10}) or rank transform can often be used to fulfill these requirements. Note that if there are animals with 0 (zero) mutant frequency values, a small constant offset value such as 0.1 should be added to every animal's mutant cell frequency before log transformation because this is otherwise not a valid operation. If the transformation does not result in homoscedasticity, weighted (variance-corrected) ANOVA and/or t-tests can be applied or the appropriateness of non-parametric methods can be considered. Finally, when more than one sex is used in a study, factorial design approaches that consider both treatment and sex are generally advantageous. This is described in greater detail below.

2.1 Factorial design

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

The analysis partitions the variability in the dataset into that between the sexes, that between the dose levels, and that related to the interaction between the sexes and the dose levels. Full details of the underlying methodology are available in many standard statistical textbooks and in the 'help' facilities provided with statistical packages.

© OECD, (2022)

The analysis proceeds by first inspecting the sex-by-dose interaction term in the ANOVA table. (Note that statisticians who take a modelling approach such as using General Linear Models may approach the analysis in a different but comparable way but will not necessarily derive the traditional ANOVA table). In the absence of a significant interaction term the combined values across sexes or across dose 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-dose-level variability into contrasts which provide for a test for linear and quadratic contrasts of the responses across the vehicle/solvent concurrent control and test chemical dose levels. When there is a significant sex x dose level interaction this term also can be partitioned into linear x sex and quadratic x sex interaction contrasts. These terms provide tests of whether the dose level 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 pairwise tests of the difference(s) between means. These comparisons could be made between the means for the two sexes and between the means for the different dose levels such as for comparisons with the negative (solvent/vehicle) concurrent control levels. In those cases where there is a significant interaction, comparisons can be made between the means of different dose levels within a sex or between the means of the sexes at the same dose level.

3. Trend test

A second type of analysis described in this Test Guideline is a trend test to identify a dose-response relationship. When conducting trend testing, MUT RET and MUT RBC data should be available from the concurrent negative control group and each of the experimental dose groups. Note that this analysis is not normally applicable for Limit Tests (i.e., single test chemical dose level studies as described in paragraph 51). In employing these analyses, care is needed in interpreting the results of some types of trend tests. For example, a simple linear trend test may fail to detect a trend when the dose-response is non-monotonic. Trend tests capable of detecting non-monotonicity such as the downturn protection test proposed by Bretz and Hothorn (57) may be useful in such cases. A graph that shows the dose response can be helpful in determining what type of trend test is appropriate and/or whether a transformation is needed.

4. Historical negative control data distribution

Distributions of historical negative control data are important for determining the minimum number of cells that should be evaluated for the mutant phenotype, and for assessing assay acceptability. A third important function is determining whether the mean MUT RET and/or MUT RBC frequency of any test-chemical-exposed treatment group exceeds the upper bounds of the historical negative control data distribution. Given their importance, guidance for building historical negative control datasets is provided below.

4.1 Building historical negative control datasets

During the course of the proficiency investigations, each laboratory should establish historical negative control ranges and distributions for MUT RET and MUT RBC frequencies. Negative control data can be obtained from two sources. First, negative control data can be derived from animals assayed as negative controls in conjunction with animals dosed with test chemicals (referred to here as concurrent negative controls). These animals typically will be dosed with the solvent/vehicle alone. Secondly, for laboratories that conduct pre-dosing *Pig-a* assays, negative control data can be obtained from these untreated, naïve animals (referred to as pre-dose control data).

© OECD, (2022)

When first acquiring data for historical negative control MUT RET and MUT RBC frequency distributions, negative controls should be consistent with published data, where they exist (1). Experimental data should continue to be added to achieve statistically robust databases that facilitate assessments of later studies' validity, as well as comparisons of test chemical-exposed animals' MUT RET and MUT RBC frequencies relative to historical negative control distributions.

For these purposes, the laboratory ideally should endeavour to acquire MUT RET and MUT RBC frequency measurements from at least 30 negative control animals from each rodent strain used for testing. These negative control data should be from solvent/vehicle concurrent control animals, and if the data are available, from naïve (pre-dose) animals. Each MUT RET and MUT RBC frequency value should be acquired from an individual animal, thus serial blood samples from the same animal are not acceptable for this purpose. Furthermore, the data should be acquired from at least 3 independent experiments that each use animals generated from different breeding cycles. To the extent possible, the experiments should be conducted under conditions comparable to regulatory studies.

Although no significant differences in negative control mutant frequencies have been detected between animals that differ in age by a few months (3), a historical range based on young adult animal mutant frequencies (*e.g.*, 2-5 months of age) may not be suitable, without further evidence, for evaluating data generated from samples taken from older animals, *e.g.*, 8 or 12 months of age.

Based on the literature, it is acceptable to initially consider sex as having no effect on MUT RET and MUT RBC frequencies of young, healthy rodents (58) (43) (7) and historical negative control frequencies can be constructed using animals of either sex. Laboratories are encouraged, however, to collect negative control data for both sexes of each rodent strain used for testing, and in this manner establish the equivalence or nonequivalence of MUT RET and MUT RBC frequency distributions between the sexes and rodent strains. So long as sex is not found to be a significant factor, a single MUT RET and single (separate) MUT RBC distribution can be assembled. If significant sex differences are found, this would indicate that sex-specific negative control distributions should be constructed.

Laboratories should use quality control methods, such as control charts *[e.g.*, I- and X-bar charts (59)], to identify how variable their data are, and to show that the methodology is 'under control' in their laboratory. Further recommendations on how to build and use the historical data (*i.e.*, criteria for inclusion and exclusion of data in historical databases) and establishing acceptability criteria for a given experiment can be found in the literature [*e.g.*, (60)] and are discussed below.

Any changes to the experimental protocol should be considered in terms of their impact on whether the resulting data remains consistent with the laboratory's existing historical negative control database. Only major inconsistencies should result in the establishment of a new historical negative control database. During the re-establishment, 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.

Negative control data from *Pig-a* assays should consist of the frequency of MUT RET and MUT RBC in each animal. Where data from individual negative control animals fall outside the historical negative control distribution, they may be acceptable for inclusion in the database provided (i) these data are not extreme (*e.g.*, an occasional 'jackpot' mutation which has an unusually high mutant frequency is to be expected and can be excluded), (ii) there is evidence that the test system is 'under control', and (iii) there is no evidence of technical or human error.

© OECD, (2022)

4.2 Characterizing historical negative control distributions

There are several valid approaches for characterizing the distribution of historical negative control data, and each laboratory should use an appropriate method for describing their data. This will generally take the form of calculating an upper and/or lower bound limit that describes the boundary within which the majority of negative control animal values are expected to fall. Factors that should be considered include sample size and whether the data are normally distributed. A brief overview of methods for calculating an upper and/or lower bound limit is provided below. Additional information can be found in (61) (62) and (63).

Inappropriate methods

Range: The range is the difference between the minimum and maximum observed value. The range does not adequately describe the historical negative control distribution for the purpose of establishing useful upper and/or lower bound limits. This is because the range will widen as the number of samples increases, and may depend on two extreme (unusual/outlier) values. A wide range may “reward” poorly performing laboratories.

Confidence interval: A confidence interval is a range of estimates that is likely to include a population value (parameter) such as the mean with a defined (*e.g.*, 95%) degree of confidence. Confidence intervals around a population mean are not useful for adequately describing the historical negative control distribution in the context of establishing useful upper and/or lower bound limits. A main issue with confidence intervals is that they narrow as the sample size increases, reflecting the increased precision of the estimate of the population value (parameter).

Appropriate methods

Control limit: In the field of Quality Control, multiples of the standard deviation, usually the mean plus and minus 3 standard deviations, are used as control limits. These values are lines plotted on a control chart, and may be accompanied by other standard deviation multiples, for instance 2x, which are referred to as warning limits. In conjunction with control charts, control and warning limits are valuable tools for assessing the degree to which a repeated process or test is ‘under control’. Assuming a normal distribution and an ‘under control’ process, 99.73% of the data should fall within 3 standard deviations, and approximately 95% of the data should fall within two standard deviations. Control and/or warning limits therefore represent a useful resource for evaluating historical negative control data, and can provide useful upper and/or lower bound limits that aid in the interpretation of study data. That being said, limits that describe where approximately 95% of normally distributed data fall (± 2 standard deviations) are generally more appropriate than those based on 3 standard deviations. The former is consistent with other OECD Test Guidelines (*e.g.*, OECD TG474), and the latter characterises exceptionally rare/unusual data points, thereby generating intervals that are too wide for the intended purposes. A rule of thumb is that ≥ 25 individual data points are sufficient to derive useful control and warning limits when the process is ‘under control’.

Prediction intervals: Prediction intervals are designed to predict one or several future observation(s) based on existing data. For example, with a 95% prediction interval, a new result would be expected to fall in the range with 95% probability. (Note that, as with that of the confidence interval, this is a simplified definition, the exact definition is more complex.) Non-normal data should be transformed as necessary, with back transformation to original units for reporting and use. Alternately, some computer software programs enable a non-parametric calculation which does not assume a normal distribution. A rule of thumb is that ≥ 30 individual data points work better, and smaller sample sizes should not be used to calculate prediction intervals.

© OECD, (2022)

Tolerance intervals: Tolerance intervals are designed to predict numerous future observations, within user-defined ‘coverage’. For example, 95% of future observations should fall within the 95% tolerance interval. (Again, this is a simplified definition, the exact definition is more complex.) Non-normal data should be transformed as necessary, with back transformation to original units for reporting and use. Alternately, some computer software programs enable a non-parametric calculation which does not assume a normal distribution. Tolerance intervals tend to be wider compared with analogous prediction intervals (since the former is designed to predict a high percentage of future values, while the latter is usually used to predict one or a few new observations). It is also important to recognise that tolerance intervals generally require much larger sample sizes compared with prediction intervals. A rule of thumb is that ≥ 100 data points work better, and smaller sample sizes should not be used to calculate tolerance intervals.

Quantiles: Quantiles are used for summarizing the rank of data points according to their size without assuming any specific probability distribution. Quantiles are widely used in many biomedical applications where non-normality because of outliers and/or skewness is common. They establish intervals based, for instance, on percentiles, to help interpret test results. Confidence intervals for quantiles can be calculated to provide estimates of uncertainty around the quantile measurement. These can help evaluate the quality of the underlying data set. Quantile confidence intervals will be especially wide for the tails of the distribution unless the sample size is large. A rule of thumb is that ≥ 100 individual data points work better, and smaller sample sizes should not be used to calculate quantiles.

Note that while the intervals described above will generally be calculated using individual animal mutant cell frequencies, the primary comparison described by criterion C (paragraph 66) considers where treatment group mean values fall relative to the upper bound value. This is a practical recommendation, acknowledging that as an *in vivo* test system, sample sizes will be larger and historical control distributions more robust when based on individual animal data. Even so, this is not the only comparison that can be made. For instance, there may be times when it would be useful to consider the relationship of individual animal mutant cell frequencies to the historical negative control upper bound limit value.

5. Longitudinal data

Experimental designs that include two or more sampling times may be used to conduct longitudinal analysis (64). These experimental designs may involve dose groups consisting of sub-groups of animals being examined at different time points or, alternatively, animals in the dose groups providing blood samples on multiple days. Designs with both dose and time as factors should be analysed taking into consideration whether the blood samples are from independent animals or are measures from the same animal on different days (*i.e.*, repeated measures design). In general, it is not necessary to develop historical control data for adult animals that differ in age by only a few months (see Annex 4, section 4.1).

There are expectations regarding the manifestation of MUT RET and MUT RBC frequencies with time, *e.g.*, increases in MUT RET frequencies generally precede increases in MUT RBC frequencies, after which increases in both MUT RET and MUT RBCs are typically found (2) (3). These expectations can contribute to a scientific-judgement/weight-of-evidence analysis in establishing the biological relevance as well as to the interpretation of the statistical results of a response. The expectation when measured soon after cessation of dosing in a 28-day study is that a positive response will show elevated frequencies of both MUT RET and MUT RBC; only MUT RET frequencies may be elevated if the sampling was conducted during the dosing period (1) (3).

6. Interpretation of study results

When assessing *Pig-a* results, there should be some assurance that the historical negative control database is of sufficient quality to provide a reasonable assessment of those responses that exceed its

© OECD, (2022)

distribution bounds. One suggested method for assessing the quality of the historical control database that form the basis of lower and/or upper bound limit calculations is the use of control charts in conjunction with Nelson rules (65). Even so, the quality of the historical control data should ultimately be evaluated in terms of consistency with published values, and consistency over multiple studies. Beyond providing evidence of an adequate historical negative control database, the current study should be deemed valid. This includes, in part, mean concurrent vehicle/solvent control treatment group MUT RET/RBC frequencies that are consistent with the historical negative control distribution. Other factors such as the number of animals evaluated per group and instances of zeros in the dataset should be consistent with the guidance given in the Test Guideline.

The analytical approaches described above are regarded as key tools for evaluating whether the test substance induces increases in MUT RET/RBC frequencies. Clearly positive (mutagenic) study results are indicated when the aforementioned three criteria (significant pairwise comparison, significant trend increase, test chemical response greater than the historical negative control distribution; see paragraph 66) are all met. Clearly negative study results are indicated when none of these criteria is met. Scientific judgement will be essential in those cases where only some of the three criteria are met. This paradigm is reinforced by an expert OECD genotoxicity working group that concluded "…data should be interpreted based both on statistics and biological relevance" (53).

In some cases, even after applying scientific judgement and possibly evaluating additional data, it will not be possible to classify a response as positive or negative. In these cases, the response is equivocal and further testing may be required to resolve the mutagenicity of the test substance. This is not as straight-forward as conducting statistical tests and referring to a rigid rubric to make final judgements, but it is considered the best scientific approach according to the aforementioned expert working group (53).

© OECD, (2022)

Annex 5 - An alternative to mutant mimics for establishing flow cytometer settings for conducting the *Pig-a* assay

An alternative to mutant mimics is to use blood samples previously collected from animals with elevated *Pig-a* mutant frequencies (such as from positive controls, if used). Such samples can be stored and used as flow cytometry standards to identify the light scatter and fluorescence characteristics of mutant versus wild-type RET and RBC. In these cases, it is usually sufficient to include 1 to 3 such blood samples at each time point studied. Examples of test chemicals that will produce appropriate increased mutant frequencies are listed in Table 1. Note that it is not necessary to administer the substance using the same route of administration as the test chemical, and the dosing and blood sampling schedule can be different, as long as samples with elevated MUT RBC and MUT RET frequencies are available at the time when experimental analyses are conducted. When blood sample(s) with elevated mutant frequencies come from previously treated rodents, blood should be stored appropriately (paragraphs 57-58). When these samples are being used in place of mutant mimics, the blood sample(s) should demonstrate levels of MUT RET and MUT RBC that are elevated sufficiently to establish the fluorescence characteristics of mutant phenotype cells. For this purpose, it is ideal for the mutant frequency in these two cell populations be at least 100 mutant cells per million erythrocytes.

© OECD, (2022)