

II

(Non-legislative acts)

REGULATIONS

COMMISSION REGULATION (EU) 2019/1390

of 31 July 2019

amending, for the purpose of its adaptation to technical progress, the Annex to Regulation (EC) No 440/2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)

(Text with EEA relevance)

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC ⁽¹⁾, and in particular Article 13(2) thereof,

Whereas:

- (1) Commission Regulation (EC) No 440/2008 ⁽²⁾ contains the test methods for the purposes of the determination of the physicochemical properties, toxicity and ecotoxicity of chemicals to be applied for the purposes of Regulation (EC) No 1907/2006.
- (2) The Organisation for Economic Cooperation and Development (OECD) develops harmonised and internationally agreed test guidelines for the testing of chemicals for regulatory purposes. The OECD regularly issues new and revised test guidelines, taking account of scientific progress in this area.
- (3) In order to take into account technical progress and, whenever possible, to reduce the number of animals used for experimental purposes in accordance with Article 13(2) of Regulation (EC) No 1907/2006, following the adoption of relevant OECD test guidelines, two new test methods for the assessment of ecotoxicity and nine new test methods for the determination of toxicity to human health should be laid down and seven test methods should be updated. Eleven of those test methods relate to in vitro tests for skin and eye irritation/corrosion, skin sensitisation, genotoxicity and endocrine effects. Stakeholders have been consulted on the proposed amendment.

⁽¹⁾ OJ L 396, 30.12.2006, p. 1.

⁽²⁾ Commission Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) (OJ L 142, 31.5.2008, p. 1).

- (4) Regulation (EC) No 440/2008 should therefore be amended accordingly.
- (5) The measures provided for in this Regulation are in accordance with the opinion of the Committee established under Article 133 of Regulation (EC) No 1907/2006,

HAS ADOPTED THIS REGULATION:

Article 1

The Annex to Regulation (EC) No 440/2008 is amended in accordance with the Annex to this Regulation.

Article 2

This Regulation shall enter into force on the twentieth day following that of its publication in the *Official Journal of the European Union*.

This Regulation shall be binding in its entirety and directly applicable in all Member States.

Done at Brussels, 31 July 2019.

For the Commission
The President
Jean-Claude JUNCKER

ANNEX

The Annex to Regulation (EC) No 440/2008 is amended as follows:

- (1) In part B, Chapter B.4 is replaced by the following:

"B.4 ACUTE DERMAL IRRITATION/CORROSION

INTRODUCTION

1. This test method is equivalent to OECD test guideline (TG) 404 (2015). OECD guidelines for testing of Chemicals are periodically reviewed to ensure that they reflect the best available science. In the review of OECD TG 404, special attention was given to possible improvements in relation to animal welfare concerns and to the evaluation of all existing information on the test chemical in order to avoid unnecessary testing in laboratory animals. The updated version of OECD TG 404 (originally adopted in 1981, revised in 1992, 2002 and 2015) includes reference to the Guidance Document on Integrated Approaches to Testing and Assessment (IATA) for Skin Irritation/Corrosion (1), proposing a modular approach for skin irritation and skin corrosion testing. The IATA describes several modules which group information sources and analysis tools, and (i) provides guidance on how to integrate and use existing testing and non-testing data for the assessment of the skin irritation and skin corrosion potentials of chemicals and (ii) proposes an approach when further testing is needed (1). In addition, where needed, the successive, instead of simultaneous, application of the three test patches to the animal in the initial *in vivo* test is recommended in that Guideline.
2. Definitions of dermal irritation and corrosion are set out in the Appendix to this test method.

INITIAL CONSIDERATIONS

3. In the interest of both sound science and animal welfare, *in vivo* testing should not be undertaken until all available data relevant to the potential dermal corrosivity/irritation of the test chemical have been evaluated in a weight-of-the-evidence (WoE) analysis as presented in the Guidance Document on Integrated Approaches to Testing and Assessment for Skin Corrosion and Irritation, i.e. over the three Parts of this guidance and their corresponding modules (1). Briefly, under Part 1 existing data is addressed over seven modules covering human data, *in vivo* data, *in vitro* data, physico-chemical properties data (e.g. pH, in particular strong acidity or alkalinity) and non-testing methods. Under Part 2, WoE analysis is performed. If this WoE is still inconclusive, Part 3 should be conducted with additional testing, starting with *in vitro* methods, and *in vivo* testing is used as last resort. This analysis should therefore decrease the need for *in vivo* testing for dermal corrosivity/irritation of test chemicals for which sufficient evidence already exists from other studies as to those two endpoints.

PRINCIPLE OF THE IN VIVO TEST

4. The test chemical to be tested is applied in a single dose to the skin of an experimental animal; untreated skin areas of the test animal serve as the control. The degree of irritation/corrosion is read and scored at specified intervals and is further described in order to provide a complete evaluation of the effects. The duration of the study should be sufficient to evaluate the reversibility or irreversibility of the effects observed.
5. Animals showing continuing signs of severe distress and/or pain at any stage of the test should be humanely killed, and the test chemical assessed accordingly. Criteria for making the decision to humanely kill moribund and severely suffering animals are the subject of a separate Guidance Document (2).

PREPARATIONS FOR THE IN VIVO TEST

Selection of animal species

6. The albino rabbit is the preferable laboratory animal, and healthy young adult rabbits are used. A rationale for using other species should be provided.

Preparation of the animals

7. Approximately 24 hours before the test, fur should be removed by closely clipping the dorsal area of the trunk of the animals. Care should be taken to avoid abrading the skin, and only animals with healthy, intact skin should be used.
8. Some strains of rabbit have dense patches of hair that are more prominent at certain times of the year. Such areas of dense hair growth should not be used as test sites.

Housing and feeding conditions

9. Animals should be individually housed. The temperature of the experimental animal room should be 20 °C (± 3 °C) for rabbits. Although the relative humidity should be at least 30 % and preferably not exceed 70 %, other than during room cleaning, the aim should be 50-60 %. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unrestricted supply of drinking water

TEST PROCEDURE

Application of the test chemical

10. The test chemical should be applied to a small area (approximately 6 cm²) of skin and covered with a gauze patch, which is held in place with non-irritating tape. In cases in which direct application is not possible (e.g. liquids or some pastes), the test chemical should first be applied to the gauze patch, which is then applied to the skin. The patch should be loosely held in contact with the skin by means of a suitable semi-occlusive dressing for the duration of the exposure period. If the test chemical is applied to the patch, it should be attached to the skin in such a manner that there is good contact and uniform distribution of the test chemical on the skin. Access by the animal to the patch and ingestion or inhalation of the test chemical should be prevented.
11. Liquid test chemicals are generally used undiluted. When testing solids (which may be pulverised, if considered necessary), the test chemical should be moistened with the smallest amount of water (or, where necessary, of another suitable vehicle) sufficient to ensure good skin contact. When vehicles other than water are used, the potential influence of the vehicle on irritation of the skin by the test chemical should be minimal, if any.
12. At the end of the exposure period, which is normally 4 hours, residual test chemical should be removed, where practicable, using water or an appropriate solvent without altering the existing response or the integrity of the epidermis.

Dose level

13. A dose of 0,5 ml of liquid or 0,5 g of solid or paste is applied to the test site.

Initial test (In vivo dermal irritation/corrosion test using one animal)

14. When a test chemical has been judged to be corrosive, irritant or non-classified on the basis of a weight of evidence analyses or of previous *in vitro* testing, further *in vivo* testing is normally not necessary. However, in the cases where additional data are felt warranted, the *in vivo* test is performed initially using one animal and applying the following approach. Up to three test patches are applied sequentially to the animal. The first patch is removed after three minutes. If no serious skin reaction is observed, a second patch is applied at a different site and removed after one hour. If the observations at this stage indicate that exposure can humanely be allowed to extend to four hours, a third patch is applied and removed after four hours, and the response is graded.
15. If a corrosive effect is observed after any of the three sequential exposures, the test is immediately terminated. If a corrosive effect is not observed after the last patch is removed, the animal is observed for 14 days, unless corrosion develops at an earlier time point.
16. In those cases in which the test chemical is not expected to produce corrosion but may be irritating, a single patch should be applied to one animal for four hours.

Confirmatory test (In vivo dermal irritation test with additional animals)

17. If a corrosive effect is not observed in the initial test, the irritant or negative response should be confirmed using up to two additional animals, each with one patch, for an exposure period of four hours. If an irritant effect is observed in the initial test, the confirmatory test may be conducted in a sequential manner, or by exposing two additional animals simultaneously. In the exceptional case, in which the initial test is not conducted, two or three animals may be treated with a single patch, which is removed after four hours. When two animals are used, if both exhibit the same response, no further testing is needed. Otherwise, the third animal is also tested. Equivocal responses may need to be evaluated using additional animals.

Observation period

18. The duration of the observation period should be sufficient to evaluate fully the reversibility of the effects observed. However, the experiment should be terminated at any time that the animal shows continuing signs of severe pain or distress. To determine the reversibility of effects, the animals should be observed up to 14 days after removal of the patches. If reversibility is seen before 14 days, the experiment should be terminated at that time.

Clinical observations and grading of skin reactions

19. All animals should be examined for signs of erythema and oedema, and the responses scored at 60 minutes, and then at 24, 48 and 72 hours after patch removal. For the initial test in one animal, the test site is also examined immediately after the patch has been removed. Dermal reactions are graded and recorded according to the grades in the Table below. If there is damage to skin which cannot be identified as irritation or corrosion at 72 hours, observations may be needed until day 14 to determine the reversibility of the effects. In addition to the observation of irritation, all local toxic effects, such as defatting of the skin, and any systemic adverse effects (e.g. effects on clinical signs of toxicity and body weight), should be fully described and recorded. Histopathological examination should be considered to clarify equivocal responses.
20. The grading of skin responses is necessarily subjective. To promote harmonisation in grading of skin response and to assist testing laboratories and those involved in making and interpreting the observations, the personnel performing the observations need to be adequately trained in the scoring system used (see Table below). An illustrated guide for grading skin irritation and other lesions could be helpful (3).

DATA AND REPORTING

21. Study results should be summarised in tabular form in the final test report and should cover all items listed in paragraph 24.

Evaluation of results

22. The dermal irritation scores should be evaluated in conjunction with the nature and severity of lesions, and their reversibility or lack of reversibility. The individual scores do not represent an absolute standard for the irritant properties of a material, as other effects of the test material are also evaluated. Instead, individual scores should be viewed as reference values, which need to be evaluated in combination with all other observations from the study.
23. Reversibility of dermal lesions should be considered in evaluating irritant responses. When responses such as alopecia (limited area), hyperkeratosis, hyperplasia and scaling, persist to the end of the 14-day observation period, the test chemical should be considered an irritant.

Test report

24. The test report must include the following information:

Rationale for in vivo testing:

- Weight-of-evidence analysis of pre-existing test data, including results from sequential testing strategy;
- Description of relevant data available from prior testing;
- Data derived at each stage of testing strategy;
- Description of *in vitro* tests performed, including details of procedures, results obtained with test/reference substances;
- Weight-of-the-evidence analysis for performing *in vivo* study.

Test chemical:

- Mono-constituent substance: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc;
- Multi-constituent substance, mixture and substances of unknown or variable composition, complex reaction products or biological materials (UVCB): characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physico-chemical properties of the constituents;
- Physical appearance, water solubility, and additional relevant physico-chemical properties;
- Source, lot number if available;
- Treatment of the test chemical/control substance prior to testing, if applicable (e.g. warming, grinding);

- Stability of the test chemical, limit date for use, or date for re-analysis if known;
- Storage conditions.

Vehicle:

- Identification, concentration (where appropriate), volume used;
- Justification for choice of vehicle.

Test animal(s):

- Species/strain used, rationale for using animal(s) other than albino rabbit;
- Number of animal(s) of each sex;
- Individual animal weight(s) at start and conclusion of test;
- Age at start of study;
- Source of animal(s), housing conditions, diet, etc.

Test conditions:

- Technique of patch site preparation;
- Details of patch materials used and patching technique;
- Details of test chemical preparation, application, and removal.

Results:

- Tabulation of irritation/corrosion response scores for each animal at all time points measured;
- Descriptions of all lesions observed;
- Narrative description of nature and degree of irritation or corrosion observed, and any histopathological findings;
- Description of other adverse local (e.g. defatting of skin) and systemic effects in addition to dermal irritation or corrosion.

Discussion of results

Conclusions

LITERATURE

- (1) OECD (2014). Guidance document on Integrated Approaches to Testing and Assessment for Skin Irritation/Corrosion. Environmental Health and Safety Publications, Series on Testing and Assessment, (No 203), Organisation for Economic Cooperation and Development, Paris.
- (2) OECD (1998) Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances, as endorsed by the 28th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, November 1998.
- (3) OECD (2000). Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Environmental Health and Safety Publications, Series on Testing and Assessment (No 19), Organisation for Economic Cooperation and Development, Paris.

Table

Grading of Skin Reactions**Erythema and Eschar Formation**

No erythema	0
Very slight erythema (barely perceptible)	1
Well defined erythema.....	2
Moderate to severe erythema	3
Severe erythema (beef redness) to eschar formation preventing grading of erythema.....	4

Maximum possible: 4

Oedema Formation

No oedema.....	0
Very slight oedema (barely perceptible).....	1
Slight oedema (edges of area well defined by definite raising).....	2
Moderate oedema (raised approximately 1 mm).....	3
Severe oedema (raised more than 1 mm and extending beyond area of exposure).....	4

Maximum possible: 4

Histopathological examination may be carried out to clarify equivocal responses.

Appendix

DEFINITIONS

Chemical is a substance or a mixture.

Dermal irritation is the production of reversible damage of the skin following the application of a test chemical for up to 4 hours.

Dermal corrosion is the production of irreversible damage of the skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discolouration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions.

Test chemical is any substance or mixture tested using this test method"

- (2) In Part B, Chapter B.17 is replaced by the following:

"B.17 *IN VITRO* MAMMALIAN CELL GENE MUTATION TESTS USING THE HPRT AND XPRT GENES

INTRODUCTION

1. This test method (TM) is equivalent to the OECD test guideline 476 (2016). Test methods are periodically reviewed in the light of scientific progress, changing regulatory needs and animal welfare. This current revised version of TM B.17 reflects nearly thirty years of experience with this test and also results from the development of a separate new method dedicated to *in vitro* mammalian cell gene mutation tests using the thymidine kinase gene. TM B.17 is part of a series of test methods on genetic toxicology. A document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to genetic toxicity OECD test guidelines has been developed by OECD (1).
2. The purpose of the *in vitro* mammalian cell gene mutation test is to detect gene mutations induced by chemicals. The cell lines used in these tests measure forward mutations in reporter genes, specifically the endogenous hypoxanthine-guanine phosphoribosyl transferase gene (Hprt in rodent cells, HPRT in human cells; collectively referred to as the Hprt gene and HPRT test in this test method), and the xanthine-guanine phosphoribosyl transferase transgene (gpt) (referred to as the XPRT test). The HPRT and XPRT mutation tests detect different spectra of genetic events. In addition to the mutational events detected by the HPRT test (e.g. base pair substitutions, frameshifts, small deletions and insertions) the autosomal location of the gpt transgene may allow the detection of mutations resulting from large deletions and possibly mitotic recombination not detected by the HPRT test because the Hprt gene is located on the X-chromosome (2) (3) (4) (5) (6) (7). The XPRT is currently less widely used than the HPRT test for regulatory purposes.
3. Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

4. Tests conducted *in vitro* generally require the use of an exogenous source of metabolic activation. The exogenous metabolic activation system does not entirely mimic *in vivo* conditions.
5. Care should be taken to avoid conditions that would lead to artefactual positive results, (i.e. possible interaction with the test system), not caused by direct interaction between the test chemicals and the genetic material of the cell; such conditions include changes in pH or osmolality (8) (9) (10), interaction with the medium components (11) (12), or excessive levels of cytotoxicity (13). Cytotoxicity exceeding the recommended top cytotoxicity levels as defined in paragraph 19 is considered excessive for the HPRT test.
6. Before use of the test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed when there is a regulatory requirement for testing of the mixture.

PRINCIPLE OF THE TEST

7. Mutant cells deficient in Hprt enzyme activity in the HPRT test or xprt enzyme activity in the XPRT test are resistant to the cytostatic effects of the purine analogue 6-thioguanine (TG). The Hprt (in the HPRT test) or gpt (in XPRT test) proficient cells are sensitive to TG, which causes the inhibition of cellular metabolism and halts further cell division. Thus, mutant cells are able to proliferate in the presence of TG, whereas normal cells, which contain the Hprt (in the HPRT test) or gpt (in XPRT test) enzyme, are not.

8. Cells in suspension or monolayer cultures are exposed to the test chemical, both with and without an exogenous source of metabolic activation (see paragraph 14), for a suitable period of time (3-6 hours), and then sub-cultured to determine cytotoxicity and to allow phenotypic expression prior to mutant selection (14) (15) (16) (17). Cytotoxicity is determined by relative survival (RS), i.e. cloning efficiency measured immediately after treatment and adjusted for any cell loss during treatment as compared to the negative control (paragraph 18 and Appendix 2). The treated cultures are maintained in growth medium for a sufficient period of time, characteristic of each cell type, to allow near-optimal phenotypic expression of induced mutations (typically a minimum of 7-9 days). Following phenotypic expression, mutant frequency is determined by seeding known numbers of cells in medium containing the selective agent to detect mutant colonies, and in medium without selective agent to determine the cloning efficiency (viability). After a suitable incubation time, colonies are counted. Mutant frequency is calculated based on the number of mutant colonies corrected by the cloning efficiency at the time of mutant selection.

DESCRIPTION OF THE METHOD

Preparations

Cells

9. The cell types used for the HPRT and XPRT tests should have a demonstrated sensitivity to chemical mutagens, a high cloning efficiency, a stable karyotype, and a stable spontaneous mutant frequency. The most commonly used cells for the HPRT test include the CHO, CHL and V79 lines of Chinese hamster cells, L5178Y mouse lymphoma cells, and TK6 human lymphoblastoid cells (18) (19). CHO-derived AS52 cells containing the gpt transgene (and having the Hprt gene deleted) are used for the XPRT test (20) (21); the HPRT test cannot be performed in AS52 cells because the hprt gene has been deleted. The use of other cell lines should be justified and validated.
10. Cell lines should be checked routinely for the stability of the modal chromosome number and the absence of Mycoplasma contamination (22) (23), and cells should not be used if contaminated or if the modal chromosome number has changed. The normal cell cycle time used in the testing laboratory should be established and should be consistent with the published cell characteristics. The spontaneous mutant frequency in the master cell stock should also be checked, and the stock should not be used if the mutant frequency is not acceptable.
11. Prior to use in this test, the cultures may need to be cleansed of pre-existing mutant cells, e.g. by culturing in HAT medium for HPRT test and MPA for XPRT test (5) (24) (See Appendix 1). The cleansed cells can be cryopreserved and then thawed to use as working stocks. The newly thawed working stock can be used for the test after normal doubling times are attained. When conducting the XPRT test, routine culture of AS52 cells should use conditions that assure the maintenance of the gpt transgene (20).

Media and culture conditions

12. Appropriate culture medium and incubation conditions (culture vessels, humidified atmosphere of 5 % CO₂, and incubation temperature of 37 °C) should be used for maintaining cultures. Cell cultures should always be maintained under conditions that ensure that they are growing in log phase. It is particularly important that media and culture conditions be chosen to ensure optimal growth of cells during the expression period and optimal cloning efficiency for both mutant and non-mutant cells.

Preparation of cultures

13. Cell lines are propagated from stock cultures, seeded in culture medium at a density such that the cells in suspensions or in monolayers will continue to grow exponentially through the treatment and expression periods (e.g. confluence should be avoided for cells growing in monolayers).

Metabolic activation

14. Exogenous metabolising systems should be used when employing cells which have inadequate endogenous metabolic capacity. The most commonly used system, that is recommended by default, unless otherwise justified, is a co-factor-supplemented post-mitochondrial fraction (S9) prepared from the livers of rodents (generally rats) treated with enzyme-inducing agents such as Aroclor 1254 (25) (26) (27) (28) or a combination of phenobarbital and β -naphthoflavone (29) (30) (31) (32). The latter combination does not conflict with the Stockholm Convention on Persistent Organic Pollutants (33) and has been shown to be as effective as Aroclor 1254 for inducing mixed-function oxidases (29) (31). The S9 fraction typically is used at concentrations ranging from 1 to 2 % (v/v) but may be increased to 10 % (v/v) in the final test medium. The choice of the type and concentration of exogenous metabolic activation system or metabolic inducer employed may be influenced by the class of substances being tested (34) (35) (36).

Test chemical preparation

15. Solid test chemicals should be prepared in appropriate solvents and diluted, if appropriate, prior to treatment of the cells (see paragraph 16). Liquid test chemicals may be added directly to the test system and/or diluted prior to treatment of the test system. Gaseous or volatile test chemicals should be tested by appropriate modifications to the standard protocols, such as treatment in sealed culture vessels (37) (38). Preparations of the test chemical should be made just prior to treatment unless stability data demonstrate the acceptability of storage.

TEST CONDITIONS

Solvents

16. The solvent should be chosen to optimise the solubility of the test chemicals without adversely impacting the conduct of the test e.g. changing cell growth, affecting the integrity of the test chemical, reacting with culture vessels, impairing the metabolic activation system. It is recommended that, wherever possible, the use of an aqueous solvent (or culture medium) should be considered first. Well established solvents are for example, water and dimethyl sulfoxide. Generally, organic solvents should not exceed 1 % (v/v) and aqueous solvents (saline or water) should not exceed 10 % (v/v) in the final treatment medium. If the solvents used are not well-established (e.g. ethanol or acetone), their use should be supported by data indicating their compatibility with the test chemicals and the test system, and their lack of genetic toxicity at the concentration used. In the absence of that supporting data, it is important to add untreated controls (see Appendix 1) to demonstrate that no deleterious or mutagenic effects are induced by the chosen solvent.

Measuring cytotoxicity and choosing exposure concentrations

17. When determining the highest test chemical concentration, concentrations that have the capability of producing artefactual positive responses, such as those producing excessive cytotoxicity (see paragraph 20), precipitation in the culture medium (see paragraph 21), or marked changes in pH or osmolality (see paragraph 5) should be avoided. If the test chemical causes a marked change in the pH of the medium at the time of addition, the pH might be adjusted by buffering the final treatment medium so as to avoid artefactual positive results and to maintain appropriate culture conditions.
18. Concentration selection is based on cytotoxicity and other considerations (see paragraphs 20-22). While the evaluation of cytotoxicity in an initial test may be useful to better define the concentrations to be used in the main experiment, an initial test is not required. Even if an initial cytotoxicity evaluation is performed, the measurement of cytotoxicity for each culture is still required in the main experiment. Cytotoxicity should be evaluated using RS, i.e. cloning efficiency (CE) of cells plated immediately after treatment, adjusted by any loss of cells during treatment, based on cell count, as compared with adjusted cloning efficiency in negative controls (assigned a survival of 100 %) (see Appendix 2 for the formula).

19. At least four test concentrations (not including the solvent and positive controls) that meet the acceptability criteria (appropriate cytotoxicity, number of cells, etc.) should be evaluated. While the use of duplicate cultures is advisable, either replicate or single treated cultures may be used at each concentration tested. The results obtained in the independent replicate cultures at a given concentration should be reported separately but can be pooled for the data analysis (17). For test chemicals demonstrating little or no cytotoxicity, concentration intervals of approximately 2 to 3 fold will usually be appropriate. Where cytotoxicity occurs, the test concentrations selected should cover a range from that producing cytotoxicity to concentrations at which there is moderate and little or no cytotoxicity. Many test chemicals exhibit steep concentration response curves and in order to cover the whole range of cytotoxicity or to study the concentration response relationship in detail, it may be necessary to use more closely spaced concentrations and more than four concentrations, in particular in situations where a repeat experiment is required (see paragraph 43). The use of more than 4 concentrations may be particularly important when using single cultures.

20. If the maximum concentration is based on cytotoxicity, the highest concentration should aim to achieve between 20 and 10 % RS. Care should be taken when interpreting positive results only found at 10 % RS or below (paragraph 43).

21. For poorly soluble test chemicals that are not cytotoxic at concentrations below the lowest insoluble concentration, the highest concentration analysed should produce turbidity or a precipitate visible by eye or with the aid of an inverted microscope at the end of the treatment with the test chemical. Even if cytotoxicity occurs above the lowest insoluble concentration, it is advisable to test at only one concentration producing turbidity or with a visible precipitate because artefactual effects may result from the precipitate. At the concentration producing a precipitate, care should be taken to assure that the precipitate does not interfere with the conduct of the test. The determination of solubility in the culture medium prior to the experiment may be useful.

22. If no precipitate or limiting cytotoxicity is observed, the highest test concentration should correspond to 10 mM, 2 mg/ml or 2 µl/ml, whichever is the lowest (39) (40). When the test chemical is not of defined composition, e.g. substance of unknown or variable composition, complex reaction products or biological materials (i.e. Chemical Substances of Unknown or Variable Composition (UVCBs)) (41), environmental extracts, etc., the top concentration may need to be higher (e.g. 5 mg/mL), in the absence of sufficient cytotoxicity, to increase the concentration of each of the components. It should be noted however that these requirements may differ for human pharmaceuticals (42).

Controls

23. Concurrent negative controls (see paragraph 16), consisting of solvent alone in the treatment medium and handled in the same way as the treatment cultures, should be included for every experimental condition.

24. Concurrent positive controls are needed to demonstrate the ability of the laboratory to identify mutagens under the conditions of the test protocol used and the effectiveness of the exogenous metabolic activation system, when applicable. Examples of positive controls are given in Table 1 below. Alternative positive control substances can be used, if justified. Because *in vitro* mammalian cell tests for genetic toxicity are sufficiently standardised, tests using treatments with and without exogenous metabolic activation may be conducted using only a positive control requiring metabolic activation. In this case, this single positive control response will demonstrate both the activity of the metabolic activation system and the responsiveness of the test system. Each positive control should be used at one or more concentrations expected to give reproducible and detectable increases over background in order to demonstrate the sensitivity of the test system, and the response should not be compromised by cytotoxicity exceeding the limits specified in this test method (see paragraph 20).

Table 1

Reference substances recommended for assessing laboratory proficiency and for selection of positive controls

Metabolic Activation condition	Locus	Substance and CAS No
Absence of exogenous metabolic activation	<i>Hprt</i>	Ethylmethanesulfonate [CAS no. 62-50-0] Ethylnitrosourea [CAS no. 759-73-9] 4-Nitroquinoline 1-oxide [CAS no. 56-57-5]
	<i>xprt</i>	Streptonigrin [CAS no. 3930-19-6] Mitomycin C [CAS no. 50-07-7]
Presence of exogenous metabolic activation	<i>Hprt</i>	3-Methylcholanthrene [CAS no. 56-49-5] 7,12-Dimethylbenzanthracene [CAS no. 57-97-6] Benzo[a]pyrene [CAS no. 50-32-8]
	<i>xprt</i>	Benzo[a]pyrene [CAS no. 50-32-8]

PROCEDURE

Treatment with test chemical

25. Proliferating cells are treated with the test chemical in the presence and absence of a metabolic activation system. Exposure should be for a suitable period of time (usually 3 to 6 hours is adequate).
26. The minimum number of cells used for each test (control and treated) culture at each stage in the test should be based on the spontaneous mutant frequency. A general guide is to treat and passage sufficient cells as to maintain 10 spontaneous mutants in every culture in all phases of the test (17). The spontaneous mutant frequency is generally between 5 and 20×10^{-6} . For a spontaneous mutant frequency of 5×10^{-6} and to maintain a sufficient number of spontaneous mutants (10 or more) even for the cultures treated at concentrations that cause 90 % cytotoxicity during treatment (10 % RS), it would be necessary to treat at least 20×10^6 cells. In addition a sufficient number of cells (but never less than 2 million) must be cultured during the expression period and plated for mutant selection (17).

Phenotypic expression time and measuring mutant frequency

27. After the treatment period, cells are cultured to allow expression of the mutant phenotype. A minimum of 7 to 9 days generally is sufficient to allow near optimal phenotypic expression of newly induced *Hprt* and *xprt* mutants (43) (44). During this period, cells are regularly sub-cultured to maintain them in exponential growth. After phenotypic expression, cells are re-plated in medium with and without selective agent (6-thioguanine) for the determination of the number of mutants and cloning efficiency at the time of selection, respectively. This plating can be accomplished using dishes for monolayer cultures or microwell plates for cells in suspension. For mutant selection, cells should be plated at a density to assure optimum mutant recovery (i.e. avoid metabolic cooperation) (17). Plates are incubated for an appropriate length of time for optimum colony growth (e.g. 7-12 days) and colonies counted. Mutant frequency is calculated based on the number of mutant colonies corrected by the cloning efficiency at the time of mutant selection (see Appendix 2 for formulas).

Proficiency of the laboratory

28. In order to establish sufficient experience with the test prior to using it for routine testing, the laboratory should have performed a series of experiments with reference positive substances acting via different mechanisms (at least one active with and one active without metabolic activation selected from the substances listed in Table 1) and various negative controls (using various solvents/vehicles). These positive and negative control responses should be consistent with the literature. This is not applicable to laboratories that have experience, i.e. that have an historical data base available as defined in paragraphs 30 to 33.
29. A selection of positive control substances (see Table 1 in paragraph 25) should be investigated in the absence and in the presence of metabolic activation, in order to demonstrate proficiency to detect mutagenic chemicals, to determine the effectiveness of the metabolic activation system and to demonstrate the appropriateness of the cell growth conditions during treatment, phenotypic expression and mutant selection and of the scoring procedures. A range of concentrations of the selected substances should be chosen so as to give reproducible and concentration-related increases above the background in order to demonstrate the sensitivity and dynamic range of the test system.

Historical control data

30. The laboratory should establish:
 - A historical positive control range and distribution,
 - A historical negative (untreated, solvent) control range and distribution.
31. When first acquiring data for an historical negative control distribution, concurrent negative controls should be consistent with published control data (22). As more experimental data are added to the control distribution, concurrent negative controls should ideally be within the 95 % control limits of that distribution (17) (45) (46).
32. The laboratory's historical negative control database should initially be built with a minimum of 10 experiments but would preferably consist of at least 20 experiments conducted under comparable experimental conditions. Laboratories should use quality control methods, such as control charts (e.g. C-charts or X-bar charts (47)), to identify how variable their positive and negative control data are, and to show that the methodology is 'under control' in their laboratory (46). Further recommendations on how to build and use the historical data (i.e. criteria for inclusion and exclusion of data in historical data and the acceptability criteria for a given experiment) can be found in the literature (45).
33. Negative control data should consist of mutant frequencies from single or preferably replicate cultures as described in paragraph 23. Concurrent negative controls should ideally be within the 95 % control limits of the distribution of the laboratory's historical negative control database (17) (45) (46). Where concurrent negative control data fall outside the 95 % control limit they may be acceptable for inclusion in the historical control distribution as long as these data are not extreme outliers and there is evidence that the test system is 'under control' (see above) and there is evidence of no technical or human failure.
34. Any changes to the experimental protocol should be considered in terms of their consistency with the laboratory's existing historical control databases. Any major inconsistencies should result in the establishment of a new historical control database.

DATA AND REPORTING

Presentation of the results

35. The presentation of results should include all of the data needed to calculate cytotoxicity (expressed as RS). The data, for both treated and control cultures, should include the number of cells at the end of treatment, the number of cells plated immediately following treatment, and the colony counts (or number of wells without colonies for the microwell method). RS for each culture should be expressed as a percentage relative to the concurrent solvent control (refer to Appendix 1 for definitions).

36. The presentation of results should also include all of the data needed to calculate the mutant frequency. Data for both treated and control cultures, should include: (1) the number of cells plated with and without selective agent (at the time the cells are plated for mutant selection), and (2) the number of colonies counted (or the number of wells without colonies for the microwell method) from the plates with and without selective agent. Mutant frequency is calculated based on the number of mutant colonies (in the plates with selective agent) corrected by the cloning efficiency (from the plates without selective agent). The mutant frequency should be expressed as the number of mutant cells per million viable cells (refer to Appendix 1 for definitions).

37. Individual culture data should be provided. Additionally, all data should be summarised in tabular form.

Acceptability Criteria

38. Acceptance of a test is based on the following criteria:
 - The concurrent negative control is considered acceptable for addition to the laboratory historical negative control database as described in paragraph 33.

 - Concurrent positive controls (see paragraph 24) should induce responses that are compatible with those generated in the historical positive control data base and produce a statistically significant increase compared with the concurrent negative control.

 - Two experimental conditions (i.e. with and without metabolic activation) were tested unless one resulted in positive results (see paragraph 25).

 - Adequate number of cells and concentrations are analysable (paragraphs 25, 26 and 19).

 - The criteria for the selection of top concentration are consistent with those described in paragraphs 20, 21 and 22.

Evaluation and interpretation of results

39. Providing that all acceptability criteria are fulfilled, a test chemical is considered to be clearly positive if, in any of the experimental conditions examined:
 - at least one of the test concentrations exhibits a statistically significant increase compared with the concurrent negative control,

 - the increase is concentration-related when evaluated with an appropriate trend test,

- any of the results are outside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limit; see paragraph 33).

When all of these criteria are met, the test chemical is then considered able to induce gene mutations in cultured mammalian cells in this test system. Recommendations for the most appropriate statistical methods can be found in the literature (46) (48).

40. Providing that all acceptability criteria are fulfilled, a test chemical is considered clearly negative if, in all experimental conditions examined:

- none of the test concentrations exhibits a statistically significant increase compared with the concurrent negative control,
- there is no concentration-related increase when evaluated with an appropriate trend test,
- all results are inside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limit; see paragraph 33).

The test chemical is then considered unable to induce gene mutations in cultured mammalian cells in this test system.

41. There is no requirement for verification of a clearly positive or negative response.
42. In cases when the response is neither clearly negative nor clearly positive as described above, or in order to assist in establishing the biological relevance of a result, the data should be evaluated by expert judgement and/or further investigations. Performing a repeat experiment possibly using modified experimental conditions (e.g. concentration spacing, other metabolic activation conditions [i.e. S9 concentration or S9 origin]) could be useful.
43. In rare cases, even after further investigations, the data set will preclude making a conclusion of positive or negative results. Therefore the test chemical response should be concluded to be equivocal (interpreted as equally likely to be positive or negative).

Test report

44. The test report should include the following information:

Test chemical:

- source, lot number, limit date for use, if available;
- stability of the test chemical itself, if known;
- solubility and stability of the test chemical in solvent, if known;
- measurement of pH, osmolality and precipitate in the culture medium to which the test chemical was added, as appropriate.

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Solvent:

- justification for choice of solvent;
- percentage of solvent in the final culture medium.

Cells:

For Laboratory master cultures:

- type, source of cell lines;
- number of passages, if available, and history in the laboratory;
- karyotype features and/or modal number of chromosomes;
- methods for maintenance of cell cultures;
- absence of mycoplasma;
- cell doubling times.

Test conditions:

- rationale for selection of concentrations and number of cultures including, e.g. cytotoxicity data and solubility limitations;
- composition of media, CO₂ concentration, humidity level;
- concentration of test chemical expressed as final concentration in the culture medium (e.g. µg or mg/ml or mM of culture medium);

- concentration (and/or volume) of solvent and test chemical added in the culture medium;
- incubation temperature;
- incubation time;
- duration of treatment;
- cell density during treatment;
- type and composition of metabolic activation system (source of S9, method of preparation of the S9 mix, the concentration or volume of S9 mix and S9 in the final culture medium, quality controls of S9);
- positive and negative control substances, final concentrations for each condition of treatment;
- length of expression period (including number of cells seeded, and subcultures and feeding schedules, if appropriate);
- identity of the selective agent and its concentration;
- criteria for acceptability of tests;
- methods used to enumerate numbers of viable and mutant cells;
- methods used for the measurements of cytotoxicity;
- any supplementary information relevant to cytotoxicity and method used;
- duration of incubation times after plating;
- criteria for considering studies as positive, negative or equivocal;
- methods used to determine pH, osmolality and precipitation.

Results:

- number of cells treated and number of cells sub-cultured for each culture;
- cytotoxicity measurements and other observations if any;
- signs of precipitation and time of the determination;

- number of cells plated in selective and non-selective medium;
- number of colonies in non-selective medium and number of resistant colonies in selective medium, and related mutant frequencies;
- concentration-response relationship, where possible;
- concurrent negative (solvent) and positive control data (concentrations and solvents);
- historical negative (solvent) and positive control data, with ranges, means and standard deviations and confidence interval (e.g. 95 %) as well as the number of data;
- statistical analyses (for individual cultures and pooled replicates if appropriate), and p-values if any.

Discussion of the results.

Conclusion

LITERATURE

- (1) OECD (2016). Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. ENV Publications. Series on Testing and Assessment, No 234, OECD, Paris.
- (2) Moore M.M., DeMarini D.M., DeSerres F.J. and Tindall, K.R. (Eds.) (1987). Banbury Report 28: Mammalian Cell Mutagenesis, Cold Spring Harbor Laboratory, New York, New York.
- (3) Chu E.H.Y. and Malling H.V. (1968). Mammalian Cell Genetics. II. Chemical Induction of Specific Locus Mutations in Chinese Hamster Cells *In Vitro*, Proc. Natl. Acad. Sci., USA, 61, 1306-1312.
- (4) Moore M.M., Harrington-Brock K., Doerr C.L. and Dearfield K.L. (1989). Differential Mutant Quantitation at the Mouse Lymphoma TK and CHO HGPRT Loci. *Mutagen. Res.*, 4, 394-403.
- (5) Aaron C.S. and Stankowski Jr. L.F. (1989). Comparison of the AS52/XPRT and the CHO/HPRT Assays: Evaluation of Six Drug Candidates. *Mutation Res.*, 223, 121-128.
- (6) Aaron C.S., Bolcsfoldi G., Glatt H.R., Moore M., Nishi Y., Stankowski L., Theiss J. and Thompson E. (1994). Mammalian Cell Gene Mutation Assays Working Group Report. Report of the International Workshop on Standardisation of Genotoxicity Test Procedures. *Mutation Res.*, 312, 235-239.
- (7) Li A.P., Gupta R.S., Heflich R.H. and Wasson J. S. (1988). A Review and Analysis of the Chinese Hamster Ovary/Hypoxanthine Guanine Phosphoribosyl Transferase System to Determine the Mutagenicity of Chemical Agents: A Report of Phase III of the U.S. Environmental Protection Agency Gene-tox Program. *Mutation Res.*, 196, 17-36.
- (8) Scott D., Galloway S.M., Marshall R.R., Ishidate M., Brusick D., Ashby J. and Myhr B.C. (1991). Genotoxicity Under Extreme Culture Conditions. A Report from ICPEMC Task Group 9. *Mutation Res.*, 257, 147-204.

- (9) Morita T., Nagaki T., Fukuda I. and Okumura K. (1992). Clastogenicity of Low pH to Various Cultured Mammalian Cells. *Mutation Res.*, 268, 297-305.
- (10) Brusick D. (1986). Genotoxic Effects in Cultured Mammalian Cells Produced by Low pH Treatment Conditions and Increased Ion concentrations, *Environ. Mutagen.*, 8, 789-886.
- (11) Nesslany F., Simar-Meintieres S., Watzinger M., Talahari I. and Marzin D. (2008). Characterization of the Genotoxicity of Nitrotriacetic Acid. *Environ. Mol. Mutation Res.*, 49, 439-452.
- (12) Long L.H., Kirkland D., Whitwell J. and Halliwell B. (2007). Different Cytotoxic and Clastogenic Effects of Epigallocatechin Gallate in Various Cell-Culture Media Due to Variable Rates of its Oxidation in the Culture Medium, *Mutation Res.*, 634, 177-183.
- (13) Kirkland D., Aardema M., Henderson L., and Müller L. (2005). Evaluation of the Ability of a Battery of Three *In Vitro* Genotoxicity Tests to Discriminate Rodent Carcinogens and Non-Carcinogens. I: Sensitivity, Specificity and Relative Predictivity. *Mutation Res.*, 5841-256.
- (14) Li A.P., Carver J.H., Choy W.N., Hsie A.W., Gupta R.S., Loveday K.S., O'Neill J.P., Riddle J.C., Stankowski L.F. Jr. and Yang L.L. (1987). A Guide for the Performance of the Chinese Hamster Ovary Cell/Hypoxanthine-Guanine Phosphoribosyl Transferase Gene Mutation Assay. *Mutation Res.*, 189, 135-141.
- (15) Liber H.L., Yandell D.W. and Little J.B. (1989). A Comparison of Mutation Induction at the TK and HPRT Loci in Human Lymphoblastoid Cells; Quantitative Differences are Due to an Additional Class of Mutations at the Autosomal TK Locus. *Mutation Res.*, 216, 9-17.
- (16) Stankowski L.F. Jr., Tindall K.R. and Hsie A.W. (1986). Quantitative and Molecular Analyses of Ethyl Methanesulfonate- and ICR 191-Induced Molecular Analyses of Ethyl Methanesulfonate- and ICR 191-Induced Mutation in AS52 Cells. *Mutation Res.*, 160, 133-147.
- (17) Arlett C.F., Smith D.M., Clarke G.M., Green M.H.L., Cole J., McGregor D.B. and Asquith J.C. (1989). Mammalian Cell Gene Mutation Assays Based upon Colony Formation. In: *Statistical Evaluation of Mutagenicity Test Data*, Kirkland, D.J. (Eds.), Cambridge University Press, pp. 66-101.
- (18) Hsie A.W., Casciano D.A., Couch D.B., Krahn D.F., O'Neill J.P., and Whitfield B.L. (1981). The Use of Chinese Hamster Ovary Cells to Quantify Specific Locus Mutation and to Determine Mutagenicity of Chemicals; a Report of the Gene-Tox Program, *Mutation Res.*, 86, 193-214.
- (19) Li A.P. (1981). Simplification of the CHO/HGPRT Mutation Assay Through the Growth of Chinese Hamster Ovary Cells as Unattached Cultures, *Mutation Res.*, 85, 165-175.
- (20) Tindall K.R., Stankowski Jr., L.F., Machanoff R., and Hsie A.W. (1984). Detection of Deletion Mutations in pSV2gpt-Transformed Cells, *Mol. Cell. Biol.*, 4, 1411-1415.
- (21) Hsie A. W., Recio L., Katz D. S., Lee C. Q., Wagner M., and Schenley R. L. (1986). Evidence for Reactive Oxygen Species Inducing Mutations in Mammalian Cells. *Proc Natl Acad Sci.*, 83(24): 9616-9620.

- (22) Lorge E., Moore M., Clements J., Donovan M. O., Honma M., Kohara A., Van Benthem J., Galloway S., Armstrong M.J., Thybaud V., Gollapudi B., Aardema M., Kim J., Sutter A., Kirkland D.J. (2015). Standardized Cell Sources and Recommendations for Good Cell Culture Practices in Genotoxicity Testing. (Manuscript in preparation).
- (23) Coecke S., Balls M., Bowe G., Davis J., Gstraunthaler G., Hartung T., Hay R., Merten O.W., Price A., Schechtman L., Stacey G. and Stokes W. (2005). Guidance on Good Cell Culture Practice. A Report of the Second ECVAM Task Force on Good Cell Culture Practice, ATLA, 33, 261-287.
- (24) Rosen M.P., San R.H.C. and Stich H.F. (1980). Mutagenic Activity of Ascorbate in Mammalian Cell Cultures, Can. Lett. 8, 299-305.
- (25) Natarajan A.T., Bates A.D, Van Buul P.P.W., Meijers M. and de Vogel N. (1976). Cytogenetic Effects of Mutagens/Carcinogens after Activation in a Microsomal System *In Vitro*, I. Induction of Chromosomal Aberrations and Sister Chromatid Exchanges by Diethylnitrosamine (DEN) and Dimethylnitrosamine (DMN) in CHO Cells in the Presence of Rat-Liver Microsomes. Mutation Res., 37, 83-90.
- (26) Abbondandolo A., Bonatti S., Corti G., Fiorio R., Loprieno N. and Mazzaccaro A. (1977). Induction of 6-Thioguanine-Resistant Mutants in V79 Chinese Hamster Cells by Mouse-Liver Microsome-Activated Dimethylnitrosamine. Mutation Res., 46, 365-373.
- (27) Ames B.N., McCann J. and Yamasaki E. (1975). Methods for Detecting Carcinogens and Mutagens with the Salmonella/Mammalian-Microsome Mutagenicity Test. Mutation Res., 31, 347-364.
- (28) Maron D.M. and Ames B.N. (1983). Revised Methods for the Salmonella Mutagenicity Test. Mutation Res., 113, 173, 215.
- (29) Elliott B.M., Combes R.D., Elcombe C.R., Gatehouse D.G., Gibson G.G., Mackay J.M. and Wolf R.C. (1992) Alternatives to Aroclor 1254-Induced S9 in *In Vitro* Genotoxicity Assays. Mutagen. 7, 175-177.
- (30) Matsushima T., Sawamura M., Hara K. and Sugimura T. (1976). A Safe Substitute for Polychlorinated Biphenyls as an Inducer of Metabolic Activation Systems. In: *In Vitro* Metabolic Activation in Mutagenesis Testing, de Serres F.J., Fouts J.R., Bend J.R. and Philpot R.M. (Eds), Elsevier, North-Holland, pp. 85-88.
- (31) Ong T.-m., Mukhtar M., Wolf C.R. and Zeiger E. (1980). Differential Effects of Cytochrome P450-Inducers on Promutagen Activation Capabilities and Enzymatic Activities of S-9 from Rat Liver, J. Environ. Pathol. Toxicol., 4, 55-65.
- (32) Johnson T.E., Umbenhauer D.R. and Galloway S.M. (1996). Human Liver S-9 Metabolic Activation: Proficiency in Cytogenetic Assays and Comparison with Phenobarbital/beta-Naphthoflavone or Aroclor 1254 Induced Rat S-9, Environ. Mol. Mutagen., 28, 51-59.
- (33) UNEP. (2001). Stockholm Convention on Persistent Organic Pollutants, United Nations Environment Programme (UNEP). Available at: [<http://www.pops.int.html>].
- (34) Tan E.-L. and Hsie A.W. (1981). Effect of Calcium Phosphate and Alumina Cy Gels on the Mutagenicity and Cytotoxicity of Dimethylnitrosamine as Studied in the CHO/HGPRT System. Mutation Res., 84, 147-156.

- (35) O'Neill J.P., Machanoff R., San Sebastian J.R., Hsie A.W. (1982). Cytotoxicity and Mutagenicity of Dimethylnitrosamine in Camalian Cells (CHO/HGPRT system): Enhancement by Calcium Phosphate. *Environ. Mol. Mutation.*, 4, 7-18.
- (36) Li, A.P. (1984). Use of Aroclor 1254-Induced Rat Liver Homogenate in the Assaying of Promutagens in Chinese Hamster Ovary Cells. *Environ. Mol. Mutation*, 4, 7-18.
- (37) Krahn D.F., Barsky F.C. and McCooey K.T. (1982). CHO/HGPRT Mutation Assay: Evaluation of Gases and Volatile Liquids. In: Tice, R.R., Costa, D.L., Schaich, K.M. (eds.) *Genotoxic Effects of Airborne Agents*. New York, Plenum, pp. 91-103.
- (38) Zamora P.O., Benson J.M., Li A.P. and Brooks A.L. (1983). Evaluation of an Exposure System Using Cells Grown on Collagen Gels for Detecting Highly Volatile Mutagens in the CHO/HGPRT Mutation Assay. *Environ. Mutagen.*,5, 795-801.
- (39) OECD (2014). Document Supporting the WNT Decision to Implement Revised Criteria for the Selection of the Top Concentration in the *In Vitro* Mammalian Cell Assays on Genotoxicity (Test Guidelines 473, 476 and 487). Available upon request from the Organisation for Economic Cooperation and Development.
- (40) Brookmire L., Chen J.J. and Levy D.D. (2013). Evaluation of the Highest Concentrations Used in the *In Vitro* Chromosome Aberrations Assay, *Environ. Mol. Mutation*, 54, 36-43.
- (41) EPA, Office of Chemical Safety and Pollution Prevention. (2011). Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials: UVCB Substances,
- (42) USFDA (2012). International Conference on Harmonisation (ICH) Guidance S2 (R1) on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended for Human Use. Available at: [<https://federalregister.gov/a/2012-13774>].
- (43) O'Neill J.P. and Hsie A.W. (1979). Phenotypic Expression Time of Mutagen-Induced 6-thioguanine resistance in Chinese hamster ovary cells (CHO/HGPRT system), *Mutation, Res.*, 59, 109-118.
- (44) Chiewchanwit T., Ma H., El Zein R., Hallberg L., and Au W.W. (1995). Induction of Deletion Mutations by Methoxyacetaldehyde in Chinese Hamster Ovary (CHO)-AS52 cells. *Mutation, Res.*, 1335(2):121-8.
- (45) Hayashi M., Dearfield K., Kasper P., Lovell D., Martus H.J., and Thybaud V. (2011). Compilation and Use of Genetic Toxicity Historical Control Data, *Mutation, Res.*, 723, 87-90.
- (46) OECD (2014). Statistical Analysis Supporting the Revision of the Genotoxicity Test Guidelines. Environmental, Health and Safety, Series on testing and assessment (No 199), Organisation for Economic Cooperation and Development, Paris.
- (47) Richardson C., Williams D.A., Allen J.A., Amphlett G., Chanter D.O., and Phillips B. (1989). Analysis of Data from *In Vitro* Cytogenetic Assays. In: *Statistical Evaluation of Mutagenicity Test Data*. Kirkland, D.J., (Ed) Cambridge University Press, Cambridge, pp. 141-154.
- (48) Fleiss J. L., Levin B., and Paik M. C. (2003). *Statistical Methods for Rates and Proportions*, Third Edition, New York: John Wiley & Sons

Appendix 1

DEFINITIONS

Base pair substitution mutagens: chemicals that cause substitution of base pairs in the DNA.

Chemical: A substance or a mixture.

Cloning efficiency: The percentage of cells plated at a low density that are able to grow into a colony that can be counted.

Concentrations: refer to final concentrations of the test chemical in culture medium

Cytotoxicity: For the assays covered in this test method, cytotoxicity is identified as a reduction in relative survival of the treated cells as compared to the negative control (see specific paragraph).

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

Frameshift mutagens: chemicals which cause the addition or deletion of single or multiple base pairs in the DNA molecule.

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

HAT medium: medium containing Hypoxanthine, Aminopterin and Thymidine, used for cleansing of Hprt mutants.

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

MPA medium: medium containing Xanthine, Adenine, Thymidine, Aminopterin and Mycophenolic acid, used for cleansing of Xprt mutants.

Mutagenic: produces a heritable change of DNA base-pair sequences(s) in genes or of the structure of chromosomes (chromosome aberrations).

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

Phenotypic expression time: The time after treatment during which the genetic alteration is fixed within the genome and any preexisting gene products are depleted to the point that the phenotypic trait is altered.

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

S9 liver fractions: supernatant of liver homogenate after 9 000g centrifugation, i.e. raw liver extract

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

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

Test chemical: Any substance or mixture tested using this test method.

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

UVCB: Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials

Appendix 2

FORMULAS FOR ASSESSMENT OF CYTOTOXICITY AND MUTANT FREQUENCY

Cytotoxicity is evaluated by relative survival, i. e., cloning efficiency (CE) of cells plated immediately after treatment adjusted by any loss of cells during treatment as compared with adjusted cloning efficiency in negative controls (assigned a survival of 100 %) (see RS formula below).

Adjusted CE for a culture treated by a test chemical is calculated as:

$$\text{Adjusted CE} = \frac{\text{Number of cells at the end of treatment}}{\text{Number of cells at the beginning of treatment}}$$

RS for a culture treated by a test chemical is calculated as:

$$\text{RS} = \frac{\text{Adjusted CE in treated culture}}{\text{Adjusted CE in the solvent control}} \times 100$$

Mutant frequency is the cloning efficiency of mutant colonies in selective medium divided by the cloning efficiency in non-selective medium measured for the same culture at the time of selection.

$$\text{Mutant frequency} = \frac{\text{Cloning efficiency of mutant colonies in selective medium}}{\text{Cloning efficiency in non – selective medium}}$$

When plates are used for cloning efficiency:

CE = Number of colonies / Number of cells plated.

When micro-well plates are used for cloning efficiency:

The number of colonies per well on micro-wells plates follows a Poisson distribution.

Cloning Efficiency = $-\text{LnP}(0)$ / Number of cells plated per well

Where $-\text{LnP}(0)$ is the probable number of empty wells out of the seeded wells and is described by the following formula

$\text{LnP}(0) = -\text{Ln}(\text{number of empty wells} / \text{number of plated wells})$

- (3) In Part B, Chapter B.22 is replaced by the following:

"B.22 RODENT DOMINANT LETHAL TEST

INTRODUCTION

1. This test method (TM) is equivalent to the OECD test guideline (TG) 478 (2016). Test methods are periodically reviewed in the light of scientific progress, changing regulatory needs, and animal welfare considerations. This modified version of the test method reflects more than thirty years of experience with this test and the potential for integrating or combining this test with other toxicity tests such as developmental, reproductive toxicity, or genotoxicity studies; however due to its limitations and the use of a large number of animals this assay is not intended for use as a primary method, but rather as a supplemental test method which can only be used when there is no alternative for regulatory requirements. Combining toxicity testing has the potential to spare large numbers of animals from use in toxicity tests. A document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to genetic toxicity OECD test guidelines has been developed by OECD (1).
2. The purpose of the Dominant lethal (DL) test is to investigate whether chemicals produce mutations resulting from chromosomal aberrations in germ cells. In addition, the dominant lethal test is relevant to assessing genotoxicity because, although they may vary among species, factors of *in vivo* metabolism, pharmacokinetics and DNA-repair processes are active and contribute to the response. Induction of a DL mutation after exposure to a test chemical indicates that the chemical has affected germinal tissue of the test animal.
3. DL mutations cause embryonic or foetal death. Induction of DL mutation after exposure to a test chemical indicates that the chemical has affected the germ cells of the test animal.
4. A DL assay is useful for confirmation of positive results of tests using somatic *in vivo* endpoints, and is a relevant endpoint for the prediction of human hazard and risk of genetic diseases transmitted through the germline. However, this assay requires a large number of animals and is labour-intensive; as a result, it is very expensive and time-consuming to conduct. Because the spontaneous frequency of dominant lethal mutations is quite high, the sensitivity of the assay for detection of small increases in the frequency of mutations is generally limited.
5. Definitions of key terms are set out in Appendix 1.

INITIAL CONSIDERATIONS

6. The test is most often conducted in mice (2) (3) (4) but other species, such as rats (5) (6) (7) (8), may in some cases be appropriate if scientifically justified. DLs generally are the result of gross chromosomal aberrations (structural and numerical abnormalities) (9) (10) (11), but gene mutations cannot be excluded. A DL mutation is a mutation occurring in a germ cell per se, or is fixed post fertilisation in the early embryo, that does not cause dysfunction of the gamete, but is lethal to the fertilised egg or developing embryo.
7. Individual males are mated sequentially to virgin females at appropriate intervals. The number of matings following treatment is dependent on the ultimate purpose of the DL study (Paragraph 23) and should ensure that all phases of male germ cell maturation are evaluated for DLs (12).
8. If there is evidence that the test chemical, or its metabolite(s), will not reach the testis, it is not appropriate to use this test.

PRINCIPLE OF THE TEST

9. Generally, male animals are exposed to a test chemical by an appropriate route of exposure and mated to untreated virgin females. Different germ cell types can be tested by the use of sequential mating intervals. Following mating, the females are euthanised after an appropriate period of time, and their uteri are examined to determine the numbers of implants and live and dead embryos. The dominant lethality of a test chemical is determined by comparing the live implants per female in the treated group with the live implants per female in the vehicle/solvent control group. The increase of dead implants per female in the treated group over the dead implants per female in the control group reflects the test-chemical-induced post-implantation loss. The post-implantation loss is calculated by determining the ratio of dead to total implants in the treated group compared to the ratio of dead to total implants in the control group. Pre-implantation loss can be estimated by comparing corpora lutea counts minus total implants or the total implants per female in treated and control groups.

VERIFICATION OF LABORATORY PROFICIENCY

10. Competence in this assay should be established by demonstrating the ability to reproduce dominant lethal frequencies from published data (e.g. (13) (14) (15) (16) (17) (18)) with positive control substances (including weak responses) such as those listed in Table 1, and vehicle controls and obtaining negative control frequencies that are consistent acceptable range of data (see references above) or with the laboratory's historical control distribution, if available.

DESCRIPTION OF THE METHOD

Preparations*Selection of animal species*

11. Commonly used laboratory strains of healthy sexually mature animals should be employed. Mice are commonly used but rats may also be appropriate. Any other appropriate mammalian species may be used, if scientific justification is provided in the report.

Animal housing and feeding conditions

12. For rodents, the temperature in the animal room should be 22 °C (\pm 3 °C). Although the relative humidity ideally should be 50-60 %, it should be at least 40 % and preferably not exceed 70 %, other than during room cleaning. Lighting should be artificial, the sequence being 12 hours light, followed by 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. The choice of diet may be influenced by the need to ensure a suitable admixture of a test chemical when administered by this route. Prior to treatment or mating, rodents should be housed in small groups (no more than five) of the same sex if no aggressive behaviour is expected or observed, preferably in solid cages with appropriate environmental enrichment. Animals may be housed individually if scientifically justified.

Preparation of the animals

13. Healthy and sexually mature male and female adult animals are randomly assigned to the control and treatment groups. The individual animals are identified uniquely using a humane, minimally invasive method (e.g. by ringing, tagging, micro-chipping, or biometric identification, but not toe and ear clipping) and acclimated to the laboratory conditions for at least five days. Cages should be arranged in such a way that possible effects due to cage placement are minimised. Cross contamination by the positive control and the test chemical should be avoided. At the commencement of the study, the weight variation of animals should be minimal and not exceed \pm 20 % of the mean weight of each sex.

Preparation of doses

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

Test Conditions*Solvent/vehicle*

15. The solvent/vehicle should not produce toxic effects at the dose volumes used, and should not be suspected of chemical reaction with the test chemical. If other than well-known solvents/vehicles are used, their inclusion should be supported with reference data indicating their compatibility. It is recommended that wherever possible, the use of an aqueous solvent/vehicle should be considered first. Examples of commonly used compatible solvents/vehicles include water, physiological saline, methylcellulose solution, carboxymethyl cellulose sodium salt solution, olive oil and corn oil.

Positive controls

16. Concurrent positive control animals should always be used unless the laboratory has demonstrated proficiency in the conduct of the test and has used the test routinely in the recent past (e.g. within the last 5 years). However, it is not necessary to treat positive control animals by the same route as animals receiving the test chemical, or sample all the mating intervals. The positive control substances should be known to produce DLs under the conditions used for the test. Except for the treatment, animals in the control groups should be handled in an identical manner to animals in the treated groups.
17. The doses of the positive control substances should be selected so as to produce weak or moderate effects that critically assess the performance and sensitivity of the assay, but which consistently produce positive dominant lethal effects. Examples of positive control substances, and appropriate doses, are included in Table 1.

Table 1

Examples of Positive Control Substances.

Substance [CAS no.] (reference no.)	Effective Dose range (mg/kg) (rodent species)	Administration Time (days)
Triethylenemelamine [51-18-3] (15)	0,25 (mice)	1
Cyclophosphamide [50-18-0] (19)	50-150 (mice)	5
Cyclophosphamide [50-18-0] (5)	25-100 (rats)	1
Ethyl methanesulphonate [62-50-0] (13)	100-300 (mice)	5
Monomeric Acrylamide [79-06-1] (17)	50 (mice)	5
Chlorambucil [305-03-3] (14)	25 (mice)	1

Negative controls

18. Negative control animals, treated with solvent or vehicle alone, and otherwise treated in the same way as the treatment groups, should be included for every sampling time (20). In the absence of historical or published control data showing that no DLs or other deleterious effects are induced by the chosen solvent/vehicle, untreated control animals should also be included for every sampling time in order to establish acceptability of the vehicle control.

PROCEDURE

Number of Animals

19. Individual males are mated sequentially at appropriate predetermined intervals (e.g. weekly intervals, Paragraphs 21 & 23) preferably to one virgin female. The number of males per group should be predetermined to be sufficient (in combination with the number of mated females at each mating interval) to provide the statistical power necessary to detect at least a doubling in DL frequency (Paragraph 44).
20. The number of females per mating interval should also be predetermined by statistical power calculations to permit the detection of at least a doubling in the DL frequency (i.e. sufficient pregnant females to provide at least 400 total implants) (20) (21) (22) (23) and that at least one dead implant per analysis unit (i.e. mating group per dose) is expected (24).

Administration Period and Mating Intervals

21. The number of mating intervals following treatment is governed by the treatment schedule and should ensure that all phases of male germ cell maturation are evaluated for DL induction (12) (25). For a single treatment up to five daily dose administrations, there should be 8 (mouse) or 10 (rat) matings conducted at weekly intervals following the last treatment. For multiple dose administrations, the number of mating intervals may be reduced in proportion to the increased time of the administration period, but maintaining the goal of evaluating all phases of spermatogenesis (e.g. after a 28-day exposure, only 4 weekly matings are sufficient to evaluate all phases of spermatogenesis in the mouse). All treatment and mating schedules should be scientifically justified.
22. Females should remain with the males for at least the duration of one oestrus cycle (e.g. one week covers one oestrus cycle in both mice and rats). Females that did not mate during a one-week interval can be used for a subsequent mating interval. Alternatively, until mating has occurred, as determined by the presence of sperm in the vagina or by the presence of a vaginal plug.
23. The exposure and mating regimen used is dependent on the ultimate purpose of the DL study. If the goal is to determine whether a given chemical induces DL mutations *per se*, then the accepted method would be to expose an entire round of spermatogenesis (e.g. 7 weeks in the mouse, 5-7 treatments per week) and mate once at the end. However, if the goal is to identify the sensitive germ cell type for DL induction, then a single or 5 day exposure followed by weekly mating is preferred.

Dose Levels

24. If a preliminary range-finding study is performed because there are no suitable data already available to aid in dose selection, it should be performed in the same laboratory, using the same species, strain, sex, and treatment regimen to be used in the main study (26). The study should aim to identify the maximum tolerated dose (MTD), defined as the highest dose that will be tolerated without evidence of study-limiting toxicity, relative to the duration of the study period (for example, abnormal behaviour or reactions, minor body weight depression or hematopoietic system cytotoxicity), but not death or evidence of pain, suffering or distress necessitating humane euthanasia (27).

25. The MTD must also not adversely affect mating success (21).
26. Test chemicals with specific biological activities at low non-toxic doses (such as hormones and mitogens), and chemicals which exhibit saturation of toxicokinetic properties may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis.
27. In order to obtain dose response information, a complete study should include a negative control group and a minimum of three dose levels generally separated by a factor of 2, but not greater than 4. If the test chemical does not produce toxicity in a range-finding study, or based on existing data, the highest dose for a single administration should be 2 000 mg/kg body weight. However, if the test chemical does cause toxicity, the MTD should be the highest dose administered and the dose levels used should preferably cover a range from the maximum to a dose producing little or no toxicity. For not-toxic chemicals, the limit dose for an administration period of 14 days or more is 1 000 mg/kg body weight/day, and for administration periods of less than 14 days the limit dose is 2 000 mg/kg body weight/day.

Administration of Doses

28. The anticipated route of human exposure should be considered when designing an assay. Therefore, routes of exposures such as dietary, drinking water, subcutaneous, intravenous, topical, inhalation, oral (by gavage), or implantation may be chosen as justified. In any case, the route should be chosen to ensure adequate exposure of the target tissue(s). Intraperitoneal injection is not normally recommended since it is not an intended route of human exposure, and should only be used with specific scientific justification. If the test chemical is admixed in diet or drinking water, especially in case of single dosing, care should be taken that the delay between food and water consumption and mating should be sufficient to allow detection of the effects (paragraph 31). 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/100g body weight except in the case of aqueous solutions where a maximum of 2 ml/100g may be used. The use of volumes greater than this (if permitted by animal welfare legislation) should be justified. Variability in test volume should be minimised by adjusting the concentration to ensure a constant volume in relation to body weight at all dose levels.

Observations

29. General clinical observations of the test animals should be made and clinical signs recorded at least once a day, preferably at the same time(s) each day and considering the peak period of anticipated effects after dosing. At least twice daily during the dosing period, all animals should be observed for morbidity and mortality. All animals should be weighed at the beginning of the study and at least once a week during repeated dose studies, and at the time of euthanasia. Measurements of food consumption should be made at least weekly. If the test chemical is administered via the drinking water, water consumption should be measured at each change of water and at least weekly. Animals exhibiting non-lethal indicators of excess toxicity should be euthanised prior to completion of the test period (27).

Tissue Collection and Processing

30. Females are euthanised in the second half of pregnancy at gestation day (GD) 13 for mice and GD 14-15 for rats. Uteri are examined for dominant lethal effects to determine the number of implants, live and dead embryos, and corpora lutea.
31. The uterine horns and ovaries are exposed for counting of corpora lutea, and fetuses are removed, counted, and weighted. Care should be taken to examine the uteri for resorptions obscured by live fetuses and to ensure that all resorptions are enumerated. Fetal mortality is recorded. The number of successfully impregnated females and the number of total implantations, pre-implantation losses, and post-implantation mortality (included early and late resorptions) also are recorded. In addition, the visible fetuses may be preserved in Bouin's fixative for at least 2 weeks followed by examination for major external malformations (28) to provide additional information on the reproductive and developmental effects of the test agent.

DATA AND REPORTING

Treatment of Results

32. Data should be tabulated to show the number of males mated, the number of pregnant females, and the number of non-pregnant females. Results of each mating, including the identity of each male and female, should be reported individually. The mating interval, dose level for treated males, and the numbers of live implants and dead implants should be enumerated for each female.
33. The post-implantation loss is calculated by determining the ratio of dead to total implants from the treated group compared to the ratio of dead to total implants from the vehicle/solvent control group.
34. Pre-implantation loss is calculated as the difference between the number of corpora lutea and the number of implants, or as a reduction in the average number of implants per female in comparison with control matings. Where pre-implantation loss is estimated, it should be reported.
35. The Dominant Lethal factor is estimated as: $(\text{post-implantation deaths}/\text{total implantations per female}) \times 100$.
36. Data on toxicity and clinical signs (as per Paragraph 29) should be reported.

Acceptability Criteria

37. The following criteria determine the acceptability of a test.
 - Concurrent negative control is consistent with published norms for historical negative control data, and the laboratory's historical control data if available (see Paragraphs 10 and 18).
 - Concurrent positive controls induce responses that are consistent with published norms for historic positive control data, or the laboratory's historical positive control database, if available, and produce a statistically significant increase compared with the negative control (see Paragraphs 17 and 18).
 - Adequate number total implants and doses have been analysed (Paragraph 20).
 - The criteria for the selection of top dose are consistent with those described in Paragraphs 24 and 27.

Evaluation and Interpretation of Results

38. At least three treated dose groups should be analysed in order to provide sufficient data for dose-response analysis.
39. Providing that all acceptability criteria are fulfilled, a test chemical is considered a clear positive if:
 - at least one of the test doses exhibits a statistically significant increase compared with the concurrent negative control;
 - the increase is dose-related in at least one experimental condition (e.g. a weekly mating interval) when evaluated with an appropriate test; and,
 - any of the results are outside of the acceptable range of negative control data, or the distribution of the laboratory's historical negative control data (e.g. Poisson-based 95 % control limit) if available.

The test chemical is then considered able to induce dominant lethal mutations in germ cells of the test animals. Recommendations for the most appropriate statistical methods are described in Paragraph 44; other recommend statistical approaches can also be found in the literature (20) (21) (22) (24) (29). Statistical tests used should consider the animal as the experimental unit.

40. Providing that all acceptability criteria are fulfilled, a test chemical is considered a clear negative if:

- none of the test doses exhibits a statistically significant increase compared with the concurrent negative control;
- there is no dose-related increase in any experimental condition; and
- all results are within acceptable range of negative control data, or the laboratory's historical negative control data (e.g. Poisson-based 95 % control limit), if available.

The test chemical is then considered unable to induce dominant lethal mutations in germ cells of the test animals.

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

42. If the response is not clearly negative or positive, and in order to assist in establishing the biological relevance of a result (e.g. a weak or borderline increase), the data should be evaluated by expert judgment and/or further investigations using the existing experimental data, such as consideration whether the positive result is outside the acceptable range of negative control data, or the laboratory's historical, negative control data (30).

43. In rare cases, even after further investigations, the data set will preclude making a conclusion of positive or negative results, and will therefore be concluded as equivocal.

44. Statistical tests used should consider the male animal as the experimental unit. While it is possible that count data (e.g. number of implants per female) may be Poisson distributed and/or proportions (e.g. proportion of dead implants) may be binomially distributed, it is often the case that such data are overdispersed (31). Accordingly, statistical analysis should first employ a test for over- underdispersion using variance tests such as Cochran's binomial variance test (32) or Tarone's $C(\alpha)$ test for binomial overdispersion (31) (33). If no departure from binomial dispersion is detected, trends in proportions across dose levels may be tested using the Cochran-Armitage trend test (34) and pairwise comparisons with the control group may be tested using Fisher's exact test (35). Likewise, if no departure from Poisson dispersion is detected, trends in counts may be tested using Poisson regression (36) and pairwise comparisons with the control group may be tested within the context of the Poisson model, using pairwise contrasts (36). If significant overdispersion or underdispersion is detected, nonparametric methods are recommended (23) (31). These include rank-based tests, such as the Jonckheere-Terpstra test for trend (37) and Mann-Whitney tests (38) for pairwise comparisons with the vehicle/solvent control group, as well as permutation, resampling, or bootstrap tests for trend and pairwise comparisons with the control group (31) (39).

45. A positive DL assay provides evidence for the genotoxicity of the test chemical in the germ cells of the treated male of the test species.

46. Consideration of whether the observed values are within or outside of the historical control range can provide guidance when evaluating the biological significance of the response (40).

Test Report

47. The test report should include the following information.

Summary.

Test chemical:

- source, lot number, limit date for use, if available;
- stability of the test chemical itself, if known;
- solubility and stability of the test chemical in solvent, if known;
- measurement of pH, osmolality, and precipitate in the culture medium to which the test chemical was added, as appropriate.

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Test chemical preparation:

- justification for choice of vehicle;
- solubility and stability of the test chemical in the solvent/vehicle, if known;
- preparation of dietary, drinking water or inhalation formulations;
- analytical determinations on formulations (e.g. stability, homogeneity, nominal concentrations) when conducted.

Test animals:

- species/strain used and justification for the choice;
- number, age and sex of animals;

- source, housing conditions, diet, etc.;
- method of uniquely identifying the animals;
- for short-term studies: individual body weight of the male animals at the start and end of the test; for studies longer than one week: individual body weights during the study and food consumption. Body weight range, mean and standard deviation for each group should be included.

Test conditions:

- positive and negative (vehicle/solvent) control data;
- data from the range-finding study;
- rationale for dose level selection;
- details of test chemical preparation;
- details of the administration of the test chemical;
- rationale for route of administration;
- methods for measurement of animal toxicity, including, where available, histopathological or hematological analyses and the frequency with which animal observations and body weights were taken;
- methods for verifying that the test chemical reached the target tissue, or general circulation, if negative results are obtained;
- actual dose (mg/kg body weight/day) calculated from diet/drinking water test chemical concentration (ppm) and consumption, if applicable;
- details of food and water quality;
- details on cage environment enrichment;
- detailed description of treatment and sampling schedules and justifications for the choices;
- method of analgesia
- method of euthanasia;
- procedures for isolating and preserving tissues;
- source and lot numbers of all kits and reagents (where applicable);

- methods for enumeration of DLs;
- mating schedule;
- methods used to determine that mating has occurred;
- time of euthanasia;
- criteria for scoring DL effects, including, corpora lutea, implantations, resorptions and pre-implantation losses, live implants, dead implants.

Results:

- animal condition prior to and throughout the test period, including signs of toxicity;
- male body weight during the treatment and mating periods;
- number of mated females;
- dose-response relationship, where possible;
- concurrent and historical negative control data with ranges, means and standard deviations;
- concurrent positive control data;
- tabulated data for each dam including: number of corpora lutea per dam; number of implantations per dam; number of resorptions and pre-implantation losses per dam; number of live implants per dam; number of dead implants per dam; fetus weights;
- the above data summarised for each mating period and dose, with Dominant Lethal frequencies;
- statistical analyses and methods applied.

Discussion of the results.

Conclusion.

LITERATURE

- (1) OECD (2016). Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. ENV Publications. Series on Testing and Assessment, No 234, OECD, Paris.
- (2) Bateman, A.J. (1977). The Dominant Lethal Assay in the Male Mouse, in Handbook of Mutagenicity Test Procedures B.J. Kilbey *et. al.*(Eds.) pp. 235-334, Elsevier, Amsterdam

- (3) Ehling U.H., Ehling, U.H., Macheimer, L., Buselmaier, E., Dycka, D., Frohberg, H., Kratochvilova, J., Lang, R., Lorke, D., Muller, D., Peh, J., Rohrborn, G., Roll, R., Schulze-Schencking, M., and Wiemann, H. (1978). Standard Protocol for the Dominant Lethal Test on Male Mice. Set up by the Work Group "Dominant" lethal mutations of the ad hoc Committee Chemogenetics, *Arch. Toxicol.*, 39, 173-185
- (4) Shelby M.D. (1996). Selecting Chemicals and Assays for Assessing Mammalian Germ Cell Mutagenicity. *Mutation Res.*, 352:159-167.
- (5) Knudsen I., Knudsen, I., Hansen, E.V., Meyer, O.A. and Poulsen, E. (1977). A proposed Method for the Simultaneous Detection of Germ-Cell Mutations Leading to Fetal Death (Dominant Lethality) and of Malformations (Male Teratogenicity) in Mammals. *Mutation Res.*, 48:267-270.
- (6) Anderson D., Hughes, J.A., Edwards, A.J. and Brinkworth, M.H. (1998). A Comparison of Male-Mediated Effects in Rats and Mice Exposed to 1,3-Butadiene. *Mutation Res.*, 397:77-74.
- (7) Shively C.A., C.A., White, D.M., Blauch, J.L. and Tarka, S.M. Jr. (1984). Dominant Lethal Testing of Theobromine in Rats. *Toxicol. Lett.* 20:325-329.
- (8) Rao K.S., Cobel-Geard, S.R., Young, J.T., Hanley, T.R. Jr., Hayes, W.C., John, J.A. and Miller, R.R. (1983). Ethyl Glycol Monomethyl Ether II. Reproductive and dominant Lethal Studies in Rats. *Fundam. Appl. Toxicol.*, 3:80-85.
- (9) Brewen J.G., Payne, H.S., Jones, K.P., and Preston, R.J. (1975). Studies on Chemically Induced Dominant Lethality. I. The Cytogenetic Basis of MMS-Induced Dominant Lethality in Post-Meiotic Male Germ Cells, *Mutation Res.*, 33, 239-249.
- (10) Marchetti F., Bishop, J.B., Cosentino, L., Moore II, D. and Wyrobek, A.J. (2004). Paternally Transmitted Chromosomal Aberrations in Mouse Zygotes Determine their Embryonic Fate. *Biol. Reprod.*, 70:616-624.
- (11) Marchetti F. and Wyrobek, A.J. (2005). Mechanisms and Consequences of Paternally Transmitted Chromosomal Aberrations. *Birth Defects Res.*, C 75:112-129.
- (12) Adler I.D. (1996). Comparison of the Duration of Spermatogenesis Between Rodents and Humans. *Mutation Res.*, 352:169-172.
- (13) Favor J., and Crenshaw J.W. (1978). EMS-Induced Dominant Lethal Dose Response Curve in DBA/1J Male Mice, *Mutation Res.*, 53: 21-27.
- (14) Generoso W.M., Witt, K.L., Cain, K.T., Hughes, L. Cacheiro, N.L.A, Lockhart, A.M.C. and Shelby, M.D. (1995). Dominant Lethal and Heritable Translocation Test with Chlorambucil and Melphalan. *Mutation Res.*, 345:167-180.
- (15) astings S.E., Huffman K.W. and Gallo M.A. (1976). The dominant Lethal Effect of Dietary Triethylenemelamine, *Mutation Res.*, 40:371-378.
- (16) James D.A. and Smith D.M. (1982). Analysis of Results from a Collaborative Study of the Dominant Lethal Assay, *Mutation Res.*, 99:303-314.
- (17) Shelby M.D., Cain, K.T., Hughes, L.A., Braden, P.W. and Generoso, W.M. (1986). Dominant Lethal Effects of Acrylamide in Male Mice. *Mutation Res.*, 173:35-40.

- (18) Sudman P.D., Rutledge, J.C., Bishop, J.B. and Generoso W.M. (1992). Bleomycin: Female-Specific Dominant Lethal Effects in Mice, *Mutation Res.*, 296: 143-156.
- (19) Holstrom L.M., Palmer A.K. and Favor, J. (1993). The Rodent Dominant Lethal Assay. In *Supplementary Mutagenicity Tests*. Kirkland D.J. and Fox M. (Eds.), Cambridge University Press, pp. 129-156.
- (20) Adler I-D., Bootman, J., Favor, J., Hook, G., Schriever-Schwemmer, G., Welzl, G., Whorton, E., Yoshimura, I. and Hayashi, M. (1998). Recommendations for Statistical Designs of *In Vivo* Mutagenicity Tests with Regard to Subsequent Statistical Analysis, *Mutation Res.*, 417:19–30.
- (21) Adler I.D., Shelby M. D., Bootman, J., Favor, J., Generoso, W., Pacchierotti, F., Shibuya, T. and Tanaka N. (1994). International Workshop on Standardisation of Genotoxicity Test Procedures. Summary Report of the Working Group on Mammalian Germ Cell Tests. *Mutation Res.*, 312:313-318.
- (22) Generoso W.M. and Piegorsch W.W. (1993). Dominant Lethal Tests in Male and Female Mice. *Methods, Toxicol.*, 3A:124-141.
- (23) Haseman J.K. and Soares E.R. (1976). The Distribution of Fetal Death in Control Mice and its Implications on Statistical Tests for Dominant Lethal Effects. *Mutation. Res.*, 41: 277-288.
- (24) Whorton E.B. Jr. (1981). Parametric Statistical Methods and Sample Size Considerations for Dominant Lethal Experiments. The Use of Clustering to Achieve Approximate Normality, *Teratogen. Carcinogen. Mutagen.*, 1:353 – 360.
- (25) Anderson D., Anderson, D., Hodge, M.C.E., Palmer, S., and Purchase, I.F.H. (1981). Comparison of Dominant Lethal and Heritable Translocation Methodologies. *Mutation. Res.*, 85:417-429.
- (26) Fielder R. J., Allen, J. A., Boobis, A. R., Botham, P. A., Doe, J., Esdaile, D. J., Gatehouse, D. G., Hodson-Walker, G., Morton, D. B., Kirkland, D. J. and Richold, M. (1992). Report of British Toxicology Society/UK Environmental Mutagen Society Working Group: Dose Setting in *In Vivo* Mutagenicity Assays. *Mutagen.*, 7:313-319.
- (27) OECD (2000). Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Environment, Health and Safety Publications, Series on Testing and Assessment (No.19.), Organisation for Economic Cooperation and Development, Paris.
- (28) Barrow M.V., Taylor W.J and Morphol J. (1969). A Rapid Method for Detecting Malformations in Rat Fetuses, 127, 291–306.
- (29) Kirkland D.J., (Ed.)(1989). *Statistical Evaluation of Mutagenicity Test Data*, Cambridge University Press.
- (30) Hayashi, M., Dearfield, K., Kasper P., Lovell D., Martus H.-J. and Thybaud V. (2011). “Compilation and Use of Genetic Toxicity Historical Control Data”, *Mutation. Res.*, 723:87-90.
- (31) Lockhart A.C., Piegorsch W.W. and Bishop J.B. (1992). Assessing Over Dispersion and Dose-Response in the Male Dominant Lethal Assay. *Mutation. Res.*, 272:35-58.
- (32) Cochran W.G. (1954). Some Methods for Strengthening the Common χ^2 Tests. *Biometrics*, 10: 417-451.

-
- (33) Tarone R.E. (1979). Testing the Goodness of Fit of the Binomial Distribution. *Biometrika*, 66: 585-590.
- (34) Margolin B.H. (1988). Test for Trend in Proportions. In *Encyclopedia of Statistical Sciences*, Volume 9, Kotz S. and Johnson N. L. (Eds.), pp. 334-336. John Wiley and Sons, New York.
- (35) Cox D.R., Analysis of Binary Data. Chapman and Hall, London (1970).
- (36) Neter J.M., Kutner, H.C., Nachtsheim, J. and Wasserman, W. (1996). Applied Linear Statistical Models, Fourth Edition, Chapters 14 and 17. McGraw-Hill, Boston
- (37) Jonckheere R. (1954). A Distribution-Free K-Sample Test Against Ordered Alternatives. *Biometrika*, 41:133-145.
- (38) Conover W.J. (1971). Practical Nonparametric Statistics. John Wiley and Sons, New York
- (39) Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- (40) Fleiss J. (1973). Statistical Methods for Rates and Proportions. John Wiley and Sons, New York.

Appendix 1

DEFINITIONS

Chemical: A substance or a mixture

Corpora luteum (lutea): the hormonal secreting structure formed on the ovary at the site of a follicle that has released the egg. The number of corpora lutea in the ovaries corresponds to the number of eggs that were ovulated.

Dominant Lethal Mutation: a mutation occurring in a germ cell, or is fixed after fertilisation, that causes embryonic or foetal death.

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

Mating interval: the time between the end of exposure and mating of treated males. By controlling this interval, chemical effects on different germ cell types can be assessed. In the mouse mating during the 1, 2, 3, 4, 5, 6, 7 and 8 week after the end of exposure measures effects in sperm, condensed spermatids, round spermatids, pachytene spermatocytes, early spermatocytes, differentiated spermatogonia, differentiating spermatogonia and stem cell spermatogonia.

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

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

Test chemical: Any substance or mixture tested using this test method.

UVCB: Chemical Substance of Unknown or Variable Composition, Complex Reaction Products and Biological Materials

Appendix 2

TIMING OF SPERMATOGENESIS IN MAMMALS

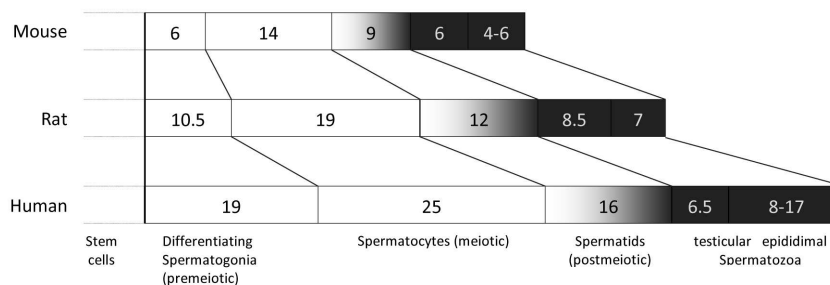


Fig.1: Comparison of the duration (days) of male germ cell development in mice, rats and humans. DNA repair does not occur during the periods indicated by shading.

A schematic of spermatogenesis in the mouse, rat and human is shown above (taken from Adler, 1996). Undifferentiated spermatogonia include: A-single; A-paired; and A-aligned spermatogonia (Hess and de Franca, 2008). A-single is considered the true stem cells; therefore, to assess effects on stem cells at least 49 days (in the mouse) must pass between the last injection of the test chemical and mating.

References

Adler, ID (1996). Comparison of the duration of spermatogenesis between rodents and humans. *Mutat Res*, 352:169-172.

Hess, RA, De Franca LR (2008). Spermatogenesis and cycle of the seminiferous epithelium. In: *Molecular Mechanisms in Spermatogenesis*, C. Yan Cheng (Ed), Landes Biosciences and Springer Science&Business Media:1-15."

- (4) In Part B, Chapter B.23 is replaced by the following:

"B.23 MAMMALIAN SPERMATOGONIAL CHROMOSOMAL ABERRATION TEST

INTRODUCTION

1. This test method (TM) is equivalent to the OECD test guideline 483 (2016). Test methods are periodically reviewed in the light of scientific progress, changing regulatory needs, and animal welfare considerations. This modified version of the test method reflects many years of experience with this assay and the potential for integrating or combining this test with other toxicity or genotoxicity studies. Combining toxicity studies has the potential to reduce the numbers of animals used in toxicity testing. This test method is part of a series of test methods on genetic toxicology. A document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to genetic toxicity OECD test guidelines has been developed by OECD (1).
2. The purpose of the *in vivo* mammalian spermatogonial chromosomal aberration test is to identify those chemicals that cause structural chromosomal aberrations in mammalian spermatogonial cells (2) (3) (4). In addition, this test is relevant to assessing genotoxicity because, although they may vary among species, factors of *in vivo* metabolism, pharmacokinetics and DNA-repair processes are active and contribute to the response. This test method is not designed to measure numerical abnormalities; the assay is not routinely used for this purpose.
3. This test measures structural chromosomal aberrations (both chromosome- and chromatid-type) in dividing spermatogonial germ cells and is, therefore, expected to be predictive of induction of heritable mutations in these germ cells.
4. Definitions of key terms are set out in the Appendix.

INITIAL CONSIDERATIONS

5. Rodents are routinely used in this test but other species may in some cases be appropriate if scientifically justified. Standard cytogenetic preparations of rodent testes generate mitotic (spermatogonia) and meiotic (spermatocyte) metaphases. Mitotic and meiotic metaphases are identified based on the morphology of the chromosomes (4). This *in vivo* cytogenetic test detects structural chromosomal aberrations in spermatogonial mitoses. Other target cells are not the subject of this test method.
6. To detect chromatid-type aberrations in spermatogonial cells, the first mitotic cell division following treatment should be examined before these aberrations are converted into chromosome-type-aberrations in subsequent cell divisions. Additional information from treated spermatocytes can be obtained by meiotic chromosome analysis for chromosomal structural aberrations at diakinesis-metaphase I and metaphase II.
7. A number of generations of spermatogonia are present in the testis (5), and these different germ cell types may have a spectrum of sensitivity to chemical treatment. Thus, the aberrations detected represent an aggregate response of treated spermatogonial cell populations. The majority of mitotic cells in testis preparations are B spermatogonia, which have a cell cycle of approximately 26 hr (3).
8. If there is evidence that the test chemical, or its metabolite(s), will not reach the testis it is not appropriate to use this test.

PRINCIPLE OF THE TEST METHOD

9. Generally, animals are exposed to the test chemical by an appropriate route of exposure and are euthanised at appropriate times after treatment. Prior to euthanasia, animals are treated with a metaphase-arresting agent (e.g. colchicine or Colcemid®). Chromosome preparations are then made from germ cells and stained, and metaphase cells are analysed for chromosome aberrations.

VERIFICATION OF LABORATORY PROFICIENCY

10. Competency in this assay should be established by demonstrating the ability to reproduce expected results for structural chromosomal aberration frequencies in spermatogonia with positive control substances (including weak responses) such as those listed in Table 1 and obtaining negative control frequencies that are consistent with acceptable range of control data in the published literature (e.g. (2)(3)(6)(7)(8)(9)(10)) or with the laboratory's historical control distribution, if available.

DESCRIPTION OF THE METHOD

Preparations*Selection of animal species*

11. Commonly used laboratory strains of healthy young adult animals should be employed. Male mice are commonly used; however, males of other appropriate mammalian species may be used when scientifically justified and to allow this test to be run in conjunction with another test method. The scientific justification for using species other than rodents should be provided in the report.

Animal Housing and feeding conditions

12. For rodents, the temperature in the animal room should be 22 °C (\pm 3 °C). Although the relative humidity ideally should be 50-60 %, it should be at least 40 % and preferably not exceed 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 an unlimited supply of drinking water. The choice of diet may be influenced by the need to ensure a suitable admixture of a test chemical when administered by this route. Rodents should be housed in small groups (no more than five per cage) if no aggressive behaviour is expected, preferably in solid floor cages with appropriate environmental enrichment. Animals may be housed individually if scientifically justified.

Preparation of the animals

13. Healthy young adult male animals (8-12 weeks old at start of treatment) are normally used, and are randomly assigned to the control and treatment groups. The individual animals are identified uniquely using a humane, minimally invasive method (e.g. by ringing, tagging, micro-chipping or biometric identification, but not ear or toe clipping) and acclimated to the laboratory conditions for at least five days. Cages should be arranged in such a way that possible effects due to cage placement are minimised. Cross contamination by the positive control and test chemical should be avoided. At the commencement of the study, the variation between individual animal weights should be minimal and not exceed \pm 20 %.

Preparation of doses

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

Test conditions - Solvent/vehicle

15. The solvent/vehicle should not produce toxic effects at the dose levels used, and should not be capable of chemical reaction with the test chemicals. If other than well-known solvents/vehicles are used, their inclusion should be supported with reference data indicating their compatibility. It is recommended that, wherever possible, the use of an aqueous solvent/vehicle should be considered first. Examples of commonly used compatible solvents/vehicles include water, physiological saline, methylcellulose solution, carboxymethyl cellulose sodium salt solution, olive oil and corn oil. In the absence of historical or published control data showing that no structural chromosomal aberrations and other deleterious effects are induced by a chosen atypical solvent/vehicle, an initial study should be conducted in order to establish the acceptability of the solvent/vehicle control.

Positive controls

16. Concurrent positive control animals should always be used unless the laboratory has demonstrated proficiency in the conduct of the test and has used the test routinely in the recent past (e.g. within the last 5 years). When a concurrent positive control group is not included, scoring controls (fixed and unstained slides) should be included in each experiment. These can be obtained by including within the scoring of the study appropriate reference samples that have been obtained and stored from a separate positive control experiment conducted periodically (e.g. every 6-18 months) in the laboratory where the test is performed; for example, during proficiency testing and on a regular basis thereafter, where necessary.
17. Positive control substances should reliably produce a detectable increase in the frequencies of cells with structural chromosomal aberrations over the spontaneous levels. Positive control doses should be chosen so that the effects are clear but do not immediately reveal the identity of the coded samples to the scorer. Examples of positive control substances are included in Table 1.

Table 1

Examples of positive control substances.

Substances [CAS No] (reference no)
Cyclophosphamide (monohydrate) [CAS no. 50-18-0 (CAS no. 6055-19-2)] (9)
Cyclohexylamine [CAS no. 108-91-8] (7)
Mitomycin C [CAS no. 50-07-7] (6)
Monomeric acrylamide [CAS 79-06-1] (10)
Triethylenemelamine [CAS 51-18-3] (8)

Negative controls

18. Negative control animals, treated with solvent or vehicle alone, and otherwise treated in the same way as the treatment groups, should be included for every sampling time. In the absence of historical or published control data showing that no chromosomal aberrations or other deleterious effects are induced by the chosen solvent/vehicle, untreated control animals also should be included for every sampling time in order to establish acceptability of the vehicle control.

PROCEDURE

Number of animals

19. Group sizes at study initiation should be established with the aim of providing a minimum of 5 male animals per group. This number of animals per group is considered to be sufficient to provide adequate statistical power (i.e. generally able to detect at least a doubling in chromosomal aberration frequency when the negative control level is 1,0 % or greater with 80 % probability at a significance level of 0,05) (3) (11). As a guide to typical maximum animal requirements, a study at two sampling times with three dose groups and a concurrent negative control group, plus a positive control group (each composed of five animals per group), would require 45 animals.

Treatment schedule

20. Test chemicals are usually administered once (i.e. as a single treatment); other dose regimens may be used, provided they are scientifically justified.
21. In the highest dose group two sampling times after treatment are used. Since the time required for uptake and metabolism of the test chemical(s), as well as its effect on cell cycle kinetics, can affect the optimum time for chromosomal aberration detection, one early and one late sampling time approximately 24 and 48 hours after treatment are used. For doses other than the highest dose, an early sampling time of 24 hours (less than or equal to the cell cycle time of B spermatogonia and thus optimising the probability of scoring first post-treatment metaphases) after treatment should be taken, unless another sampling time is known to be more appropriate and justified.
22. Other sampling times may be used. For example in the case of chemicals that exert S-independent effects, earlier sampling times (i.e. less than 24 hr) may be appropriate.
23. A repeat dose treatment regimen can be used, such as in conjunction with a test on another endpoint that uses a 28 day administration period (e.g., TM B.58); however, additional animal groups would be required to accommodate different sampling times. Accordingly, the appropriateness of such a schedule needs to be justified scientifically on a case-by-case basis.
24. Prior to euthanasia, animals are injected intraperitoneally with an appropriate dose of a metaphase arresting chemical (e.g. Colcemid® or colchicine). Animals are sampled at an appropriate interval thereafter. For mice and rats, this interval is approximately 3 - 5 hours.

Dose levels

25. If a preliminary range-finding study is performed because there are no suitable data already available to aid in dose selection, it should be performed in the same laboratory, using the same species, strain, and treatment regimen to be used in the main study, according to recommendations for conducting dose range-finding studies (12). This study should aim to identify the maximum tolerated dose (MTD), defined as the dose inducing slight toxic effects relative to the duration of the study period (for example, abnormal behaviour or reactions, minor body weight depression or hematopoietic system cytotoxicity) but not death or evidence of pain, suffering or distress necessitating euthanasia of the animals (13).
26. The highest dose may also be defined as a dose that produces some indication of toxicity in the spermatogonial cells (e.g. a reduction in the ratio of spermatogonial mitoses to first and second meiotic metaphases). This reduction should not exceed 50 %.

27. Test chemicals with specific biological activities at low non-toxic doses (such as hormones and mitogens), and chemicals which exhibit saturation of toxicokinetic properties may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis.

28. In order to obtain dose response information, a complete study should include a negative control group (paragraph 18) and a minimum of three dose levels generally separated by a factor of 2, but by no greater than 4. If the test chemical does not produce toxicity in a range-finding study or based on existing data, the highest dose for a single administration should be 2 000 mg/kg body weight. However, if the test chemical does cause toxicity, the MTD should be the highest dose administered, and the dose levels used should preferably cover a range from the maximum to a dose producing little or no toxicity. When target tissue (i.e. testis) toxicity is observed at all dose levels tested, further study at non-toxic doses is advisable. Studies intending to more fully characterise the quantitative dose-response information may require additional dose groups. For certain types of test chemicals (e.g. human pharmaceuticals) covered by specific requirements, these limits may vary. If the test chemical does produce toxicity, the limit dose plus two lower doses (as described above) should be selected. The limit dose for an administration period of 14 days or more is 1 000 mg/kg body weight/day, and for administration periods of less than 14 days, the limit dose is 2 000 mg/kg/body weight/day.

Administration of doses

29. The anticipated route of human exposure should be considered when designing an assay. Therefore, routes of exposure such as dietary, drinking water, topical subcutaneous, intravenous, oral (by gavage), inhalation, or implantation may be chosen as justified. In any case, the route should be chosen to ensure adequate exposure of the target tissue. Intraperitoneal injection is not normally recommended unless scientifically justified since it is not usually a physiologically relevant route of human exposure. If the test chemical is admixed in diet or drinking water, especially in case of single dosing, care should be taken that the delay between food and water consumption and sampling should be sufficient to allow detection of the effects (see paragraph 33). 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/100g body weight except in the case of aqueous solutions where a maximum of 2 ml/100g body weight may be used. The use of volumes greater than this (if permitted by animal welfare legislation) should be justified. Variability in test volume should be minimised by adjusting the concentration to ensure a constant volume in relation to body weight at all dose levels.

Observations

30. General clinical observations of the test animals should be made and clinical signs recorded at least once a day, preferably at the same time(s) each day and considering the peak period of anticipated effects after dosing. At least twice daily, all animals should be observed for morbidity and mortality. All animals should be weighed at study initiation, at least once a week during repeated-dose studies, and at euthanasia. In studies of at least one-week duration, measurements of food consumption should be made at least weekly. If the test chemical is administered via the drinking water, water consumption should be measured at each change of water and at least weekly. Animals exhibiting non-lethal indicators of excess toxicity should be euthanised prior to completion of the test period (13).

Chromosome preparation

31. Immediately after euthanasia, germ cell suspensions are obtained from one, or both, testes, exposed to hypotonic solution and fixed following established protocols (e.g. (2) (14) (15)). The cells are then spread on slides and stained (16) (17). All slides should be coded so that their identity is not available to the scorer.

Analysis

32. At least 200 well spread metaphases should be scored for each animal (3) (11). If the historical negative control frequency is < 1 %, more than 200 cells/animal should be scored to increase the statistical power (3). Staining methods that permit the identification of the centromere should be used.

33. Chromosome and chromatid-type aberrations should be recorded separately and classified by sub-types (breaks, exchanges). Gaps should be recorded, but not considered, when determining whether a chemical induces significant increases in the incidence of cells with chromosomal aberrations. Procedures in use in the laboratory should ensure that analysis of chromosomal aberrations is performed by well-trained scorers. Recognising that slide preparation procedures often result in the breakage of a proportion of metaphases with a resulting loss of chromosomes, the cells scored should, therefore, contain a number of centromeres not less than $2n \pm 2$, where n is the haploid number of chromosomes for that species.
34. Although the purpose of the test is to detect structural chromosomal aberrations, it is important to record the frequencies of polyploid cells and cells with endoreduplicated chromosomes when these events are seen (see Paragraph 44).

DATA AND REPORTING

Treatment of results

35. Individual animal data should be presented in tabular form. For each animal the number of cells with structural chromosomal aberration(s) and the number of chromosome aberrations per cell should be evaluated. Chromatid- and chromosome-type aberrations classified by sub-types (breaks, exchanges) should be listed separately with their numbers and frequencies for experimental and control groups. Gaps are recorded separately. The frequency of gaps is reported but generally not included in the analysis of the total structural chromosomal aberration frequency. Percentage of polyploidy and cells with endoreduplicated chromosomes are reported when seen.
36. Data on toxicity and clinical signs (as per Paragraph 30) should be reported.

Acceptability Criteria

37. The following criteria determine the acceptability of a test.
 - Concurrent negative control is consistent with published norms for historical negative control data, which are generally expected to be $> 0\%$ and $\leq 1,5\%$ cells with chromosomal aberrations, and the laboratory's historical control data if available (see Paragraphs 10 and 18).
 - Concurrent positive controls induce responses that are consistent with published norms for historical positive control data, or the laboratory's historical positive control database, if available, and produce a statistically significant increase compared with the negative control (see Paragraphs 17, 18).
 - Adequate numbers of cells and doses have been analysed (see Paragraphs 28 and 32).
 - The criteria for the selection of top dose are consistent with those described in Paragraphs 25, and 26.
38. If both mitosis and meiosis are observed, the ratio of spermatogonial mitoses to first and second meiotic metaphases should be determined as a measure of cytotoxicity for all treated and negative control animals in a total sample of 100 dividing cells per animal. If only mitosis is observed, the mitotic index should be determined in at least 1 000 cells for each animal.

Evaluation and interpretation of results

39. At least three treated dose groups should be analysed in order to provide sufficient data for dose-response analysis.

40. Providing that all acceptability criteria are fulfilled, a test chemical is considered a clear positive if:

- at least one of the test doses exhibits a statistically significant increase compared with the concurrent negative control;
- the increase is dose-related at least at one sampling time; and,
- any of the results are outside acceptable range of negative control data, or the distribution of the laboratory's historical negative control data (e.g. Poisson-based 95 % control limit) if available.

The test chemical is then considered able to induce chromosomal aberrations in spermatogonial cells of the test animals. Recommendations for the most appropriate statistical methods can also be found in the literature (11) (18). Statistical tests used should consider the animal as the experimental unit.

41. Providing that all acceptability criteria are fulfilled, a test chemical is considered a clear negative if:

- none of the test doses exhibits a statistically significant increase compared with the concurrent negative control;
- there is no dose-related increase in any experimental condition; and,
- all results are within acceptable range of negative control data, or the laboratory's historical negative control data (e.g. Poisson-based 95 % control limit), if available.

The test chemical is then considered unable to induce chromosomal aberrations in the spermatogonial cells of the test animals. Recommendations for the most appropriate statistical methods can also be found in the literature (11) (18). A negative result does not exclude the possibility that the chemical may induce chromosomal aberrations at later developmental phases not studied, or gene mutations.

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

43. If the response is not clearly negative or positive, and in order to assist in establishing the biological relevance of a result (e.g. a weak or borderline increase), the data should be evaluated by expert judgment and/or further investigations using the existing experimental data, such as consideration whether the positive result is outside the acceptable range of negative control data, or the laboratory's historical negative control data (19).

44. In rare cases, even after further investigations, the data set will preclude making a conclusion of positive or negative results, and will therefore be concluded as equivocal.

45. An increase in the number of polyploid cells may indicate that the test chemical has the potential to inhibit mitotic processes and to induce numerical chromosomal aberrations (20). An increase in the number of cells with endoreduplicated chromosomes may indicate that the test chemical has the potential to inhibit cell cycle progress (21) (22), which is a different mechanism of inducing numerical chromosome changes than inhibition of mitotic processes (see Paragraph 2). Therefore incidence of polyploid cells and cells with endoreduplicated chromosomes should be recorded separately.

Test report

46. The test report should include the following information:

*Summary.**Test chemical:*

- source, lot number, limit date for use, if available;
- stability of the test chemical itself, if known;
- solubility and stability of the test chemical in solvent, if known;
- measurement of pH, osmolality, and precipitate in the culture medium to which the test chemical was added, as appropriate.

Mono-constituents substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Test chemical preparation:

- justification for choice of vehicle;
- solubility and stability of the test chemical in solvent/vehicle.
- preparation of dietary, drinking water or inhalation formulations;
- analytical determinations on formulations (e.g. stability, homogeneity, nominal concentrations when conducted).

Test animals:

- species/strain used and justification for use;
- number and age of animals;
- source, housing conditions, diet, etc.;

- method for uniquely identifying the animals

- for short-term studies: individual weight of the animals at the start and end of the test; for studies longer than one week: individual body weights during the study and food consumption. Body weight range, mean and standard deviation for each group should be included.

Test conditions:

- positive and negative (vehicle/solvent) control data;

- data from range finding study, if conducted;

- rationale for dose level selection;

- rationale for route of administration;

- details of test chemical preparation;

- details of the administration of the test chemical;

- rationale for sacrifice times;

- methods for measurement of animal toxicity, including, where available, histopathological or hematological analyses and the frequency with which animal observations and body weights were taken;

- methods for verifying that the test chemical reached the target tissue, or general circulation, if negative results are obtained;

- actual dose (mg/kg body weight/day) calculated from diet/drinking water test chemical concentration (ppm) and consumption, if applicable;

- details of food and water quality;

- detailed description of treatment and sampling schedules and justifications for the choices;

- method of euthanasia;

- method of analgesia (where used)

- procedures for isolating tissues;

- identity of metaphase arresting chemical, its concentration and duration of treatment;

- methods of slide preparation;

- criteria for scoring aberrations;
- number of cells analysed per animal;
- criteria for considering studies as positive, negative or equivocal.

Results:

- animal condition prior to and throughout the test period, including signs of toxicity;
- body and organ weights at sacrifice (if multiple treatments are employed, body weights taken during the treatment regimen);
- signs of toxicity;
- mitotic index;
- ratio of spermatogonial mitoses cells to first and second meiotic metaphases, or other evidence of exposure to the target tissue;
- type and number of aberrations, given separately for each animal;
- total number of aberrations per group with means and standard deviations;
- number of cells with aberrations per group with means and standard deviations;
- dose-response relationship, where possible;
- statistical analyses and methods applied;
- concurrent negative control data;
- historical negative control data with ranges, means, standard deviations, and 95 % confidence interval (where available), or published historical negative control data used for acceptability of the test results;
- concurrent positive control data;
- changes in ploidy, if seen, including frequencies of polyploidy and/or endoreduplicated cells.

*Discussion of the results**Conclusion***LITERATURE**

- (1) OECD (2016). Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. ENV Publications. Series on Testing and Assessment, No 234, OECD, Paris
- (2) Adler, I.-D. (1984). Cytogenetic Tests in Mammals. In: Mutagenicity Testing: a Practical Approach. Ed. S. Venitt and J. M. Parry. IRL Press, Oxford, Washington DC, pp. 275-306.
- (3) Adler I.-D., Shelby M. D., Bootman, J., Favor, J., Generoso, W., Pacchierotti, F., Shibuya, T. and Tanaka N. (1994). International Workshop on Standardisation of Genotoxicity Test Procedures. Summary Report of the Working Group on Mammalian Germ Cell Tests. *Mutation Res.*, 312, 313-318.
- (4) Russo, A. (2000). *In Vivo* Cytogenetics: Mammalian Germ Cells. *Mutation Res.*, 455, 167-189.
- (5) Hess, R.A. and de Franca L.R. (2008). Spermatogenesis and Cycle of the Seminiferous Epithelium. In: *Molecular Mechanisms in Spermatogenesis*, Cheng C.Y. (Ed.) Landes Bioscience and Springer Science+Business Media, pp. 1-15.
- (6) Adler, I.-D. (1974). Comparative Cytogenetic Study after Treatment of Mouse Spermatogonia with Mitomycin C, *Mutation. Res.*, 23(3): 368-379. Adler, I.D. (1986). Clastogenic Potential in Mouse Spermatogonia of Chemical Mutagens Related to their Cell-Cycle Specifications. In: *Genetic Toxicology of Environmental Chemicals, Part B: Genetic Effects and Applied Mutagenesis*, Ramel C., Lambert B. and Magnusson J. (Eds.) Liss, New York, pp. 477-484.
- (7) Cattanaach, B.M., and Pollard C.E. (1971). Mutagenicity Tests with Cyclohexylamine in the Mouse, *Mutation Res.*, 12, 472-474.
- (8) Cattanaach, B.M., and Williams, C.E. (1971). A search for Chromosome Aberrations Induced in Mouse Spermatogonia by Chemical Mutagens, *Mutation Res.*, 13, 371-375.
- (9) Rathenburg, R. (1975). Cytogenetic Effects of Cyclophosphamide on Mouse Spermatogonia, *Human-genetik* 29, 135-140.
- (10) Shiraishi, Y. (1978). Chromosome Aberrations Induced by Monomeric Acrylamide in Bone Marrow and Germ Cells of Mice, *Mutation Res.*, 57(3): 313-324.
- (11) Adler I.-D., Bootman, J., Favor, J., Hook, G., Schriever-Schwemmer, G., Welzl, G., Whorton, E., Yoshimura, I. and Hayashi, M. (1998). Recommendations for Statistical Designs of *In Vivo* Mutagenicity Tests with Regard to Subsequent Statistical Analysis, *Mutation Res.*, 417, 19-30.
- (12) Fielder, R. J., Allen, J. A., Boobis, A. R., Botham, P. A., Doe, J., Esdaile, D. J., Gatehouse, D. G., Hodson-Walker, G., Morton, D. B., Kirkland, D. J. and Richold, M. (1992). Report of British Toxicology Society/UK Environmental Mutagen Society Working Group: Dose setting in *In Vivo* Mutagenicity Assays. *Mutagenesis*, 7, 313-319.

- (13) OECD (2000). Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation, Series on Testing and Assessment, (No 19.), Organisation for Economic Cooperation and Development, Paris.
- (14) Yamamoto, K. and Kikuchi, Y. (1978). A New Method for Preparation of Mammalian Spermatogonial Chromosomes. *Mutation Res.*, 52, 207-209.
- (15) Hsu, T.C., Elder, F. and Pathak, S. (1979). Method for Improving the Yield of Spermatogonial and Meiotic Metaphases in Mammalian Testicular Preparations. *Environ. Mutagen.*, 1, 291-294.
- (16) Evans, E.P., Breckon, G., and Ford, C.E. (1964). An Air-Drying Method for Meiotic Preparations from Mammalian Testes. *Cytogenetics and Cell Genetics*, 3, 289-294.
- (17) Richold, M., Ashby, J., Bootman, J., Chandley, A., Gatehouse, D.G. and Henderson, L. (1990). *In Vivo* Cytogenetics Assays, In: D.J. Kirkland (Ed.) Basic Mutagenicity Tests, UKEMS Recommended Procedures. UKEMS Subcommittee on Guidelines for Mutagenicity Testing. Report. Part I revised. Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney, pp. 115-141.
- (18) Lovell, D.P., Anderson, D., Albanese, R., Amphlett, G.E., Clare, G., Ferguson, R., Richold, M., Papworth, D.G. and Savage, J.R.K. (1989). Statistical Analysis of *In Vivo* Cytogenetic Assays In: D.J. Kirkland (Ed.) Statistical Evaluation of Mutagenicity Test Data. UKEMS Sub-Committee on Guidelines for Mutagenicity Testing, Report, Part III. Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney, pp. 184-232.
- (19) Hayashi, M., Dearfield, K., Kasper, P., Lovell, D., Martus, H.-J. and Thybaud, V. (2011). Compilation and Use of Genetic Toxicity Historical Control Data. *Mutation Res.*, 723, 87-90.
- (20) Warr T.J., Parry E.M. and Parry J.M. (1993). A Comparison of Two *In Vitro* Mammalian Cell Cytogenetic Assays for the Detection of Mitotic Aneuploidy Using 10 Known or Suspected Aneugens, *Mutation Res.*, 287, 29-46.
- (21) Huang, Y., Change, C. and Trosko, J.E. (1983). Aphidicolin-Induced Endoreduplication in Chinese Hamster Cells. *Cancer Res.*, 43, 1362-1364.
- (22) Locke-Huhle, C. (1983). Endoreduplication in Chinese Hamster Cells during Alpha-Radiation Induced G2 Arrest. *Mutation Res.*, 119, 403-413.

Appendix

DEFINITIONS

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

Centromere: Region(s) of a chromosome with which spindle fibers are associated during cell division, allowing orderly movement of daughter chromosomes to the poles of the daughter cells.

Chemical: A substance or a mixture

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

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

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

Clastogen: any chemical which causes structural chromosomal aberrations in populations of cells or organisms.

Gap: an achromatic lesion smaller than the width of one chromatid, and with minimum misalignment of the chromatids.

Genotoxic: a general term encompassing all types of DNA or chromosome damage, including breaks, deletions, adducts, nucleotides modifications and linkages, rearrangements, mutations, chromosome aberrations, and aneuploidy. Not all types of genotoxic effects result in mutations or stable chromosome damage.

Mitotic index (MI): the ratio of cells in metaphase divided by the total number of cells observed in a population of cells; an indication of the degree of proliferation of that population.

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

Mutagenic: produces a heritable change of DNA base-pair sequence(s) in genes or of the structure of chromosomes (chromosome aberrations).

Numerical abnormality: a change in the number of chromosomes from the normal number characteristic of the animals utilised.

Polyploidy: a multiple of the haploid chromosome number (n) other than the diploid number (i.e., $3n$, $4n$ and so on).

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

Test chemical: Any substance or mixture tested using this test method.

UVCB: Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials"

- (5) In Part B, Chapter B.40 is replaced by the following:

"B.40 IN VITRO SKIN CORROSION: TRANSCUTANEOUS ELECTRICAL RESISTANCE TEST METHOD (TER)

INTRODUCTION

1. This test method (TM) is equivalent to OECD test guideline (TG) 430 (2015). Skin corrosion refers to the production of irreversible damage to the skin manifested as visible necrosis through the epidermis and into the dermis, following the application of a test chemical [as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (1) and the European Union (EU) Regulation 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) ⁽¹⁾]. This updated test method B.40 provides an *in vitro* procedure allowing the identification of non-corrosive and corrosive substances and mixtures in accordance with UN GHS (1) and CLP
2. The assessment of skin corrosivity has typically involved the use of laboratory animals (TM B.4, equivalent to OECD TG 404 originally adopted in 1981, and revised in 1992, 2002 and 2015) (2). In addition to the present TM B.40, other *in vitro* test methods for testing of skin corrosion potential of chemicals have been validated and adopted as TM B.40bis (equivalent to OECD TG 431) (3) and TM B.65 (equivalent to OECD TG 435) (4), that are also able to identify sub-categories of corrosive chemicals when required. Several validated *in vitro* test methods have been adopted as TM B.46 (equivalent to OECD TG 439 (5), to be used for the testing of skin irritation. An OECD guidance document on Integrated Approaches to Testing and Assessment (IATA) for Skin Corrosion and Irritation describes several modules which group various information sources and analysis tools and provides guidance on (i) how to integrate and use existing testing and non-testing data for the assessment of skin irritation and skin corrosion potentials of chemicals and (ii) proposes an approach when further testing is needed (6).
3. This test method addresses the human health endpoint skin corrosion. It is based on the rat skin transcutaneous electrical resistance (TER) test method, which utilises skin discs to identify corrosives by their ability to produce a loss of normal *stratum corneum* integrity and barrier function. The corresponding OECD test guideline was originally adopted in 2004 and updated in 2015 to refer to the IATA guidance document.
4. In order to evaluate *in vitro* skin corrosion testing for regulatory purposes, pre-validation studies (7) followed by a formal validation study of the rat skin TER test method for assessing skin corrosion were conducted (8) (9) (10) (11). The outcome of these studies led to the recommendation that the TER test method (designated the Validated Reference Method – VRM) could be used for regulatory purposes for the assessment of *in vivo* skin corrosivity (12) (13) (14).
5. Before a proposed similar or modified *in vitro* TER test method for skin corrosion other than the VRM can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined to ensure its similarity to the VRM, in accordance with the requirements of the Performance Standards (PS) (15). OECD Mutual Acceptance of Data will only be guaranteed after any proposed new or updated test method following the PS have been reviewed and included in the corresponding OECD test guideline.

DEFINITIONS

6. Definitions used are provided in the Appendix.

INITIAL CONSIDERATIONS

7. A validation study (10) and other published studies (16) (17) have reported that the rat skin TER test method is able to discriminate between known skin corrosives and non-corrosives with an overall sensitivity of 94 % (51/54) and specificity of 71 % (48/68) for a database of 122 substances.

⁽¹⁾ Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006, OJ L 353/1, 31.12.2008

8. This test method addresses *in vitro* skin corrosion. It allows the identification of non-corrosive and corrosive test chemicals in accordance with the UN GHS/CLP. A limitation of this test method, as demonstrated by the validation studies (8) (9) (10) (11), is that it does not allow the sub-categorisation of corrosive substances and mixtures in accordance with the UN GHS/ CLP. The applicable regulatory framework will determine how this test method will be used. While this test method does not provide adequate information on skin irritation, it should be noted that TM B.46 specifically addresses the health effect skin irritation *in vitro* (5). For a full evaluation of local skin effects after a single dermal exposure, the OECD Guidance Document on IATA should be consulted (6).

9. A wide range of chemicals representing mainly substances has been tested in the validation underlying this test method and the empirical database of the validation study amounted to 60 substances covering a wide range of chemical classes (8) (9). On the basis of the overall data available, the test method is applicable to a wide range of chemical classes and physical states including liquids, semi-solids, solids and waxes. However, since for specific physical states test items with suitable reference data are not readily available, it should be noted that a comparably small number of waxes and corrosive solids were assessed during validation. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. In cases where evidence can be demonstrated on the non-applicability of the test method to a specific category of substances, the test method should not be used for that specific category of substances. In addition, this test method is assumed to be applicable to mixtures as an extension of its applicability to substances. However, due to the fact that mixtures cover a wide spectrum of categories and composition, and that only limited information is currently available on the testing of mixtures, in cases where evidence can be demonstrated on the non-applicability of the test method to a specific category of mixtures (e.g. following a strategy as proposed by Eskes *et al.*, 2012) (18), the test method should not be used for that specific category of mixtures. Before use of the test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture. Gases and aerosols have not been assessed yet in validation studies (8) (9). While it is conceivable that these can be tested using the TER test method, the current test method does not allow testing of gases and aerosols.

PRINCIPLE OF THE TEST

10. The test chemical is applied for up to 24 hours to the epidermal surfaces of skin discs in a two-compartment test system in which the skin discs function as the separation between the compartments. The skin discs are taken from humanely killed rats aged 28-30 days. Corrosive chemicals are identified by their ability to produce a loss of normal *stratum corneum* integrity and barrier function, which is measured as a reduction in the TER below a threshold level (16) (see paragraph 32). For rat skin TER, a cut-off value of 5k has been selected based on extensive data for a wide range of substances where the vast majority of values were either clearly well above (often > 10 k), or well below (often < 3 k) this value (16). Generally, test chemicals that are non-corrosive in animals but are irritant or non-irritant do not reduce the TER below this cut-off value. Furthermore, use of other skin preparations or other equipment may alter the cut-off value, necessitating further validation.

11. A dye-binding step is incorporated into the test procedure for confirmation testing of positive results in the TER including values around 5 k. The dye-binding step determines if the increase in ionic permeability is due to physical destruction of the *stratum corneum*. The TER method utilising rat skin has shown to be predictive of *in vivo* corrosivity in the rabbit assessed under TM B.4 (2).

DEMONSTRATION OF PROFICIENCY

12. Prior to routine use of the rat skin TER test method that adheres to this test method, laboratories should demonstrate technical proficiency by correctly classifying the twelve Proficiency Substances recommended in Table 1. In situations where a listed substance is unavailable or where justifiable, another substance for which adequate *in vivo* and *in vitro* reference data are available may be used (e.g. from the list of reference chemicals (16)) provided that the same selection criteria as described in Table 1 is applied.

Table 1
List of Proficiency Substances ⁽¹⁾

Substance	CASRN	Chemical Class ⁽²⁾	UN GHS/CLP Cat. Based on <i>In Vivo</i> Results ⁽³⁾	VRM Cat. Based on <i>In Vitro</i> Results	Physical State	pH ⁽⁴⁾
<i>In Vivo</i> Corrosives						
N,N'-Dimethyldipropylenetriamine	10563-29-8	organic base	1A	6 × C	L	8,3
1,2-Diaminopropane	78-90-0	organic base	1A	6 × C	L	8,3
Sulfuric acid (10 %)	7664-93-9	inorganic acid	(1A/1B/1C)	5 × C 1 × NC	L	1,2
Potassium hydroxide (10 % aq.)	1310-58-3	inorganic base	(1A/1B/1C)	6 × C	L	13,2
Octanoic (Caprylic) acid	124-07-2	organic acid	1B/1C	4 × C 2 × NC	L	3,6
2-tert-Butylphenol	88-18-6	phenol	1B/1C	4 × C 2 × NC	L	3,9
<i>In Vivo</i> Non-corrosives						
Isostearic acid	2724-58-5	organic acid	NC	6 × NC	L	3,6
4-Amino-1,2,4-triazole	584-13-4	organic base	NC	6 × NC	S	5,5
Phenethyl bromide	103-63-9	electrophile	NC	6 × NC	L	3,6
4-(Methylthio)-benzaldehyde	3446-89-7	electrophile	NC	6 × NC	L	6,8
1,9-Decadiene	1647-16-1	neutral organic	NC	6 × NC	L	3,9
Tetrachloroethylene	127-18-4	neutral organic	NC	6 × NC	L	4,5

Abbreviations: aq = aqueous; CASRN = Chemical Abstracts Service Registry Number; VRM = Validated Reference Method; C = corrosive; NC = not corrosive.

⁽¹⁾ The proficiency substances, sorted first by corrosives versus non-corrosives, then by corrosive subcategory and then by chemical class, were selected from the substances used in the ECVAM validation study of the rat skin TER test method (8) (9). Unless otherwise indicated, the substances were tested at the purity level obtained when purchased from a commercial source (8). The selection included, to the extent possible, substances that: (i) are representative of the range of corrosivity responses (e.g. non-corrosives; weak to strong corrosives) that the VRM is capable of measuring or predicting; (ii) are representative of the chemical classes used in the validation study; (iii) reflect the performance characteristics of the VRM; (iv) have chemical structures that are well-defined; (v) induce definitive results in the *in vivo* reference test method; (vi) are commercially available; and (vii) are not associated with prohibitive disposal costs.

⁽²⁾ Chemical class assigned by Barratt *et al.* (8).

⁽³⁾ The corresponding UN Packing groups are I, II and III, respectively, for the UN GHS/CLP categories 1A, 1B and 1C.

⁽⁴⁾ The pH values were obtained from Fentem *et al.* (9) and Barratt *et al.* (8).

PROCEDURE

13. Standard Operating Procedures (SOP) for the rat skin TER skin corrosion test method are available (19). The rat skin TER test methods covered by this test method should comply with the following conditions:

Animals

14. Rats should be used because the sensitivity of their skin to substances in this test method has been previously demonstrated (12) and is the only skin source that has been formally validated (8) (9). The age (when the skin is collected) and strain of the rat is particularly important to ensure that the hair follicles are in the dormant phase before adult hair growth begins.
15. The dorsal and flank hair from young, approximately 22 day-old, male or female rats (Wistar-derived or a comparable strain), is carefully removed with small clippers. Then, the animals are washed by careful wiping, whilst submerging the clipped area in antibiotic solution (containing, for example, streptomycin, penicillin, chloramphenicol, and amphotericin, at concentrations effective in inhibiting bacterial growth). Animals are washed with antibiotics again on the third or fourth day after the first wash and are used within 3 days of the second wash, when the *stratum corneum* has recovered from the hair removal.

Preparation of the skin discs

16. Animals are humanely killed when 28-30 days old; this age is critical. The dorso-lateral skin of each animal is then removed and stripped of excess subcutaneous fat by carefully peeling it away from the skin. Skin discs, with a diameter of approximately 20-mm each, are removed. The skin may be stored before discs are used where it is shown that positive and negative control data are equivalent to that obtained with fresh skin.
17. Each skin disc is placed over one of the ends of a PTFE (polytetrafluoroethylene) tube, ensuring that the epidermal surface is in contact with the tube. A rubber 'O' ring is press-fitted over the end of the tube to hold the skin in place and excess tissue is trimmed away. The rubber 'O' ring is then carefully sealed to the end of the PTFE tube with petroleum jelly. The tube is supported by a spring clip inside a receptor chamber containing MgSO_4 solution (154 mM) (Figure 1). The skin disc should be fully submerged in the MgSO_4 solution. As many as 10-15 skin discs can be obtained from a single rat skin. Tube and 'O' ring dimensions are shown in Figure 2.
18. Before testing begins, the TER of two skin discs are measured as a quality control procedure for each animal skin. Both discs should give electrical resistance values greater than 10 k for the remainder of the discs to be used for the test method. If the resistance value is less than 10 k, the remaining discs from that skin should be discarded.

Application of the test chemical and control substances

19. Concurrent positive and negative controls should be used for each run (experiment) to ensure adequate performance of the experimental model. Skin discs from a single animal should be used in each run (experiment). The suggested positive and negative control test chemicals are 10 M hydrochloric acid and distilled water, respectively.
20. Liquid test chemicals (150 μl) are applied uniformly to the epidermal surface inside the tube. When testing solid materials, a sufficient amount of the solid is applied evenly to the disc to ensure that the whole surface of the epidermis is covered. Deionised water (150 μl) is added on top of the solid and the tube is gently agitated. In order to achieve maximum contact with the skin, solids may need to be warmed to 30° C to melt or soften the test chemical, or ground to produce a granular material or powder.

21. Three skin discs are used for each test and control chemical in each testing run (experiment). Test chemicals are applied for 24 hours at 20-23° C. The test chemical is removed by washing with a jet of tap water at up to room temperature until no further material can be removed.

TER measurements

22. The skin impedance is measured as TER by using a low-voltage, alternating current Wheatstone bridge (18). General specifications of the bridge are 1-3 Volt operating voltage, a sinus or rectangular shaped alternating current of 50 - 1 000 Hz, and a measuring range of at least 0,1-30 k. The databridge used in the validation study measured inductance, capacitance and resistance up to values of 2 000H, 2 000 F, and 2 M, respectively at frequencies of 100Hz or 1kHz, using series or parallel values. For the purposes of the TER corrosivity assay measurements are recorded in resistance, at a frequency of 100 Hz and using series values. Prior to measuring the electrical resistance, the surface tension of the skin is reduced by adding a sufficient volume of 70 % ethanol to cover the epidermis. After a few seconds, the ethanol is removed from the tube and the tissue is then hydrated by the addition of 3 ml MgSO₄ solution (154mM). The databridge electrodes are placed on either side of the skin disc to measure the resistance in kΩ/skin disc (Figure 1). Electrode dimensions and the length of the electrode exposed below the crocodile clips are shown in Figure 2. The clip attached to the inner electrode is rested on the top of the PTFE tube during resistance measurement to ensure that a consistent length of electrode is submerged in the MgSO₄ solution. The outer electrode is positioned inside the receptor chamber so that it rests on the bottom of the chamber. The distance between the spring clip and the bottom of the PTFE tube is maintained as a constant (Figure 2), because this distance affects the resistance value obtained. Consequently, the distance between the inner electrode and the skin disc should be constant and minimal (1-2 mm).
23. If the measured resistance value is greater than 20 k, this may be due to the remains of the test chemical coating the epidermal surface of the skin disc. Further removal of this coating can be attempted, for example, by sealing the PTFE tube with a gloved thumb and shaking it for approximately 10 seconds; the MgSO₄ solution is discarded and the resistance measurement is repeated with fresh MgSO₄.
24. The properties and dimensions of the test apparatus and the experimental procedure used may influence the TER values obtained. The 5 k corrosive threshold was developed from data obtained with the specific apparatus and procedure described in this test method. Different threshold and control values may apply if the test conditions are altered or a different apparatus is used. Therefore, it is necessary to calibrate the methodology and resistance threshold values by testing a series of Proficiency Substances chosen from the substances used in the validation study (8) (9), or from similar chemical classes to the substances being investigated. A set of suitable Proficiency Substances is identified in Table 1.

Dye Binding Methods

25. Exposure of certain non-corrosive materials can result in a reduction of resistance below the cut-off of 5 kΩ allowing the passage of ions through the *stratum corneum*, thereby reducing the electrical resistance (9). For example, neutral organics and substances that have surface-active properties (including detergents, emulsifiers and other surfactants) can remove skin lipids making the barrier more permeable to ions. Thus, if TER values produced by such chemicals are less than or around 5 kΩ in the absence of visually perceptible damage of the skin discs, an assessment of dye penetration should be carried out on the control and treated tissues to determine if the TER values obtained were the result of increased skin permeability, or skin corrosion (7) (9). In case of the latter where the *stratum corneum* is disrupted, the dye sulforhodamine B, when applied to the skin surface rapidly penetrates and stains the underlying tissue. This particular dye is stable to a wide range of substances and is not affected by the extraction procedure described below.

Sulforhodamine B dye application and removal

26. Following TER assessment, the magnesium sulphate is discarded from the tube and the skin is carefully examined for obvious damage. If there is no obvious major damage (e.g. perforation), 150 l of a 10 % (w/v) dilution in distilled water of the dye sulforhodamine B (Acid Red 52; C.I. 45100; CAS number 3520-42-1), is applied to the epidermal surface of each skin disc for 2 hours. These skin discs are then washed with tap water at up to room temperature for approximately 10 seconds to remove any excess/unbound dye. Each skin disc is carefully removed from the PTFE tube and placed in a vial (e.g. a 20-ml glass scintillation vial) containing deionised water (8 ml). The vials are agitated gently for 5 minutes to remove any additional unbound dye.

This rinsing procedure is then repeated, after which the skin discs are removed and placed into vials containing 5ml of 30 % (w/v) sodium dodecyl sulphate (SDS) in distilled water and are incubated overnight at 60° C.

27. After incubation, each skin disc is removed and discarded and the remaining solution is centrifuged for 8 minutes at 21° C (relative centrifugal force $\sim 175 \times g$). A 1ml sample of the supernatant is diluted 1 in 5 (v/v) [i.e. 1ml + 4ml] with 30 % (w/v) SDS in distilled water. The optical density (OD) of the solution is measured at 565 nm.

Calculation of dye content

28. The sulforhodamine B dye content per disc is calculated from the OD values (9) (sulforhodamine B dye molar extinction coefficient at 565nm = $8,7 \times 10^4$; molecular weight = 580). The dye content is determined for each skin disc by the use of an appropriate calibration curve and mean dye content is then calculated for the replicates.

Acceptability Criteria

29. The mean TER results are accepted if the concurrent positive and negative control values fall within the acceptable ranges for the method in the testing laboratory. The acceptable resistance ranges for the methodology and apparatus described above are given in the following table:

Control	Substance	Resistance range (k)
Positive	10 M Hydrochloric acid	0,5 – 1,0
Negative	Distilled water	10 – 25

30. The mean dye binding results are accepted on condition that concurrent control values fall within the acceptable ranges for the method. Suggested acceptable dye content ranges for the control substances for the methodology and apparatus described above are given in the following table:

Control	Substance	Dye content range (g/disc)
Positive	10 M Hydrochloric acid	40 – 100
Negative	Distilled water	15 – 35

Interpretation of results

31. The cut-off TER value distinguishing corrosive from non-corrosive test chemicals was established during test method optimisation, tested during a pre-validation phase, and confirmed in a formal validation study.
32. The prediction model for rat skin TER skin corrosion test method (9) (19), associated with the UN GHS/CLP classification system, is given below:

The test chemical is considered to be non-corrosive to skin:

- i) if the mean TER value obtained for the test chemical is greater than ($>$) 5 k Ω , or
- ii) the mean TER value obtained for the test chemical is less than or equal to (\leq) 5 k Ω , and
 - the skin discs show no obvious damage(e.g. perforation), and
 - the mean disc dye content is less than ($<$) the mean disc dye content of the 10 M HCl positive control obtained concurrently (see paragraph 30 for positive control values).

The test chemical is considered to be corrosive to skin:

- i) if the mean TER value obtained for the test chemical is less than or equal to (\leq) 5 k Ω and the skin discs are obviously damaged(e.g. perforated), or
 - ii) the mean TER value obtained for the test chemical is less than or equal to (\leq) 5 k Ω , and
 - the skin discs show no obvious damage(e.g. perforation), but
 - the mean disc dye content is greater than or equal to (\geq) the mean disc dye content of the 10 M HCl positive control obtained concurrently (see paragraph 30 for positive control values).
33. A testing run (experiment) composed of at least three replicate skin discs should be sufficient for a test chemical when the classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements and/or mean TER equal to 5 ± 0.5 k Ω , a second independent testing run (experiment) should be considered, as well as a third one in case of discordant results between the first two testing runs (experiments).

DATA AND REPORTING

Data

34. Resistance values (k Ω) and dye content values ($\mu\text{g}/\text{disc}$), where appropriate, for the test chemical, as well as for positive and negative controls should be reported in tabular form, including data for each individual replicate disc in each testing run (experiment) and mean values \pm SD. All repeat experiments should be reported. Observed damage in the skin discs should be reported for each test chemical.

Test report

35. The test report should include the following information:

Test Chemical and Control Substances:

- Mono-constituent substance: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc;

- Multi-constituent substance, UVCB and mixture: characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physico-chemical properties of the constituents;
- Physical appearance, water solubility, and additional relevant physico-chemical properties;
- Source, lot number if available;
- Treatment of the test chemical/control substance prior to testing, if applicable (e.g. warming, grinding);
- Stability of the test chemical, limit date for use, or date for re-analysis if known;
- Storage conditions.

Test Animals:

- Strain and sex used;
- Age of the animals when used as donor animals;
- Source, housing condition, diet, etc.;
- Details of the skin preparation.

Test Conditions:

- Calibration curves for test apparatus;
- Calibration curves for dye binding test performance, band pass used for measuring OD values, and OD linearity range of measuring device (e.g. spectrophotometer), if appropriate;
- Details of the test procedure used for TER measurements;
- Details of the test procedure used for the dye binding assessment, if appropriate;
- Test doses used, duration of exposure period(s) and temperature(s) of exposure;
- Details on washing procedure used after the exposure period;
- Number of replicate skin discs used per test chemical and controls (positive and negative control);
- Description of any modification of the test procedure;

- Reference to historical data of the model. This should include, but is not limited to:
 - i) Acceptability of the positive and negative control TER values (in k Ω) with reference to positive and negative control resistance ranges
 - ii) Acceptability of the positive and negative control dye content values (in $\mu\text{g}/\text{disc}$) with reference to positive and negative control dye content ranges
 - iii) Acceptability of the test results with reference to historical variability between skin disc replicates
- Description of decision criteria/prediction model applied.

Results:

- Tabulation of data from the TER and dye binding assays (if appropriate) for individual test chemicals and controls, for each testing run (experiment) and each skin disc replicate (individual animals and individual skin samples), means, SDs and CVs;
- Description of any effects observed;
- The derived classification with reference to the prediction model/decision criteria used.

Discussion of the results

Conclusions

LITERATURE

- (1) United Nations (UN) (2013). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Second Revised Edition, UN New York and Geneva, 2013. Available at: [http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html].
- (2) Chapter B.4 of this Annex, Acute Dermal Irritation, Corrosion.
- (3) Chapter B.40bis of this Annex, *In Vitro* Skin Model.
- (4) Chapter B.65 of this Annex, *In Vitro* Membrane Barrier Test Method.
- (5) Chapter B.46 of this Annex, *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method.
- (6) OECD (2014). Guidance document on Integrated Approaches to Testing and Assessment for Skin Irritation/Corrosion. Environment, Health and Safety Publications, Series on Testing and Assessment, (No 203), Organisation for Economic Cooperation and Development, Paris.

- (7) Botham P.A., Chamberlain M., Barratt M.D., Curren R.D., Esdaile D.J., Gardner J.R., Gordon V.C., Hildebrand B., Lewis R.W., Liebsch M., Logemann P., Osborne R., Ponc M., Regnier J.F., Steiling W., Walker A.P., and Balls M. (1995). A Prevalidation Study on *In Vitro* Skin Corrosivity Testing. The Report and Recommendations of ECVAM Workshop 6. *ATLA* 23, 219-255.
- (8) Barratt M.D., Brantom P.G., Fentem J.H., Gerner I., Walker A.P., and Worth A.P. (1998). The ECVAM International Validation Study on *In Vitro* Tests for Skin Corrosivity. 1. Selection and Distribution of the Test Chemicals. *Toxic.In Vitro* 12, 471-482.
- (9) Fentem J.H., Archer G.E.B., Balls M., Botham P.A., Curren R.D., Earl L.K., Esdaile D.J., Holzhütter H.-G., and Liebsch M. (1998). The ECVAM International Validation Study on *In Vitro* Tests For Skin Corrosivity. 2. Results and Evaluation by the Management Team. *Toxic.In Vitro* 12, 483- 524.
- (10) Balls M., Blaauboer B.J., Fentem J.H., Bruner L., Combes R.D., Ekwall B., Fielder R.J., Guillouzo A., Lewis R.W., Lovell D.P., Reinhardt C.A., Repetto G., Sladowski D., Spielmann H., and Zucco F. (1995). Practical Aspects of the Validation of Toxicity Test Procedures. The Report and Recommendations of ECVAM Workshops. *ATLA* 23, 129-147.
- (11) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (1997). Validation and Regulatory Acceptance of Toxicological Test Methods. NIH Publication No 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.
- (12) EC-ECVAM (1998). Statement on the Scientific Validity of the Rat Skin Transcutaneous Electrical Resistance (TER) Test (an *In Vitro* Test for Skin Corrosivity), Issued by the ECVAM Scientific Advisory Committee (ESAC10), 3 April 1998.
- (13) ECVAM (1998). ECVAM News & Views. *ATLA* 26, 275-280.
- (14) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2002). ICCVAM Evaluation of EpiDerm™ (EPI-200), EPISKIN™ (SM), and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: *In Vitro* Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No 02-4502. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.
- (15) OECD (2015). Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Transcutaneous Electrical Resistance (TER) Test Method for Skin Corrosion in Relation to TG 430. Environmental Health and Safety Publications, Series on Testing and Assessment No 218. Organisation for Economic Cooperation and Development, Paris.
- (16) Oliver G.J.A., Pemberton M.A., and Rhodes C. (1986). An *In Vitro* Skin Corrosivity Test -Modifications and Validation. *Fd. Chem. Toxicol.* 24, 507-512.
- (17) Botham P.A., Hall T.J., Dennett R., McCall J.C., Basketter D.A., Whittle E., Cheeseman M., Esdaile D.J., and Gardner J. (1992). The Skin Corrosivity Test *In Vitro*: Results of an Interlaboratory Trial. *Toxicol. In Vitro* 6, 191-194.
- (18) Eskes C., Detappe V., Koëter H., Kreysa J., Liebsch M., Zuang V., Amcoff P., Barroso J., Cotovio J., Guest R., Hermann M., Hoffmann S., Masson P., Alépée N., Arce L.A., Brüschweiler B., Catone T., Cihak R., Clouzeau J., D'Abrosca F., Delveaux C., Derouette J.P., Engelking O., Facchini D., Fröhlicher M., Hofmann M., Hopf N., Molinari J., Oberli A., Ott M., Peter R., Sá-Rocha V.M., Schenk D., Tomicic C., Vanparys P., Verdon B., Wallenhorst T., Winkler G.C. and Depallens O. (2012). Regulatory Assessment of *In Vitro* Skin Corrosion and Irritation Data Within the European Framework: Workshop Recommendations. *Regul.Toxicol.Pharmacol.* 62, 393-403.

- (19) TER SOP (December 2008). INVITTOX Protocol (No 115) Rat Skin Transcutaneous Electrical Resistance (TER) Test.
- (20) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environment, Health and Safety Publications, Series on Testing and Assessment (No 34), Organisation for Economic Cooperation and Development, Paris.

Figure 1

Apparatus for the Rat Skin Ter Assay

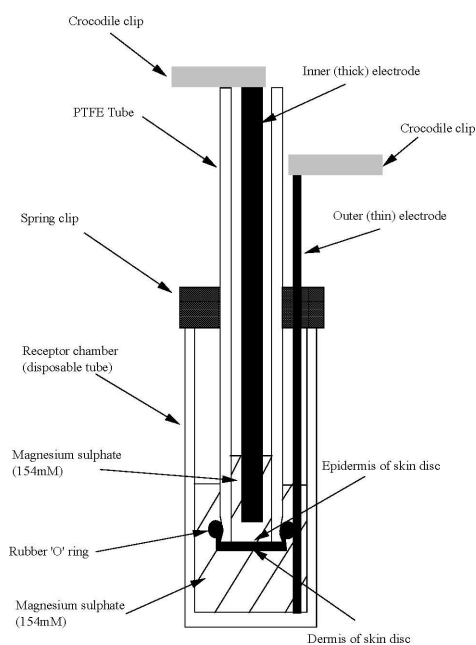
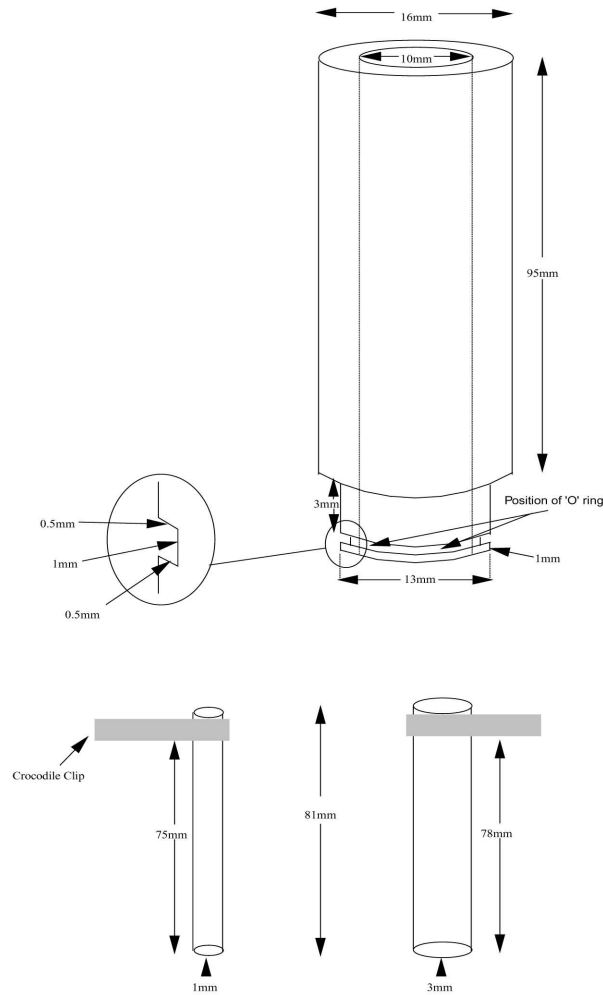


Figure 2

Dimensions Of The Polytetrafluoroethylene (PtfE) And Receptor Tubes And Electrodes Used



Critical factors of the apparatus shown above:

- The inner diameter of the PTFE tube,
- The length of the electrodes relative to the PTFE tube and receptor tube, such that the skin disc should not be touched by the electrodes and that a standard length of electrode is in contact with the MgSO_4 solution,
- The amount of MgSO_4 solution in the receptor tube should give a depth of liquid, relative to the level in the PTFE tube, as shown in Figure 1,
- The skin disc should be fixed well enough to the PTFE tube, such that the electrical resistance is a true measure of the skin properties.

Appendix

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with “concordance” to mean the proportion of correct outcomes of a test method (20).

C: Corrosive.

Chemical: A substance or a mixture.

Concordance: A measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (20).

GHS (Globally Harmonized System of Classification and Labelling of Chemicals (UN)): A system proposing the classification of chemicals (substances and mixtures) according to standardised types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

IATA: Integrated Approach on Testing and Assessment.

Mixture: A mixture or solution composed of two or more substances.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration ≥ 10 % (w/w) and < 80 % (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

NC: Non corrosive.

OD: Optical Density.

PC: Positive Control, a replicate containing all components of a test system and treated with a substance known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of reliability and accuracy, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals.

Relevance: Description of relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test method correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (20).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility (20).

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (20).

Skin corrosion in vivo: The production of irreversible damage of the skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (20).

Substance: A chemical element and its compounds in the natural state or obtained by any production process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

(Testing) run: A single test chemical concurrently tested in a minimum of three replicate skin discs.

Test chemical: Any substance or mixture tested using this test method.

Transcutaneous Electrical Resistance (TER): is a measure of the electrical impedance of the skin, as a resistance value in kilo Ohms. A simple and robust method of assessing barrier function by recording the passage of ions through the skin using a Wheatstone bridge apparatus.

UVCB: Substances of unknown or variable composition, complex reaction products or biological materials."

- (6) In Part B, Chapter B.40bis is replaced by the following:

"B.40bis **IN VITRO SKIN CORROSION: RECONSTRUCTED HUMAN EPIDERMIS (RhE) TEST METHOD**

INTRODUCTION

1. This test method (TM) is equivalent to OECD test guideline (TG) 431 (2016). Skin corrosion refers to the production of irreversible damage to the skin manifested as visible necrosis through the *epidermis* and into the *dermis*, following the application of a test chemical [as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (1) and the European Union (EU) Regulation 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) ⁽¹⁾]. This updated test method B.40bis provides an *in vitro* procedure allowing the identification of non-corrosive and corrosive substances and mixtures in accordance with UN GHS and CLP. It also allows a partial sub-categorisation of corrosives.
2. The assessment of skin corrosion potential of chemicals has typically involved the use of laboratory animals (TM B.4, equivalent to OECD TG 404; originally adopted in 1981 and revised in 1992, 2002 and 2015) (2). In addition to the present test method B.40bis, two other *in vitro* test methods for testing corrosion potential of chemicals have been validated and adopted as TM B.40 (equivalent to OECD TG 430) (3) and TM B.65 (equivalent to OECD TG 435) (4). Furthermore the *in vitro* TM B.46 (equivalent to OECD TG 439) (5) has been adopted for testing skin irritation potential. A OECD guidance document on Integrated Approaches to Testing and Assessment (IATA) for Skin Corrosion and Irritation describes several modules which group information sources and analysis tools, and provides guidance on (i) how to integrate and use existing testing and non-testing data for the assessment of skin irritation and skin corrosion potentials of chemicals and (ii) proposes an approach when further testing is needed (6).
3. This test method addresses the human health endpoint skin corrosion. It makes use of reconstructed human *epidermis* (RhE) (obtained from human derived non-transformed epidermal keratinocytes) which closely mimics the histological, morphological, biochemical and physiological properties of the upper parts of the human skin, *i.e.* the *epidermis*. The corresponding OECD test guideline was originally adopted in 2004 and updated in 2013 to include additional test methods using the RhE models and the possibility to use the methods to support the sub-categorisation of corrosive chemicals, and updated in 2015 to refer to the IATA guidance document and introduce the use of an alternative procedure to measure viability.
4. Four validated commercially available RhE models are included in this test method. Prevalidation studies (7), followed by a formal validation study for assessing skin corrosion (8)(9)(10) have been conducted (11) (12) for two of these commercially available test models, EpiSkin™ Standard Model (SM) and EpiDerm™ Skin Corrosivity Test (SCT) (EPI-200) (referred to in the following text as the Validated Reference Methods - VRMs). The outcome of these studies led to the recommendation that the two VRMs mentioned above could be used for regulatory purposes for distinguishing corrosive (C) from non-corrosive (NC) substances, and that the EpiSkin™ could moreover be used to support sub-categorisation of corrosive substances (13)(14)(15). Two other commercially available *in vitro* skin corrosion RhE test models have shown similar results to the EpiDerm™ VRM according to PS-based validation (16)(17)(18). These are the SkinEthic™ RHE ⁽²⁾ and epiCS® (previously named EST-1000) that can also be used for regulatory purposes for distinguishing corrosive from noncorrosive substances (19)(20). Post validation studies performed by the RhE model producers in the years 2012 to 2014 with a refined protocol correcting interferences of unspecific MTT reduction by the test chemicals improved the performance of both discrimination of C/NC as well as supporting subcategorisation of corrosives (21)(22). Further statistical analyses of the post-validation data generated with EpiDerm™ SCT, SkinEthic™ RHE and EpiCS® have been performed to identify alternative predictions models that improved the predictive capacity for sub-categorisation (23).

⁽¹⁾ Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006, OJ L 353/1, 31.12.2008

⁽²⁾ The abbreviation RhE (=Reconstructed human Epidermis) is used for all models based on RhE technology. The abbreviation RHE as used in conjunction with the SkinEthic™ model means the same, but, as part of the name of this specific test method as marketed, is spelled all in capitals.

5. Before a proposed similar or modified *in vitro* RhE test method for skin corrosion other than the VRMs can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined to ensure its similarity to the VRMs, in accordance with the requirements of the Performance Standards (PS) (24) set out in accordance with the principles of OECD guidance document No 34 (25). The Mutual Acceptance of Data will only be guaranteed after any proposed new or updated test method following the PS have been reviewed and included in the corresponding test guideline. The test models included in that test guideline can be used to address countries' requirements for test results on *in vitro* test method for skin corrosion, while benefiting from the Mutual Acceptance of Data.

DEFINITIONS

6. Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS

7. This test method allows the identification of non-corrosive and corrosive substances and mixtures in accordance with the UN GHS and CLP. This test method further supports the sub-categorisation of corrosive substances and mixtures into optional sub-category 1A, in accordance with the UN GHS (1), as well as a combination of sub-categories 1B and 1C (21)(22)(23). A limitation of this test method is that it does not allow discriminating between skin corrosive sub-category 1B and sub-category 1C in accordance with the UN GHS and CLP due to the limited set of well-known *in vivo* corrosive sub-category 1C chemicals. EpiSkin™, EpiDerm™ SCT, SkinEthic™ RHE and epiCS® test models are able to sub-categorise (i.e. 1A versus 1B-and-1C versus NC).
8. A wide range of chemicals representing mainly individual substances has been tested in the validation supporting the test models included in this test method when they are used for identification of non-corrosives and corrosives; the empirical database of the validation study amounted to 60 chemicals covering a wide range of chemical classes (8)(9)(10). Testing to demonstrate sensitivity, specificity, accuracy and within-laboratory-reproducibility of the assay for sub-categorisation was performed by the test method developers and results were reviewed by the OECD (21) (22) (23). On the basis of the overall data available, the test method is applicable to a wide range of chemical classes and physical states including liquids, semi-solids, solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other prior treatment of the sample is required. In cases where evidence can be demonstrated on the non-applicability of test models included in this test method to a specific category of test chemicals, they should not be used for that specific category of test chemicals. In addition, this test method is assumed to be applicable to mixtures as an extension of its applicability to substances. However, due to the fact that mixtures cover a wide spectrum of categories and composition, and that only limited information is currently available on the testing of mixtures, in cases where evidence can be demonstrated on the non-applicability of the test method to a specific category of mixtures (e.g. following a strategy as proposed in (26)), the test method should not be used for that specific category of mixtures. Before use of the test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture. Gases and aerosols have not been assessed yet in validation studies (8)(9)(10). While it is conceivable that these can be tested using RhE technology, the current test method does not allow testing of gases and aerosols.
9. Test chemicals absorbing light in the same range as MTT formazan and test chemicals able to directly reduce the vital dye MTT (to MTT formazan) may interfere with the tissue viability measurements and need the use of adapted controls for corrections. The type of adapted controls that may be required will vary depending on the type of interference produced by the test chemical and the procedure used to measure MTT formazan (see paragraphs 25-31).

10. While this test method does not provide adequate information on skin irritation, it should be noted that TM B.46 specifically addresses the health effect skin irritation *in vitro* and is based on the same RhE test system, though using another protocol (5). For a full evaluation of local skin effects after a single dermal exposure, the OECD Guidance Document on Integrated Approaches for Testing and Assessment should be consulted (6). This IATA approach includes the conduct of *in vitro* tests for skin corrosion (such as described in this test method) and skin irritation before considering testing in living animals. It is recognised that the use of human skin is subject to national and international ethical considerations and conditions.

PRINCIPLE OF THE TEST

11. The test chemical is applied topically to a three-dimensional RhE model, comprised of non-transformed, human-derived epidermal keratinocytes, which have been cultured to form a multi-layered, highly differentiated model of the human *epidermis*. It consists of organised basal, spinous and granular layers, and a multi-layered *stratum corneum* containing intercellular lamellar lipid layers representing main lipid classes analogous to those found *in vivo*.
12. The RhE test method is based on the premise that corrosive chemicals are able to penetrate the *stratum corneum* by diffusion or erosion, and are cytotoxic to the cells in the underlying layers. Cell viability is measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1], into a blue formazan salt that is quantitatively measured after extraction from tissues (27). Corrosive chemicals are identified by their ability to decrease cell viability below defined threshold levels (see paragraphs 35 and 36). The RhE-based skin corrosion test method has shown to be predictive of *in vivo* skin corrosion effects assessed in rabbits according to the TM B.4 (2).

DEMONSTRATION OF PROFICIENCY

13. Prior to routine use of any of the four validated RhE test models that adhere to this test method, laboratories should demonstrate technical proficiency by correctly classifying the twelve Proficiency Substances listed in Table 1. In case of the use of a method for sub-classification, also the correct sub-categorisation should be demonstrated. In situations where a listed substance is unavailable or where justifiable, another substance for which adequate *in vivo* and *in vitro* reference data are available may be used (e.g. from the list of reference chemicals (24)) provided that the same selection criteria as described in Table 1 is applied.

Table 1

List of Proficiency Substances ⁽¹⁾

Substance	CASRN	Chemical Class ⁽²⁾	UN GHS/CLP Cat. Based on <i>In Vivo</i> results ⁽³⁾	VRM Cat. Based on <i>In Vitro</i> results ⁽⁴⁾	MTT Reducer ⁽⁵⁾	Physical State
Sub-category 1A <i>In Vivo</i> Corrosivesg						
Bromoacetic acid	79-08-3	Organic acid	1A	(3) 1A	—	S
Boron trifluoride dihydrate	13319-75-0	Inorganic acid	1A	(3) 1A	—	L
Phenol	108-95-2	Phenol	1A	(3) 1A	—	S
Dichloroacetylchloride	79-36-7	Electrophile	1A	(3) 1A	—	L
Combination of sub-categories 1B-and-1C <i>In Vivo</i> Corrosives						
Glyoxylic acid monohydrate	563-96-2	Organic acid	1B-and-1C	(3) 1B-and-1C	—	S

Substance	CASRN	Chemical Class ⁽²⁾	UN GHS/CLP Cat. Based on <i>In Vivo</i> results ⁽³⁾	VRM Cat. Based on <i>In Vitro</i> results ⁽⁴⁾	MTT Reducer ⁽⁵⁾	Physical State
Sub-category 1A <i>In Vivo</i> Corrosivesg						
Lactic acid	598-82-3	Organic acid	1B-and-1C	(3) 1B-and-1C	—	L
Ethanolamine	141-43-5	Organic base	1B	(3) 1B-and-1C	Y	Viscous
Hydrochloric acid (14,4 %)	7647-01-0	Inorganic acid	1B-and-1C	(3) 1B-and-1C	—	L
<i>In Vivo</i> Non Corrosives						
Phenethyl bromide	103-63-9	Electrophile	NC	(3) NC	Y	L
4-Amino-1,2,4-triazole	584-13-4	Organic base	NC	(3) NC	—	S
4-(methylthio)-benzaldehyde	3446-89-7	Electrophile	NC	(3) NC	Y	L
Lauric acid	143-07-7	Organic acid	NC	(3) NC	—	S

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; VRM = Validated Reference Method; NC = Not Corrosive; Y = yes; S = solid; L = liquid

(1) The proficiency substances, sorted first by corrosives versus non-corrosives, then by corrosive sub-category and then by chemical class, were selected from the substances used in the ECVAM validation studies of EpiSkin™ and EpiDerm™ (8) (9) (10) and from post-validation studies based on data provided by EpiSkin™ (22), EpiDerm™, SkinEthic™ and epiCS® developers (23). Unless otherwise indicated, the substances were tested at the purity level obtained when purchased from a commercial source (8) (10). The selection includes, to the extent possible, substances that: (i) are representative of the range of corrosivity responses (e.g. noncorrosives; weak to strong corrosives) that the VRMs are capable of measuring or predicting; (ii) are representative of the chemical classes used in the validation studies; (iii) have chemical structures that are well-defined; (iv) induce reproducible results in the VRM; (v) induce definitive results in the *in vivo* reference test method; (vi) are commercially available; and (vii) are not associated with prohibitive disposal costs.

(2) Chemical class assigned by Barratt *et al.* (8).

(3) The corresponding UN Packing groups are I, II and III, respectively, for the UN GHS/CLP 1A, 1B and 1C.

(4) The VRM *in vitro* predictions reported in this table were obtained with the EpiSkin™ and the EpiDerm™ test models (VRMs) during post-validation testing performed by the test method developers.

(5) The viability values obtained in the ECVAM Skin Corrosion Validation Studies were not corrected for direct MTT reduction (killed controls were not performed in the validation studies). However, the post-validation data generated by the test method developers that are presented in this table were acquired with adapted controls (23).

14. As part of the proficiency exercise, it is recommended that the user verifies the barrier properties of the tissues after receipt as specified by the RhE model manufacturer. This is particularly important if tissues are shipped over long distance/time periods. Once a test method has been successfully established and proficiency in its use has been demonstrated, such verification will not be necessary on a routine basis. However, when using a test method routinely, it is recommended to continue to assess the barrier properties in regular intervals.

PROCEDURE

15. The following is a generic description of the components and procedures of the RhE test models for skin corrosion assessment covered by this test method. The RhE models endorsed as scientifically valid for use within this test method, *i.e.* the EpiSkin™ (SM), EpiDerm™ (EPI-200), SkinEthic™ RHE and epiCS® models (16)(17)(19)(28)(29)(30)(31)(32)(33), can be obtained from commercial sources. Standard Operating Procedures (SOPs) for these four RhE models are available (34)(35)(36)(37), and their main test method components are summarised in Appendix 2. It is recommended that the relevant SOP be consulted when implementing and using one of these models in the laboratory. Testing with the four RhE test models covered by this test method should comply with the following:

RHE TEST METHOD COMPONENTS

General Conditions

16. Non-transformed human keratinocytes should be used to reconstruct the epithelium. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum*, *stratum granulosum*) should be present under a functional *stratum corneum*. The *stratum corneum* should be multi-layered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic benchmark chemicals, e.g. sodium dodecyl sulphate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a benchmark chemical reduces the viability of the tissues by 50 % (IC₅₀) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50 % (ET₅₀) upon application of the benchmark chemical at a specified, fixed concentration (see paragraph 18). The containment properties of the RhE model should prevent the passage of material around the *stratum corneum* to the viable tissue, which would lead to poor modelling of skin exposure. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, or fungi.

Functional Conditions*Viability*

17. The assay used for quantifying tissue viability is the MTT-assay (27). The viable cells of the RhE tissue construct reduce the vital dye MTT into a blue MTT formazan precipitate, which is then extracted from the tissue using isopropanol (or a similar solvent). The OD of the extraction solvent alone should be sufficiently small, i.e., OD < 0,1. The extracted MTT formazan may be quantified using either a standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure (38). The RhE model users should ensure that each batch of the RhE model used meets defined criteria for the negative control. An acceptability range (upper and lower limit) for the negative control OD values should be established by the RhE model developer/supplier. Acceptability ranges for the negative control OD values for the four validated RhE test models included in this test method are given in Table 2. An HPLC/UPLC- Spectrophotometry user should use the negative control OD ranges provided in Table 2 as the acceptance criterion for the negative control. It should be documented that the tissues treated with negative control are stable in culture (provide similar OD measurements) for the duration of the exposure period.

Table 2

Acceptability ranges for negative control OD values to control batch quality

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM)	> 0,6	< 1,5
EpiDerm™ SCT (EPI-200)	> 0,8	< 2,8
SkinEthic™ RHE	> 0,8	< 3,0
epiCS®	> 0,8	< 2,8

Barrier function

18. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of certain cytotoxic benchmark chemicals (e.g. SDS or Triton X-100), as estimated by IC₅₀ or ET₅₀ (Table 3). The barrier function of each batch of the RhE model used should be demonstrated by the RhE model developer/vendor upon supply of the tissues to the end user (see paragraph 21).

Morphology

19. Histological examination of the RhE model should be performed demonstrating multi-layered human *epidermis*-like structure containing *stratum basale*, *stratum spinosum*, *stratum granulosum* and *stratum corneum* and exhibits lipid profile similar to lipid profile of human epidermis. Histological examination of each batch of the RhE model used demonstrating appropriate morphology of the tissues should be provided by the RhE model developer/vendor upon supply of the tissues to the end user (see paragraph 21).

Reproducibility

20. Test method users should demonstrate reproducibility of the test methods over time with the positive and negative controls. Furthermore, the test method should only be used if the RhE model developer/supplier provides data demonstrating reproducibility over time with corrosive and non-corrosive chemicals from e.g. the list of Proficiency Substances (Table 1). In case of the use of a test method for subcategorisation, the reproducibility with respect to sub-categorisation should also be demonstrated.

Quality control (QC)

21. The RhE model should only be used if the developer/supplier demonstrates that each batch of the RhE model used meets defined production release criteria, among which those for *viability* (paragraph 17), *barrier function* (paragraph 18) and *morphology* (paragraph 19) are the most relevant. These data are provided to the test method users, so that they are able to include this information in the test report. Only results produced with QC accepted tissue batches can be accepted for reliable prediction of corrosive classification. An acceptability range (upper and lower limit) for the IC₅₀ or the ET₅₀ is established by the RhE model developer/supplier. The acceptability ranges for the four validated test models are given in Table 3.

Table 3

QC batch release criteria

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM) (18 hours treatment with SDS) (33)	IC ₅₀ = 1,0 mg/ml	IC ₅₀ = 3,0 mg/ml
EpiDerm™ SCT (EPI-200) (1 % Triton X-100) (34)	ET ₅₀ = 4,0 hours	ET ₅₀ = 8,7 hours
SkinEthic™ RHE (1 % Triton X-100) (35)	ET ₅₀ = 4,0 hours	ET ₅₀ = 10,0 hours
epiCS® (1 % Triton X-100) (36)	ET ₅₀ = 2,0 hours	ET ₅₀ = 7,0 hours

Application of the Test Chemical and Control Chemicals

22. At least two tissue replicates should be used for each test chemical and controls for each exposure time. For liquid as well as solid chemicals, sufficient amount of test chemical should be applied to uniformly cover the epidermis surface while avoiding an infinite dose, i.e. a minimum of 70 µl/cm² or 30 mg/cm² should be used. Depending on the models, the epidermis surface should be moistened with deionised or distilled water before

application of solid chemicals, to improve contact between the test chemical and the epidermis surface (34)(35)(36)(37). Whenever possible, solids should be tested as a fine powder. The application method should be appropriate for the test chemical (see e.g. references (34-37)). At the end of the exposure period, the test chemical should be carefully washed from the *epidermis* with an aqueous buffer, or 0,9 % NaCl. Depending on which of the four validated RhE test model is used, two or three exposure periods are used per test chemical (for all four valid RhE models: 3 min and 1 hour; for EpiSkin™ an additional exposure time of 4 hours). Depending on the RhE test model used and the exposure period assessed, the incubation temperature during exposure may vary between room temperature and 37°C.

23. Concurrent negative and positive controls (PC) should be used in each run to demonstrate that viability (with negative controls), barrier function and resulting tissue sensitivity (with the PC) of the tissues are within a defined historical acceptance range. The suggested PC chemicals are glacial acetic acid or 8N KOH depending upon the RhE model used. It should be noted that 8N KOH is a direct MTT reducer that might require adapted controls as described in paragraphs 25 and 26. The suggested negative controls are 0,9 % (w/v) NaCl or water.

Cell Viability Measurements

24. The MTT assay, which is a quantitative assay, should be used to measure cell viability under this test method (27). The tissue sample is placed in MTT solution of appropriate concentration (0.3 or 1 mg/ml) for 3 hours. The precipitated blue formazan product is then extracted from the tissue using a solvent (e.g. isopropanol, acidic isopropanol), and the concentration of formazan is measured by determining the OD at 570 nm using a filter band pass of maximum ± 30 nm, or by an HPLC/UPLC- spectrophotometry procedure (see paragraphs 30 and 31)(38).
25. Test chemicals may interfere with the MTT assay, either by direct reduction of the MTT into blue formazan, and/or by colour interference if the test chemical absorbs, naturally or due to treatment procedures, in the same OD range of formazan (570 ± 30 nm, mainly blue and purple chemicals). Additional controls should be used to detect and correct for a potential interference from these test chemicals such as the non-specific MTT reduction (NSMTT) control and the non-specific colour (NSC) control (see paragraphs 26 to 30). This is especially important when a specific test chemical is not completely removed from the tissue by rinsing or when it penetrates the *epidermis*, and is therefore present in the tissues when the MTT viability test is performed. Detailed description of how to correct direct MTT reduction and interferences by colouring agents is available in the SOPs for the test models (34)(35)(36)(37).
26. To identify direct MTT reducers, each test chemical should be added to freshly prepared MTT medium (34) (35) (36) (37). If the MTT mixture containing the test chemical turns blue/purple, the test chemical is presumed to directly reduce the MTT, and further functional check on non-viable epidermis should be performed, independently of using the standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure. This additional functional check employs killed tissues that possess only residual metabolic activity but absorb the test chemical in similar amount as viable tissues. Each MTT reducing chemical is applied on at least two killed tissue replicates per exposure time, which undergo the whole skin corrosion test. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the MTT reducer minus the percent non-specific MTT reduction obtained with the killed tissues exposed to the same MTT reducer, calculated relative to the negative control run concurrently to the test being corrected (%NSMTT).
27. To identify potential interference by coloured test chemicals or test chemicals that become coloured when in contact with water or isopropanol and decide on the need for additional controls, spectral analysis of the test chemical in water (environment during exposure) and/or isopropanol (extracting solution) should be performed. If the test chemical in water and/or isopropanol absorbs light in the range of 570 ± 30 nm, furthercolorant

controls should be performed or, alternatively, an HPLC/UPLC- spectrophotometry procedure should be used in which case these controls are not required (see paragraphs 30 and 31). When performing the standard absorbance (OD) measurement, each interfering coloured test chemical is applied on at least two viable tissue replicates per exposure time, which undergo the entire skin corrosion test but are incubated with medium instead of MTT solution during the MTT incubation step to generate a non-specific colour (NSC_{living}) control. The NSC_{living} control needs to be performed concurrently per exposure time per coloured test chemical (in each run) due to the inherent biological variability of living tissues. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the interfering test chemical and incubated with MTT solution minus the percent nonspecific colour obtained with living tissues exposed to the interfering test chemical and incubated with medium without MTT, run concurrently to the test being corrected (%NSC_{living}).

28. Test chemicals that are identified as producing both direct MTT reduction (see paragraph 26) and colour interference (see paragraph 27) will also require a third set of controls, apart from the NSMTT and NSC_{living} controls described in the previous paragraphs, when performing the standard absorbance (OD) measurement. This is usually the case with darkly coloured test chemicals interfering with the MTT assay (e.g., blue, purple, black) because their intrinsic colour impedes the assessment of their capacity to directly reduce MTT as described in paragraph 26. These test chemicals may bind to both living and killed tissues and therefore the NSMTT control may not only correct for potential direct MTT reduction by the test chemical, but also for colour interference arising from the binding of the test chemical to killed tissues. This could lead to a double correction for colour interference since the NSC_{living} control already corrects for colour interference arising from the binding of the test chemical to living tissues. To avoid a possible double correction for colour interference, a third control for non-specific colour in killed tissues (NSC_{killed}) needs to be performed. In this additional control, the test chemical is applied on at least two killed tissue replicates per exposure time, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step. A single NSC_{killed} control is sufficient per test chemical regardless of the number of independent tests/runs performed, but should be performed concurrently to the NSMTT control and, where possible, with the same tissue batch. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT minus %NSC_{living} plus the percent non-specific colour obtained with killed tissues exposed to the interfering test chemical and incubated with medium without MTT, calculated relative to the negative control run concurrently to the test being corrected (%NSC_{killed}).
29. It is important to note that non-specific MTT reduction and non-specific colour interferences may increase the readouts of the tissue extract above the linearity range of the spectrophotometer. On this basis, each laboratory should determine the linearity range of their spectrophotometer with MTT formazan (CAS # 57360-69-7) from a commercial source before initiating the testing of test chemicals for regulatory purposes. In particular, the standard absorbance (OD) measurement using a spectrophotometer is appropriate to assess direct MTT-reducers and colour interfering test chemicals when the ODs of the tissue extracts obtained with the test chemical without any correction for direct MTT reduction and/or colour interference are within the linear range of the spectrophotometer or when the uncorrected percent viability obtained with the test chemical already defined it as a corrosive (see paragraphs 35 and 36). Nevertheless, results for test chemicals producing %NSMTT and/or %NSC_{living} > 50 % of the negative control should be taken with caution.
30. For coloured test chemicals which are not compatible with the standard absorbance (OD) measurement due to too strong interference with the MTT assay, the alternative HPLC/UPLC- spectrophotometry procedure to measure MTT formazan may be employed (see paragraph 31) (37). The HPLC/UPLC-spectrophotometry system allows for the separation of the MTT formazan from the test chemical before its quantification (38). For this reason, NSC_{living} or NSC_{killed} controls are never required when using HPLC/UPLC-spectrophotometry, independently of the chemical being tested. NSMTT controls should nevertheless be used if the test chemical is suspected to directly reduce MTT or has a colour that impedes the assessment of the capacity to directly reduce MTT (as described in paragraph 26). When using HPLC/UPLC-spectrophotometry to measure MTT formazan, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. For test chemicals able to directly reduce MTT, true tissue viability is calculated as the percent tissue viability obtained

with living tissues exposed to the test chemical minus %NSMTT. Finally, it should be noted that direct MTT-reducers that may also be colour interfering, which are retained in the tissues after treatment and reduce MTT so strongly that they lead to ODs (using standard OD measurement) or peak areas (using UPLC/HPLC-spectrophotometry) of the tested tissue extracts that fall outside of the linearity range of the spectrophotometer cannot be assessed, although these are expected to occur in only very rare situations.

31. HPLC/UPLC-spectrophotometry may be used also with all types of test chemicals (coloured, non-coloured, MTT-reducers and non-MTT reducers) for measurement of MTT formazan (38). Due to the diversity of HPLC/UPLC-spectrophotometry systems, qualification of the HPLC/UPLC-spectrophotometry system should be demonstrated before its use to quantify MTT formazan from tissue extracts by meeting the acceptance criteria for a set of standard qualification parameters based on those described in the U.S. Food and Drug Administration guidance for industry on bio-analytical method validation (38)(39). These key parameters and their acceptance criteria are shown in Appendix 4. Once the acceptance criteria defined in Appendix 4 have been met, the HPLC/UPLC-spectrophotometry system is considered qualified and ready to measure MTT formazan under the experimental conditions described in this test method.

Acceptability Criteria

32. For each test method using valid RhE models, tissues treated with the negative control should exhibit OD reflecting the quality of the tissues as described in table 2 and should not be below historically established boundaries. Tissues treated with the PC, *i.e.* glacial acetic acid or 8N KOH, should reflect the ability of the tissues to respond to a corrosive chemical under the conditions of the test model (see Appendix 2). The variability between tissue replicates of test chemical and/or control chemicals should fall within the accepted limits for each valid RhE model requirements (see Appendix 2) (*e.g.* the difference of viability between the two tissue replicates should not exceed 30 %). If either the negative control or PC included in a run fall out of the accepted ranges, the run is considered as not qualified and should be repeated. If the variability of test chemicals falls outside of the defined range, its testing should be repeated.

Interpretation of Results and Prediction Model

33. The OD values obtained for each test chemical should be used to calculate percentage of viability relative to the negative control, which is set at 100 %. In case HPLC/UPLC-spectrophotometry is used, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. The cut-off percentage cell viability values distinguishing corrosive from non-corrosive test chemical (or discriminating between different corrosive sub-categories) are defined below in paragraphs 35 and 36 for each of the test models covered by this test method and should be used for interpreting the results.
34. A single testing run composed of at least two tissue replicates should be sufficient for a test chemical when the resulting classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements, a second run may be considered, as well as a third one in case of discordant results between the first two runs.

35. The prediction model for the EpiSkin™ skin corrosion test model (9)(34)(22), associated with the UN GHS/CLP classification system, is shown in Table 4:

Table 4

EpiSkin™ prediction model

Viability measured after exposure time points (t=3, 60 and 240 minutes)	Prediction to be considered
< 35 % after 3 min exposure	Corrosive: • Optional sub-category 1A (*)
≥ 35 % after 3 min exposure AND < 35 % after 60 min exposure OR ≥ 35 % after 60 min exposure AND < 35 % after 240 min exposure	Corrosive: • A combination of optional sub-categories 1B-and-1C
≥ 35 % after 240 min exposure	Non-corrosive

(*) According to the data generated in view of assessing the usefulness of the RhE test models for supporting subcategorisation, it was shown that around 22 % of the sub-category 1A results of the EpiSkin™ test model may actually constitute sub-category 1B or sub-category 1C substances/mixtures (i.e. overclassifications) (see Appendix 3).

36. The prediction models for the EpiDerm™ SCT (10)(23)(35), the SkinEthic™ RHE (17)(18) (23) (36), and the epiCS® (16)(23)(37) skin corrosion test models, associated with the UN GHS/CLP classification system, are shown in Table 5:

Table 5

EpiDerm™ SCT, SkinEthic™ RHE and epiCS®

Viability measured after exposure time points (t=3 and 60 minutes)	Prediction to be considered
STEP 1 for EpiDerm™ SCT, for SkinEthic™ RHE and epiCS®	
< 50 % after 3 min exposure	Corrosive
≥ 50 % after 3 min exposure AND < 15 % after 60 min exposure	Corrosive
≥ 50 % after 3 min exposure AND ≥ 15 % after 60 min exposure	Non-corrosive
STEP 2 for EpiDerm™ SCT - for substances/mixtures identified as Corrosive in step 1	
< 25 % after 3 min exposure	Optional sub-category 1A *

Viability measured after exposure time points (t=3 and 60 minutes)	Prediction to be considered
STEP 1 for EpiDerm™ SCT, for SkinEthic™ RHE and epiCS®	
≥ 25 % after 3 min exposure	A combination of optional sub-categories 1B and 1C
STEP 2 for SkinEthic™ RHE - for substances/mixtures identified as Corrosive in step 1	
< 18 % after 3 min exposure	Optional sub-category 1A *
≥ 18 % after 3 min exposure	A combination of optional sub-categories 1B and 1C
STEP 2 for epiCS® - for substances/mixtures identified as Corrosive in step 1	
< 15 % after 3 min exposure	Optional sub-category 1A *
≥ 15 % after 3 min exposure	A combination of optional sub-categories 1B and 1C

DATA AND REPORTING

Data

37. For each test, data from individual tissue replicates (e.g. OD values and calculated percentage cell viability for each test chemical, including classification) should be reported in tabular form, including data from repeat experiments as appropriate. In addition, means and ranges of viability and CVs between tissue replicates for each test should be reported. Observed interactions with MTT reagent by direct MTT reducers or coloured test chemicals should be reported for each tested chemical.

Test report

38. The test report should include the following information:

Test Chemical and Control Chemicals:

- Mono-constituent substance: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc;
- Multi-constituent substance, UVCB and mixture: characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents;
- Physical appearance, water solubility, and any additional relevant physicochemical properties;
- Source, lot number if available;
- Treatment of the test chemical/control substance prior to testing, if applicable (e.g. warming, grinding);
- Stability of the test chemical, limit date for use, or date for re-analysis if known;

— Storage conditions.

RhE model and protocol used and rationale for it (if applicable)

Test Conditions:

— RhE model used (including batch number);

— Calibration information for measuring device (e.g. spectrophotometer), wavelength and band pass (if applicable) used for quantifying MTT formazan, and linearity range of measuring device;

— Description of the method used to quantify MTT formazan;

— Description of the qualification of the HPLC/UPLC-spectrophotometry system, if applicable;

— Complete supporting information for the specific RhE model used including its performance. This should include, but is not limited to:

i) Viability;

ii) Barrier function;

iii) Morphology;

iv) Reproducibility and predictive capacity;

v) Quality controls (QC) of the model;

— Reference to historical data of the model. This should include, but is not limited to acceptability of the QC data with reference to historical batch data;

— Demonstration of proficiency in performing the test method before routine use by testing of the proficiency substances.

Test Procedure:

— Details of the test procedure used (including washing procedures used after exposure period);

— Doses of test chemical and control chemicals used;

— Duration of exposure period(s) and temperature(s) of exposure;

— Indication of controls used for direct MTT-reducers and/or colouring test chemicals, if applicable;

- Number of tissue replicates used per test chemical and controls (PC, negative control, and NSMTT, NSCliving and NSCKilled, if applicable), per exposure time;
- Description of decision criteria/prediction model applied based on the RhE model used;
- Description of any modifications of the test procedure (including washing procedures).

Run and Test Acceptance Criteria:

- Positive and negative control mean values and acceptance ranges based on historical data;
- Acceptable variability between tissue replicates for positive and negative controls;
- Acceptable variability between tissue replicates for test chemical.

Results:

- Tabulation of data for individual test chemicals and controls, for each exposure period, each run and each replicate measurement including OD or MTT formazan peak area, percent tissue viability, mean percent tissue viability, differences between replicates, SDs and/or CVs if applicable;
- If applicable, results of controls used for direct MTT-reducers and/or colouring test chemicals including OD or MTT formazan peak area, %NSMTT, %NSCliving, %NSCKilled, differences between tissue replicates, SDs and/or CVs (if applicable), and final correct percent tissue viability;
- Results obtained with the test chemical(s) and control chemicals in relation to the defined run and test acceptance criteria;
- Description of other effects observed;
- The derived classification with reference to the prediction model/decision criteria used.

Discussion of the results

Conclusions

LITERATURE

- (1) UN (2013). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Fifth Revised Edition, UN New York and Geneva. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html
- (2) Chapter B.4 of this Annex, Acute Dermal Irritation, Corrosion.
- (3) Chapter B.40 of this Annex, *In Vitro* Skin Corrosion.

- (4) Chapter B.65 of this Annex, *In Vitro* Membrane Barrier Test Method.
- (5) Chapter B.46 of this Annex, *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method.
- (6) OECD (2014). Guidance Document on Integrated Approaches to Testing and Assessment of Skin Irritation/Corrosion. Environment, Health and Safety Publications, Series on Testing and Assessment, (No 203) Organisation for Economic Cooperation and Development, Paris.
- (7) Botham P.A., Chamberlain M., Barratt M.D., Curren R.D., Esdaile D.J., Gardner J.R., Gordon V.C., Hildebrand B., Lewis R.W., Liebsch M., Logemann P., Osborne R., Ponc M., Regnier J.F., Steiling W., Walker A.P., and Balls M. (1995). A Prevalidation Study on *In Vitro* Skin Corrosivity Testing. The report and Recommendations of ECVAM Workshop 6. *ATLA* 23:219-255.
- (8) Barratt M.D., Brantom P.G., Fentem J.H., Gerner I., Walker A.P., and Worth A.P. (1998). The ECVAM International Validation Study on *In Vitro* Tests for Skin Corrosivity. 1. Selection and distribution of the Test Chemicals. *Toxicol.In Vitro* 12:471-482.
- (9) Fentem J.H., Archer G.E.B., Balls M., Botham P.A., Curren R.D., Earl L.K., Esdaile D.J., Holzthutter H.-G., and Liebsch M. (1998). The ECVAM International Validation Study on *In Vitro* Tests for Skin Corrosivity. 2. Results and Evaluation by the Management Team. *Toxicol.in Vitro* 12:483-524.
- (10) Liebsch M., Traue D., Barrabas C., Spielmann H., Uphill, P., Wilkins S., Wiemann C., Kaufmann T., Remmele M. and Holzthütter H. G. (2000). The ECVAM Prevalidation Study on the Use of EpiDerm for Skin Corrosivity Testing, *ATLA* 28: 371-401.
- (11) Balls M., Blaauboer B.J., Fentem J.H., Bruner L., Combes R.D., Ekwall B., Fielder R.J., Guillouzo A., Lewis R.W., Lovell D.P., Reinhardt C.A., Repetto G., Sladowski D., Spielmann H. et Zucco F. (1995). Practical Aspects of the Validation of Toxicity Test Procedures. The Report and Recommendations of ECVAM Workshops, *ATLA* 23:129-147.
- (12) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (1997). Validation and Regulatory Acceptance of Toxicological Test Methods. NIH Publication No 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.
- (13) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2002). ICCVAM evaluation of EpiDerm™ (EPI-200), EPISKIN™ (SM), and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: *In Vitro* Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No 02-4502. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.
- (14) EC-ECVAM (1998). Statement on the Scientific Validity of the EpiSkin™ Test (an *In Vitro* Test for Skin Corrosivity), Issued by the ECVAM Scientific Advisory Committee (ESAC10), 3 April 1998.
- (15) EC-ECVAM (2000). Statement on the Application of the EpiDerm™ Human Skin Model for Skin Corrosivity Testing, Issued by the ECVAM Scientific Advisory Committee (ESAC14), 21 March 2000.
- (16) Hoffmann J., Heisler E., Karpinski S., Losse J., Thomas D., Siefken W., Ahr H.J., Vohr H.W. and Fuchs H.W. (2005). Epidermal-Skin-Test 1000 (EST-1000)-A New Reconstructed Epidermis for *In Vitro* Skin Corrosivity Testing. *Toxicol.In Vitro* 19: 925-929.

- (17) Kandárová H., Liebsch M., Spielmann, H., Genschow E., Schmidt E., Traue D., Guest R., Whittingham A., Warren N, Gamer A.O., Remmele M., Kaufmann T., Wittmer E., De Wever B., and Rosdy M. (2006). Assessment of the Human Epidermis Model SkinEthic RHE for *In Vitro* Skin Corrosion Testing of Chemicals According to New OECD TG 431. *Toxicol. In Vitro* 20: 547-559.
- (18) Tornier C., Roquet M. and Fraissinette A.B. (2010). Adaptation of the Validated SkinEthic™ Reconstructed Human Epidermis (RHE) Skin Corrosion Test Method to 0,5 cm² Tissue Sample. *Toxicol. In Vitro* 24: 1379-1385.
- (19) EC-ECVAM (2006). Statement on the Application of the SkinEthic™ Human Skin Model for Skin Corrosivity Testing, Issued by the ECVAM Scientific Advisory Committee (ESAC25), 17 November 2006.
- (20) EC-ECVAM (2009). ESAC Statement on the Scientific Validity of an *In-Vitro* Test Method for Skin Corrosivity Testing: the EST-1000, Issued by the ECVAM Scientific Advisory Committee (ESAC30), 12 June 2009.
- (21) OECD (2013). Summary Document on the Statistical Performance of Methods in OECD Test Guideline 431 for Sub-categorisation. Environment, Health, and Safety Publications, Series on Testing and Assessment (No 190). Organisation for Economic Cooperation and Development, Paris.
- (22) Alépée N., Grandidier M.H., and Cotovio J. (2014). Sub-Categorisation of Skin Corrosive Chemicals by the EpiSkin™ Reconstructed Human Epidermis Skin Corrosion Test Method According to UN GHS: Revision of OECD Test Guideline 431. *Toxicol. In Vitro* 28:131-145.
- (23) Desprez B., Barroso J., Griesinger C., Kandárová H., Alépée N., and Fuchs, H. (2015). Two Novel Prediction Models Improve Predictions of Skin Corrosive Sub-categories by Test Methods of OECD Test Guideline No 431. *Toxicol. In Vitro* 29:2055-2080.
- (24) OECD (2015). Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Reconstructed Human Epidermis (RHE) Test Methods For Skin Corrosion in Relation to OECD TG 431. Environmental Health and Safety Publications, Series on Testing and Assessment (No 219). Organisation for Economic Cooperation and Development, Paris
- (25) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environment, Health and Safety Publications, Series on Testing and Assessment (No 34), Organisation for Economic Cooperation and Development, Paris.
- (26) Eskes C. *et al.* (2012). Regulatory Assessment of *In Vitro* Skin Corrosion and Irritation Data Within the European Framework: Workshop Recommendations. *Regul.Toxicol.Pharmacol.* 62:393-403.
- (27) Mosmann T. (1983). Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. *J. Immunol. Methods* 65:55-63.
- (28) Tinois E., *et al.* (1994). The Episkin Model: Successful Reconstruction of Human *Epidermis In Vitro*. In: *In Vitro Skin Toxicology*. Rougier A., Goldberg A.M and Maibach H.I. (Eds): 133-140.
- (29) Cannon C. L., Neal P.J., Southee J.A., Kubilus J. and Klausner M. (1994), New Epidermal Model for Dermal Irritancy Testing. *Toxicol.in Vitro* 8:889 - 891.
- (30) Ponc M., Boelsma E, Weerheim A, Mulder A, Bouwstra J and Mommaas M. (2000). Lipid and Ultrastructural Characterization of Reconstructed Skin Models. *Inter. J. Pharmaceu.* 203:211 - 225.

- (31) Tinois E., Tillier, J., Gaucherand, M., Dumas, H., Tardy, M. and Thivolet J. (1991). *In Vitro* and Post - Transplantation Differentiation of Human Keratinocytes Grown on the Human Type IV Collagen Film of a Bilayered Dermal Substitute. *Exp. Cell Res.* 193:310-319.
- (32) Parenteau N.L., Bilbo P, Nolte CJ, Mason VS and Rosenberg M. (1992). The Organotypic Culture of Human Skin Keratinocytes and Fibroblasts to Achieve Form and Function. *Cytotech.* 9:163-171.
- (33) Wilkins L.M., Watson SR, Prosky SJ, Meunier SF and Parenteau N.L. (1994). Development of a Bilayered Living Skin Construct for Clinical Applications. *Biotech. Bioeng.* 43/8:747-756.
- (34) EpiSkin™ SOP (December 2011). INVITTOX Protocol (No 118). EpiSkin™ Skin Corrosivity Test.
- (35) EpiDerm™ SOP (February 2012). Version MK-24-007-0024 Protocol for: *In Vitro* EpiDerm™ Skin Corrosion Test (EPI-200-SCT), for Use with MatTek Corporation's Reconstructed Human Epidermal Model EpiDerm.
- (36) SkinEthic™ RHE SOP (January 2012). INVITTOX Protocol SkinEthic™ Skin Corrosivity Test.
- (37) EpiCS® SOP (January 2012). Version 4.1 *In Vitro* Skin Corrosion: Human Skin Model Test Epidermal Skin Test 1000 (epiCS®) CellSystems.
- (38) Alépée N., Barroso J., De Smedt A., De Wever B., Hibatallah J., Klaric M., Mewes K.R., Millet M., Pfannenbecker U., Tailhardat M., Templier M., and McNamee P. Use of HPLC/UPLC- spectrophotometry for Detection of MTT Formazan in *In Vitro* Reconstructed Human Tissue (RhT)- based Test Methods Employing the MTT Assay to Expand their Applicability to Strongly Coloured Test Chemicals. *Toxicol. In Vitro* 29: 741-761.
- (39) US FDA (2001). Guidance for Industry: Bioanalytical Method Validation. U.S. Department of Health and Human Services, Food and Drug Administration. (May 2001). Available at: [<http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf>].

Appendix 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with “concordance” to mean the proportion of correct outcomes of a test method (25).

Cell viability: Parameter measuring total activity of a cell population *e.g.* as ability of cellular mitochondrial dehydrogenases to reduce the vital dye MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue), which depending on the endpoint measured and the test design used, correlates with the total number and/or vitality of living cells.

Chemical: A substance or a mixture.

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (25).

ET₅₀: Can be estimated by determination of the exposure time required to reduce cell viability by 50 % upon application of the benchmark chemical at a specified, fixed concentration, see also IC₅₀.

GHS (Globally Harmonized System of Classification and Labelling of Chemicals): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

HPLC: High Performance Liquid Chromatography.

IATA: Integrated Approach on Testing and Assessment.

IC₅₀: Can be estimated by determination of the concentration at which a benchmark chemical reduces the viability of the tissues by 50 % (IC₅₀) after a fixed exposure time, see also ET₅₀.

Infinite dose: Amount of test chemical applied to the *epidermis* exceeding the amount required to completely and uniformly cover the *epidermis* surface.

Mixture: A mixture or solution composed of two or more substances in which they do not react.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

MTT: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide.

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration > 10 % (w/w) and < 80 % (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

NC: Non corrosive.

NSC_{killed} control: Non-Specific Colour control in killed tissues.

NSC_{living} control: Non-Specific Colour control in living tissues.

NSMTT: Non-Specific MTT reduction.

OD: Optical Density

PC: Positive Control, a replicate containing all components of a test system and treated with a chemical known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of reliability and accuracy, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (25).

Relevance: Description of relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test method correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (25).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility (25).

Run: A run consists of one or more test chemicals tested concurrently with a negative control and with a PC.

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (25).

Skin corrosion in vivo: The production of irreversible damage of the skin; namely, visible necrosis through the *epidermis* and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (25).

Substance: A chemical element and its compounds in the natural state or obtained by any production process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Test chemical: Any substance or mixture tested using this test method.

UPLC: Ultra-High Performance Liquid Chromatography.

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

Appendix 2

MAIN COMPONENTS OF THE RHE TEST MODELS VALIDATED FOR SKIN CORROSION TESTING

Test Model Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
Model surface	0,38 cm ²	0,63 cm ²	0,5 cm ²	0,6 cm ²
Number of tissue replicates	At least 2 per exposure time	2-3 per exposure time	At least 2 per exposure time	At least 2 per exposure time
Treatment doses and application	Liquids and viscous: 50 µl ± 3 µl (131.6 µl/cm ²) Solids: 20 ± 2 mg (52.6 mg/cm ²) + 100 µl ± 5µl NaCl solution (9 g/l) Waxy/sticky: 50 ± 2 mg (131.6 mg/cm ²) with a nylon mesh	Liquids: 50 µl (79.4 µl/cm ²) with or without a nylon mesh Pre-test compatibility of test chemical with nylon mesh Semisolids: 50 µl (79.4 µl/cm ²) Solids: 25 µl H ₂ O (or more if necessary) + 25 mg (39.7 mg/cm ²) Waxes: flat "disc like" piece of ca. 8 mm diameter placed atop the tissue wetted with 15 µl H ₂ O.	Liquids and viscous: 40 µl ± 3µl (80 µl/cm ²) using nylon mesh Pre-test compatibility of test chemical with nylon mesh Solids: 20 µl ± 2µl H ₂ O + 20 ± 3 mg (40 mg/cm ²) Waxy/sticky: 20 ± 3 mg (40 mg/cm ²) using nylon mesh	Liquids: 50 µl (83.3 µl/cm ²) using nylon mesh Pre-test compatibility of test chemical with nylon mesh Semisolids: 50 µl (83.3 µl/cm ²) Solids: 25 mg (41.7 mg/cm ²) + 25 µl H ₂ O (or more if necessary) Waxy: flat "cookie like" piece of ca. 8 mm diameter placed atop the tissue wetted with 15 µl H ₂ O

Test Model Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
Pre-check for direct MTT reduction	<p>50 µl (liquid) or 20 mg (solid)+ 2 ml MTT</p> <p>0.3 mg/ml solution for 180 ± 5 min at 37°C, 5 % CO₂, 95 % RH</p> <p>→ if solution turns blue/purple, water-killed adapted controls should be performed</p>	<p>50 µl (liquid) or 25 mg (solid)+ 1 ml MTT</p> <p>1 mg/ml solution for 60 min at 37°C, 5 % CO₂, 95 % RH</p> <p>→ if solution turns blue/purple, freeze-killed adapted controls should be performed</p>	<p>40 µl (liquid) or 20 mg (solid)+ 1 ml MTT</p> <p>1 mg/ml solution for 180± 15 min at 37°C, 5 % CO₂, 95 % RH</p> <p>→ if solution turns blue/purple, freeze-killed adapted controls should be performed</p>	<p>50 µl (liquid) or 25 mg (solid)+ 1 ml MTT</p> <p>1 mg/ml solution for 60 min at 37°C, 5 % CO₂, 95 % RH</p> <p>→ if solution turns blue/purple, freeze-killed adapted controls should be performed</p>
Pre-check for colour interference	<p>10 µl (liquid) or 10 mg (solid) + 90 µl H₂O mixed for 15 min at RT</p> <p>→ if solution becomes coloured, living adapted controls should be performed</p>	<p>50 µl (liquid) or 25 mg (solid) + 300 µl H₂O for 60 min at 37°C, 5 % CO₂, 95 % RH</p> <p>→ if solution becomes coloured, living adapted controls should be performed</p>	<p>40 µl (liquid) or 20mg (solid) + 300 µl H₂O mixed for 60 min at RT</p> <p>→ if test chemical is coloured, living adapted controls should be performed</p>	<p>50 µl (liquid) or 25 mg (solid) + 300 µl H₂O for 60 min at 37°C, 5 % CO₂, 95 % RH</p> <p>→ if solution becomes coloured, living adapted controls should be performed</p>
Exposure time and temperature	<p>3 min, 60 min (± 5 min) and 240 min (± 10 min)</p> <p>In ventilated cabinet Room Temperature (RT, 18-28 °C)</p>	<p>3 min at RT, and 60 min at 37°C, 5 % CO₂, 95 % RH</p>	<p>3 min at RT, and 60 min at 37°C, 5 % CO₂, 95 % RH</p>	<p>3 min at RT, and 60 min at 37°C, 5 % CO₂, 95 % RH</p>
Rinsing	<p>2.5 ml 1x PBS (2 ml/throwing)</p>	<p>20 times with a constant soft stream of 1x PBS</p>	<p>20 times with a constant soft stream of 1x PBS</p>	<p>20 times with a constant soft stream of 1x PBS</p>

Test Model Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	eptCS®
Negative control	50 µl NaCl solution (9 g/l) Tested with every exposure time	50 µl H ₂ O Tested with every exposure time	40 µl H ₂ O Tested with every exposure time	50 µl H ₂ O Tested with every exposure time
Positive control	50 µl Glacial acetic acid Tested only for 4 hours	50 µl 8N KOH Tested with every exposure time	40 µl 8N KOH Tested only for 1 hour	50 µl 8N KOH Tested with every exposure time
MTT solution	2 ml 0,3 mg/ml	300 µl 1 mg/ml	300 µl 1 mg/ml	300 µl 1 mg/ml
MTT incubation time and temperature	180 min (± 15 min) at 37°C, 5 % CO ₂ , 95 % RH	180 min at 37°C, 5 % CO ₂ , 95 % RH	180 min (± 15 min) at 37°C, 5 % CO ₂ , 95 % RH	180 min at 37°C, 5 % CO ₂ , 95 % RH
Extraction solvent	500 µl acidified isopropanol (0.04 N HCl in isopropanol) (isolated tissue fully immersed)	2 ml isopropanol (extraction from top and bottom of insert)	1.5 ml isopropanol (extraction from top and bottom of insert)	2 ml isopropanol (extraction from top and bottom of insert)
Extraction time and temperature	Overnight at RT, protected from light	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT

Test Model Components	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®
OD reading	570 nm (545 - 595 nm) without reference filter	570 nm (or 540 nm) without reference filter	570 nm (540 - 600 nm) without reference filter	540 - 570 nm without reference filter
Tissue Quality Control	18 hours treatment with SDS 1.0 mg/ml \leq IC ₅₀ \leq 3.0 mg/ml	Treatment with 1 % Triton X-100 4.08 hours \leq ET ₅₀ \leq 8.7 hours	Treatment with 1 % Triton X-100 4.0 hours \leq ET ₅₀ \leq 10.0 hours	Treatment with 1 % Triton X-100 2.0 hours \leq ET ₅₀ \leq 7.0 hours
Acceptability Criteria	<p>1. Mean OD of the tissue replicates treated with the negative control (NaCl) should be \geq 0.6 and \leq 1.5 for every exposure time</p> <p>2. Mean viability of the tissue replicates exposed for 4 hours with the positive control (glacial acetic acid), expressed as % of the negative control, should be \leq 20 %</p> <p>3. In the range 20-100 % viability and for ODs \geq 0.3, difference of viability between the two tissue replicates should not exceed 30 %.</p>	<p>1. Mean OD of the tissue replicates treated with the negative control (H₂O) should be \geq 0.8 and \leq 2.8 for every exposure time</p> <p>2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be $<$ 15 %</p> <p>3. In the range 20 - 100 % viability, the Coefficient of Variation (CV) between tissue replicates should be 30 %</p>	<p>1. Mean OD of the tissue replicates treated with the negative control (H₂O) should be \geq 0.8 and \leq 3.0 for every exposure time</p> <p>2. Mean viability of the tissue replicates exposed for 1 hour (and 4 hours, if applicable) with the positive control (8N KOH), expressed as % of the negative control, should be 15 %</p> <p>3. In the range 20-100 % viability, and for ODs \geq 0.3, difference of viability between the two tissue replicates should not exceed 30 %</p>	<p>1. Mean OD of the tissue replicates treated with the negative control (H₂O) should be \geq 0.8 and \leq 2.8 for every exposure time</p> <p>2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be 20 %</p> <p>3. In the range 20-100 % viability, and for ODs \geq 0.3, difference of viability between the two tissue replicates should not exceed 30 %</p>

Appendix 3

PERFORMANCE OF TEST MODELS FOR SUB-CATEGORISATION

The table below provides the performances of the four test models calculated based on a set of 80 chemicals tested by the four test developers. Calculations were performed by the OECD Secretariat, reviewed and agreed by an expert subgroup (21) (23).

EpiSkin™, EpiDerm™, SkinEthic™ and epiCS® test models are able to sub-categorise (i.e. 1A versus 1B-and-1C versus NC)

Performances, overclassification rates, underclassification rates, and accuracy (Predictive capacity) of the four test models based on a set of 80 chemicals all tested over 2 or 3 runs in each test model:

STATISTICS ON PREDICTIONS OBTAINED ON THE ENTIRE SET OF CHEMICALS				
(n= 80 chemicals tested over 2 independent runs for epiCS® or 3 independent runs for EpiDerm™ SCT, EpiSkin™ and SkinEthic™ RHE, i.e. respectively 159 (*) or 240 classifications)				
	EpiSkin™	EpiDerm™	SkinEthic™	epiCS®
Overclassifications:				
1B-and-1C overclassified 1A	21,50 %	29,0 %	31,2 %	32,8 %
NC overclassified 1B-and-1C	20,7 %	23,4 %	27,0 %	28,4 %
NC overclassified 1A	0,00 %	2,7 %	0,0 %	0,00 %
overclassified Corr.	20,7 %	26,1 %	27,0 %	28,4 %
Global overclassification rate (all categories)	17,9 %	23,3 %	24,5 %	25,8 %
Underclassifications:				
1A underclassified 1B-and-1C	16,7 %	16,7 %	16,7 %	12,5 %
1A underclassified NC	0,00 %	0,00 %	0,00 %	0,00 %
1B-and-1C underclassified NC	2,2 %	0,00 %	7,5 %	6,6 %
Global underclassification rate (all categories)	3,3 %	2,5 %	5,4 %	4,4 %
Correct Classifications:				
1A correctly classified	83,3 %	83,3 %	83,3 %	87,5 %
1B-and-1C correctly classified	76,3 %	71,0 %	61,3 %	60,7 %
NC correctly classified	79,3 %	73,9 %	73,0 %	71,62 %
Overall Accuracy	78,8 %	74,2 %	70 %	69,8 %

NC: Non-corrosive

(*) one chemical was tested once in epiCS® because of no availability (23)

Appendix 4

Key parameters and acceptance criteria for qualification of an HPLC/UPLC-spectrophotometry system for measurement of MTT formazan extracted from RhE tissue

Parameter	Protocol Derived from FDA Guidance (37)(38)	Acceptance Criteria
Selectivity	Analysis of isopropanol, living blank (isopropanol extract from living RhE tissues without any treatment), dead blank (isopropanol extract from killed RhE tissues without any treatment)	$\text{Area}_{\text{interference}} \leq 20 \% \text{ of Area}_{\text{LLOQ}}^{(1)}$
Precision	Quality Controls (i.e., MTT formazan at 1,6 µg/ml, 16 µg/ml and 160 µg/ml) in isopropanol (n=5)	$\text{CV} \leq 15 \% \text{ or } \leq 20 \% \text{ for the LLOQ}$
Accuracy	Quality Controls in isopropanol (n=5)	$\% \text{Dev} \leq 15 \% \text{ or } \leq 20 \% \text{ for LLOQ}$
Matrix Effect	Quality Controls in living blank (n=5)	$85 \% \leq \text{Matrix Effect } \% \leq 115 \%$
Carryover	Analysis of isopropanol after an ULOQ ⁽²⁾ standard	$\text{Area}_{\text{interference}} \leq 20 \% \text{ of Area}_{\text{LLOQ}}$
Reproducibility (intra-day)	3 independent calibration curves (based on 6 consecutive 1/3 dilutions of MTT formazan in isopropanol starting at ULOQ, i.e., 200 µg/ml); Quality Controls in isopropanol (n=5)	Calibration Curves: $\% \text{Dev} \leq 15 \% \text{ or } \leq 20 \% \text{ for LLOQ}$ Quality Controls: $\% \text{Dev} \leq 15 \% \text{ and } \text{CV} \leq 15 \%$
Reproducibility (inter-day)	Day 1: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 2: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 3: 1 calibration curve and Quality Controls in isopropanol (n=3)	
Short Term Stability of MTT Formazan in RhE Tissue Extract	Quality Controls in living blank (n=3) analysed the day of the preparation and after 24 hours of storage at room temperature	$\% \text{Dev} \leq 15 \%$
Long Term Stability of MTT Formazan in RhE Tissue Extract, if required	Quality Controls in living blank (n=3) analysed the day of the preparation and after several days of storage at a specified temperature (e.g., 4 °C, -20 °C, -80 °C)	$\% \text{Dev} \leq 15 \%$

⁽¹⁾ LLOQ: Lower Limit of Quantification, defined to cover 1-2 % tissue viability, i.e., 0,8 µg/ml.

⁽²⁾ ULOQ: Upper Limit of Quantification, defined to be at least two times higher than the highest expected MTT formazan concentration in isopropanol extracts from negative controls i.e., 200 µg/ml.

(7) In Part B, Chapter B.46 is replaced by the following:

"B.46 IN VITRO SKIN IRRITATION: RECONSTRUCTED HUMAN EPIDERMIS TEST METHOD

INTRODUCTION

1. This test method (TM) is equivalent to OECD test guideline (TG) 439 (2015). Skin irritation refers to the production of reversible damage to the skin following the application of a test chemical for up to 4 hours [as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS)](1) and the European Union (EU) Regulation 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) (1). This test method provides an *in vitro* procedure that may be used for the hazard identification of irritant chemicals (substances and mixtures) in accordance with UN GHS/CLP Category 2 (2). In regions that do not adopt the optional UN GHS Category 3 (mild irritants), this test method can also be used to identify non-classified chemicals. Therefore, depending on the regulatory framework and the classification system in use, this test method may be used to determine the skin irritancy of chemicals either as a stand-alone replacement test for *in vivo* skin irritation testing or as a partial replacement test within a testing strategy (3).
2. The assessment of skin irritation has typically involved the use of laboratory animals [TM B.4, equivalent to OECD TG 404 originally adopted in 1981 and revised in 1992, 2002 and 2015] (4). For the testing of corrosivity, three validated *in vitro* test methods have been adopted as EU TM B.40 (equivalent to OECD TG 430), TM B.40bis (equivalent to OECD TG 431) and TM B.65 (equivalent to OECD TG 435) (5) (6) (7). An OECD guidance document on Integrated Approaches to Testing and Assessment (IATA) for Skin Corrosion and Irritation describes several modules which group information sources and analysis tools, and provides guidance on (i) how to integrate and use existing test and non-test data for the assessment of skin irritation and skin corrosion potentials of chemicals and (ii) proposes an approach when further testing is needed (3).
3. This test method addresses the human health endpoint skin irritation. It is based on the *in vitro* test system of reconstructed human *epidermis* (RhE), which closely mimics the biochemical and physiological properties of the upper parts of the human skin, *i.e.* the *epidermis*. The RhE test system uses human derived non-transformed keratinocytes as cell source to reconstruct an epidermal model with representative histology and cytoarchitecture. Performance Standards (PS) are available to facilitate the validation and assessment of similar and modified RhE-based test methods, in accordance with the principles of the OECD guidance document No 34 (8) (9). The corresponding test guideline was originally adopted in 2010, updated in 2013 to include additional RhE models, and updated in 2015 to refer to the IATA guidance document and introduce the use of an alternative procedure to measure viability.
4. Pre-validation, optimisation and validation studies have been completed for four commercially available *in vitro* test models (10) (11) (12) (13) (14) (15) (16) (17) (18) (19) (20) (21) (22) (23) (24) (25) (26) (27) (28) based on the RhE test system (sensitivity 80 %, specificity 70 %, and accuracy 75 %). These four test models are included in this TM and are listed in Appendix 2, which also provides information on the type of validation study used to validate the respective test methods. As noted in Appendix 2, the Validated Reference Method (VRM) have been used to develop the present test method and the Performance Standards (8).
5. OECD Mutual Acceptance of Data will only be guaranteed for test models validated according to the Performance Standards (8), if these test models have been reviewed and adopted by OECD. The test models included in this test method and the corresponding OECD TG can be used indiscriminately to address countries' requirements for test results from *in vitro* test methods for skin irritation, while benefiting from the Mutual Acceptance of Data.
6. Definitions of terms used in this document are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

7. A limitation of the test method, as demonstrated by the full prospective validation study assessing and characterising RhE test methods (16), is that it does not allow the classification of chemicals to the optional UN GHS Category 3 (mild irritants) (1). Thus, the regulatory framework in member countries will decide how this test method will be used. For the EU, Category 3 has not been taken up in CLP. For a full evaluation of local skin effects after a single dermal exposure, the OECD Guidance Document on Integrated Approaches for Testing and Assessment should be consulted (3). It is recognised that the use of human skin is subject to national and international ethical considerations and conditions.

(1) Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006, OJ L 353/1, 31.12.2008

8. This test method addresses the human health endpoint skin irritation. While this test method does not provide adequate information on skin corrosion, it should be noted that TM B.40bis (equivalent to OECD TG 431) on skin corrosion is based on the same RhE test system, though using another protocol (6). This test method is based on RhE-models using human keratinocytes, which therefore represent *in vitro* the target organ of the species of interest. It moreover directly covers the initial step of the inflammatory cascade/mechanism of action (cell and tissue damage resulting in localised trauma) that occurs during irritation *in vivo*. A wide range of chemicals has been tested in the validation underlying this test method and the database of the validation study amounted to 58 chemicals in total (16) (18) (23). The test method is applicable to solids, liquids, semi-solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other pre-treatment of the sample is required. Gases and aerosols have not been assessed yet in a validation study (29). While it is conceivable that these can be tested using RhE technology, the current test method does not allow testing of gases and aerosols.
9. Before use of the test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture. However, due to the fact that mixtures cover a wide spectrum of categories and composition, and that only limited information is currently available on the testing of mixtures, in cases where evidence can be demonstrated on the non-applicability of the test method to a specific category of mixtures (e.g. following a strategy as proposed in Eskes *et al.* 2012 (30)), the test method should not be used for that specific category of mixtures. Similar care should be taken in case specific chemical classes or physico-chemical properties are found not to be applicable to the current test method.
10. Test chemicals absorbing light in the same range as MTT formazan and test chemicals able to directly reduce the vital dye MTT (to MTT formazan) may interfere with the cell viability measurements and need the use of adapted controls for corrections (see paragraphs 28-34).
11. A single testing run composed of three replicate tissues should be sufficient for a test chemical when the classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements and/or mean percent viability equal to $50 \pm 5 \%$, a second run should be considered, as well as a third one in case of discordant results between the first two runs.

PRINCIPLE OF THE TEST

12. The test chemical is applied topically to a three-dimensional RhE model, comprised of non-transformed human-derived epidermal keratinocytes, which have been cultured to form a multilayered, highly differentiated model of the human *epidermis*. It consists of organised basal, spinous and granular layers, and a multilayered *stratum corneum* containing intercellular lamellar lipid layers representing main lipid classes analogous to those found *in vivo*.
13. Chemical-induced skin irritation, manifested mainly by erythema and oedema, is the result of a cascade of events beginning with penetration of the chemicals through the *stratum corneum* where they may damage the underlying layers of keratinocytes and other skin cells. The damaged cells may either release inflammatory mediators or induce an inflammatory cascade which also acts on the cells in the *dermis*, particularly the stromal and endothelial cells of the blood vessels. It is the dilation and increased permeability of the endothelial cells that produce the observed erythema and oedema (29). Notably, the RhE-based test methods, in the absence of any vascularisation in the *in vitro* test system, measure the initiating events in the cascade, e.g. cell / tissue damage (16) (17), using cell viability as readout.
14. Cell viability in RhE models is measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue; CAS number 298-93-1], into a blue formazan salt that is quantitatively measured after extraction from tissues (31). Irritant chemicals are identified by their ability to decrease cell viability below defined threshold levels (*i.e.* $\leq 50 \%$, for UN GHS/CLP Category 2). Depending on the regulatory framework and applicability of the test method, test chemicals that produce cell viabilities above the defined threshold level, may be considered non-irritants (*i.e.* $> 50 \%$, No Category).

DEMONSTRATION OF PROFICIENCY

15. Prior to routine use of any of the four validated test models that adhere to this test method (Appendix 2), laboratories should demonstrate technical proficiency, using the ten Proficiency Substances listed in Table 1. In situations where, for instance, a listed substance is unavailable, another substance for which adequate *in vivo* and *in vitro* reference data are available may be used (e.g. from the list of reference chemicals (8)) provided that the same selection criteria as described in Table 1 are applied. Using an alternative proficiency substance should be justified.
16. As part of the proficiency testing, it is recommended that users verify the barrier properties of the tissues after receipt as specified by the RhE model producer. This is particularly important if tissues are shipped over long distance/time periods. Once a test method has been successfully established and proficiency in its use has been acquired and demonstrated, such verification will not be necessary on a routine basis. However, when using a test method routinely, it is recommended to continue to assess the barrier properties at regular intervals.

Table 1

Proficiency Substances ⁽¹⁾

Substance	CAS NR	<i>In vivo</i> score ⁽²⁾	Physical state	UN GHS Category
NON-CLASSIFIED SUBSTANCES (UN GHS No Category)				
naphthalene acetic acid	86-87-3	0	Solid	No Cat.
isopropanol	67-63-0	0,3	Liquid	No Cat.
methyl stearate	112-61-8	1	Solid	No Cat.
heptyl butyrate	5870-93-9	1,7	Liquid	No Cat.(Optional Cat. 3) ⁽³⁾
hexyl salicylate	6259-76-3	2	Liquid	No Cat.(Optional Cat. 3) ⁽³⁾
CLASSIFIED SUBSTANCES (UN GHS Category 2)				
cyclamen aldehyde	103-95-7	2,3	Liquid	Cat. 2
1-bromohexane	111-25-1	2,7	Liquid	Cat. 2
potassium hydroxide (5 % aq.)	1310-58-3	3	Liquid	Cat. 2
1-methyl-3-phenyl-1-piperazine	5271-27-2	3,3	Solid	Cat. 2
heptanal	111-71-7	3,4	Liquid	Cat. 2

⁽¹⁾ The Proficiency Substances are a subset of the substances used in the validation study and the selection is based on the following criteria; (i), the chemicals substances are commercially available; (ii), they are representative of the full range of Draize irritancy scores (from non-irritant to strong irritant); (iii), they have a well-defined chemical structure; (iv), they are representative of the chemical functionality used in the validation process; (v) they provided reproducible *in vitro* results across multiple testing and multiple laboratories; (vi) they were correctly predicted *in vitro*, and (vii) they are not associated with an extremely toxic profile (e.g. carcinogenic or toxic to the reproductive system) and they are not associated with prohibitive disposal costs.

⁽²⁾ *In vivo* score in accordance with TM B.4 (4).

⁽³⁾ Under this test method, the UN GHS optional Category 3 (mild irritants) (1) is considered as No Category.

PROCEDURE

17. The following is a description of the components and procedures of a RhE test method for skin irritation assessment (See also Appendix 3 for parameters related to each test model). Standard Operating Procedures (SOPs) for the four models complying with this test method are available (32) (33) (34) (35).

RHE TEST METHOD COMPONENTS

General conditions

18. Non -transformed human keratinocytes should be used to reconstruct the epithelium. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum*, *stratum granulosum*) should be present under a functional *stratum corneum*. *Stratum corneum* should be multilayered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic benchmark chemicals, e.g. sodium dodecyl sulphate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a benchmark chemical reduces the viability of the tissues by 50 % (IC₅₀) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50 % (ET₅₀) upon application of the benchmark chemical at a specified, fixed concentration. The containment properties of the RhE model should prevent the passage of material around the *stratum corneum* to the viable tissue, which would lead to poor modelling of skin exposure. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, or fungi.

Functional conditions*Viability*

19. The assay used for quantifying viability is the MTT-assay (31). The viable cells of the RhE tissue construct can reduce the vital dye MTT into a blue MTT formazan precipitate which is then extracted from the tissue using isopropanol (or a similar solvent). The optical density (OD) of the extraction solvent alone should be sufficiently small, i.e. OD < 0.1. The extracted MTT formazan may be quantified using either a standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure (36). The RhE model users should ensure that each batch of the RhE model used meets defined criteria for the negative control. An acceptability range (upper and lower limit) for the negative control OD values (in the Skin Irritation test method conditions) are established by the RhE model developer/supplier. Acceptability ranges for the four validated RhE models included in this test method are given in Table 2. An HPLC/UPLC-Spectrophotometry user should use the negative control OD ranges provided in Table 2 as the acceptance criterion for the negative control. It should be documented that the tissues treated with the negative control are stable in culture (provide similar viability measurements) for the duration of the test exposure period.

Table 2

Acceptability ranges for negative control OD values of the test models included in this TM

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM)	≥ 0,6	≤ 1,5
EpiDerm™ SIT (EPI-200)	≥ 0,8	≤ 2,8
SkinEthic™ RHE	≥ 0,8	≤ 3,0
LabCyte EPI-MODEL24 SIT	≥ 0,7	≤ 2,5

Barrier function

20. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of cytotoxic benchmark chemicals, e.g. SDS or Triton X-100, as estimated by IC₅₀ or ET₅₀ (Table 3).

Morphology

21. Histological examination of the RhE model should be provided demonstrating human *epidermis*-like structure (including multilayered *stratum corneum*).

Reproducibility

22. The results of the positive and negative controls of the test method should demonstrate reproducibility over time.

Quality control (QC)

23. The RhE model should only be used if the developer/supplier demonstrates that each batch of the RhE model used meets defined production release criteria, among which those for *viability* (paragraph 19), *barrier function* (paragraph 20) and *morphology* (paragraph 21) are the most relevant. These data should be provided to the test method users, so that they are able to include this information in the test report. An acceptability range (upper and lower limit) for the IC₅₀ or the ET₅₀ should be established by the RhE model developer/supplier. Only results produced with qualified tissues can be accepted for reliable prediction of irritation classification. The acceptability ranges for the four test models included in this TM are given in Table 3.

Table 3

QC batch release criteria of the test models included in this TM

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM) (18 hours treatment with SDS) (32)	IC ₅₀ = 1,0 mg/ml	IC ₅₀ = 3,0 mg/ml
EpiDerm™ SIT (EPI-200) (1 % Triton X-100) (33)	ET ₅₀ = 4,0 hr	ET ₅₀ = 8,7 hr
SkinEthic™ RHE (1 % Triton X-100) (34)	ET ₅₀ = 4,0 hr	ET ₅₀ = 10,0 hr
LabCyte EPI-MODEL24 SIT (18 hours treatment with SDS) (35)	IC ₅₀ = 1,4 mg/ml	IC ₅₀ = 4,0 mg/ml

Application of the Test Chemical and Control Chemicals

24. At least three replicates should be used for each test chemical and for the controls in each run. For liquid as well as solid chemicals, sufficient amount of test chemical should be applied to uniformly cover the *epidermis* surface while avoiding an infinite dose, *i.e.* ranging from 26 to 83 l/cm² or mg/cm² (see Appendix 3). For solid chemicals, the *epidermis* surface should be moistened with deionised or distilled water before application, to improve contact between the test chemical and the *epidermis* surface. Whenever possible, solids should be tested as a fine powder. A nylon mesh may be used as a spreading aid in some cases (see Appendix 3). At the end of the exposure period, the test chemical should be carefully washed from the *epidermis* surface with aqueous buffer, or 0,9 % NaCl. Depending on the RhE test models used, the exposure period ranges between 15 and 60 minutes, and the incubation temperature between 20 and 37 °C. These exposure periods and temperatures are optimised for each individual RhE test method and represent the different intrinsic properties of the test models (*e.g.* barrier function) (see Appendix 3).
25. Concurrent negative control (NC) and positive control (PC) should be used in each run to demonstrate that *viability* (using the NC), *barrier function* and resulting tissue sensitivity (using the PC) of the tissues are within a defined historical acceptance range. The suggested PC is 5 % aqueous SDS. The suggested NCs is either water or phosphate buffered saline (PBS).

Cell Viability Measurements

26. According to the test procedure, it is essential that the viability measurement is not performed immediately after exposure to the test chemical, but after a sufficiently long post-treatment incubation period of the rinsed tissue in fresh medium. This period allows both for recovery from weak cytotoxic effects and for appearance of clear cytotoxic effects. A 42 hours post-treatment incubation period was found optimal during test optimisation of two of the RhE-based test models underlying this test method (11) (12) (13) (14) (15).
27. The MTT assay is a standardised quantitative method which should be used to measure cell viability under this test method. It is compatible with use in a three-dimensional tissue construct. The tissue sample is placed in MTT solution of appropriate concentration (e.g. 0,3 - 1 mg/ml) for 3 hours. The MTT is converted into blue formazan by the viable cells. The precipitated blue formazan product is then extracted from the tissue using a solvent (e.g. isopropanol, acidic isopropanol), and the concentration of formazan is measured by determining the OD at 570 nm using a filter band pass of maximum ± 30 nm or, by using an HPLC/UPLC-spectrophotometry procedure (see paragraph 34) (36).
28. Optical properties of the test chemical or its chemical action on MTT (e.g. chemicals may prevent or reverse the colour generation as well as cause it) may interfere with the assay leading to a false estimate of viability. This may occur when a specific test chemical is not completely removed from the tissue by rinsing or when it penetrates the *epidermis*. If a test chemical acts directly on the MTT (e.g. MTT-reducer), is naturally coloured, or becomes coloured during tissue treatment, additional controls should be used to detect and correct for test chemical interference with the viability measurement technique (see paragraphs 29 and 33). Detailed description of how to correct direct MTT reduction and interferences by colouring agents is available in the SOPs for the four validated models included in this test method (32) (33) (34) (35).
29. To identify direct MTT reducers, each test chemical should be added to freshly prepared MTT solution. If the MTT mixture containing the test chemical turns blue/purple, the test chemical is presumed to directly reduce MTT and a further functional check on non-viable RhE tissues should be performed, independently of using the standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure. This additional functional check employs killed tissues that possess only residual metabolic activity but absorb the test chemical in a similar way as viable tissues. Each MTT reducing test chemical is applied on at least two killed tissue replicates which undergo the entire testing procedure to generate a non-specific MTT reduction (NSMTT) (32) (33) (34) (35). A single NSMTT control is sufficient per test chemical regardless of the number of independent tests/runs performed. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the MTT reducer minus the percent non-specific MTT reduction obtained with the killed tissues exposed to the same MTT reducer, calculated relative to the negative control run concurrently to the test being corrected (%NSMTT).
30. To identify potential interference by coloured test chemicals or test chemicals that become coloured when in contact with water or isopropanol and decide on the need for additional controls, spectral analysis of the test chemical in water (environment during exposure) and/or isopropanol (extracting solution) should be performed. If the test chemical in water and/or isopropanol absorbs light in the range of 570 ± 30 nm, further colorant controls should be performed or, alternatively, an HPLC/UPLC-spectrophotometry procedure should be used in which case these controls are not required (see paragraphs 33 and 34). When performing the standard absorbance (OD) measurement, each interfering coloured test chemical is applied on at least two viable tissue replicates, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step to generate a non-specific colour (NSC_{living}) control. The NSC_{living} control needs to be performed concurrently to the testing of the coloured test chemical and in case of multiple testing, an independent NSC_{living} control needs to be conducted with each test performed (in each run) due to the inherent biological variability of living tissues. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the interfering test chemical and incubated with MTT solution minus the percent non-specific colour obtained with living tissues exposed to the interfering test chemical and incubated with medium without MTT, run concurrently to the test being corrected (%NSC_{living}).
31. Test chemicals that are identified as producing both direct MTT reduction (see paragraph 29) and colour interference (see paragraph 30) will also require a third set of controls, apart from the NSMTT and NSC_{living} controls described in the previous paragraphs, when performing the standard absorbance (OD) measurement. This is usually the case with darkly coloured test chemicals interfering with the MTT assay (e.g. blue, purple, black) because their intrinsic colour impedes the assessment of their capacity to directly reduce MTT as described in paragraph 29. These test chemicals may bind to both living and killed tissues and therefore the

NSMTT control may not only correct for potential direct MTT reduction by the test chemical, but also for colour interference arising from the binding of the test chemical to killed tissues. This could lead to a double correction for colour interference since the NSC_{living} control already corrects for colour interference arising from the binding of the test chemical to living tissues. To avoid a possible double correction for colour interference, a third control for non-specific colour in killed tissues (NSC_{killed}) needs to be performed. In this additional control, the test chemical is applied on at least two killed tissue replicates, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step. A single NSC_{killed} control is sufficient per test chemical regardless of the number of independent tests/runs performed, but should be performed concurrently to the NSMTT control and, where possible, with the same tissue batch. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT minus NSC_{living} plus the percent non-specific colour obtained with killed tissues exposed to the interfering test chemical and incubated with medium without MTT, calculated relative to the negative control run concurrently to the test being corrected (NSC_{killed}).

32. It is important to note that non-specific MTT reduction and non-specific colour interferences may increase the readouts of the tissue extract above the linearity range of the spectrophotometer. On this basis, each laboratory should determine the linearity range of their spectrophotometer with MTT formazan (CAS # 57360-69-7) from a commercial source before initiating the testing of test chemicals for regulatory purposes. The standard absorbance (OD) measurement using a spectrophotometer is appropriate to assess direct MTT-reducers and colour interfering test chemicals when the ODs of the tissue extracts obtained with the test chemical without any correction for direct MTT reduction and/or colour interference are within the linear range of the spectrophotometer or when the uncorrected percent viability obtained with the test chemical is already $\leq 50\%$. Nevertheless, results for test chemicals producing %NSMTT and/or $NSC_{\text{living}} \geq 50\%$ of the negative control should be taken with caution as this is the cut-off used to distinguish classified from not classified chemicals (see paragraph 36).
33. For coloured test chemicals which are not compatible with the standard absorbance (OD) measurement due to too strong interference with the MTT assay, the alternative HPLC/UPLC-spectrophotometry procedure to measure MTT formazan may be employed (see paragraph 34) (36). The HPLC/UPLC-spectrophotometry system allows for the separation of the MTT formazan from the test chemical before its quantification (36). For this reason, NSC_{living} or NSC_{killed} controls are never required when using HPLC/UPLC-spectrophotometry, independently of the chemical being tested. NSMTT controls should nevertheless be used if the test chemical is suspected to directly reduce MTT or has a colour that impedes the assessment of the capacity to directly reduce MTT (as described in paragraph 29). When using HPLC/UPLC-spectrophotometry to measure MTT formazan, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. For test chemicals able to directly reduce MTT, true tissue viability is calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT. Finally, it should be noted that direct MTT-reducers that may also be colour interfering, which are retained in the tissues after treatment and reduce MTT so strongly that they lead to ODs (using standard OD measurement) or peak areas (using UPLC/HPLC-spectrophotometry) of the tested tissue extracts that fall outside of the linearity range of the spectrophotometer cannot be assessed, although these are expected to occur in only very rare situations.
34. HPLC/UPLC-spectrophotometry may be used also with all types of test chemicals (coloured, non-coloured, MTT-reducers and non-MTT reducers) for measurement of MTT formazan (36). Due to the diversity of HPLC/UPLC-spectrophotometry systems, qualification of the HPLC/UPLC-spectrophotometry system should be demonstrated before its use to quantify MTT formazan from tissue extracts by meeting the acceptance criteria for a set of standard qualification parameters based on those described in the U.S. Food and Drug Administration guidance for industry on bio-analytical method validation (36) (37). These key parameters and their acceptance criteria are shown in Appendix 4. Once the acceptance criteria defined in Appendix 4 have been met, the HPLC/UPLC-spectrophotometry system is considered qualified and ready to measure MTT formazan under the experimental conditions described in this test method.

Acceptability Criteria

35. For each test method using valid RhE model batches (see paragraph 23), tissues treated with the negative control should exhibit OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes. Control OD values should not be below historically established boundaries. Similarly, tissues treated with the PC, i.e. 5 % aqueous SDS, should reflect their ability to respond to an irritant chemical under the conditions of the test method (see Appendix 3 and for further information SOPs of the four test models included in this TG (32) (33) (34) (35)). Associated and appropriate measures of variability between tissue replicates, i.e. standard deviations (SD) should fall within the acceptance limits established for the test model used (see Appendix 3).

Interpretation of Results and Prediction Model

36. The OD values obtained with each test chemical can be used to calculate the percentage of viability normalised to the negative control, which is set to 100 %. In case HPLC/UPLC-spectrophotometry is used, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. The cut-off value of percentage cell viability distinguishing irritant from non-classified test chemicals and the statistical procedure(s) used to evaluate the results and identify irritant chemicals should be clearly defined, documented, and proven to be appropriate (see SOPs of the test models for information). The cut-off values for the prediction of irritation are given below:
- The test chemical is identified as requiring classification and labelling according to UN GHS/CLP (Category 2 or Category 1) if the mean percent tissue viability after exposure and post-treatment incubation is less than or equal (\leq) to 50 %. Since the RhE test models covered by this test method cannot resolve between UN GHS/CLP Categories 1 and 2, further information on skin corrosion will be required to decide on its final classification [see also the OECD Guidance Document on IATA (3)]. In case the test chemical is found to be non-corrosive (e.g. based on TM.40, B.40bis or B.65), and shows tissue viability after exposure and post-treatment incubation is less than or equal (\leq) to 50 %, the test chemical is considered to be irritant to skin in accordance with UN GHS/CLP Category 2.
 - Depending on the regulatory framework in member countries, the test chemical may be considered as non-irritant to skin in accordance with UN GHS/CLP No Category if the tissue viability after exposure and post-treatment incubation is more than ($>$) 50 %.

DATA AND REPORTING

Data

37. For each run, data from individual replicate tissues (e.g. OD values and calculated percentage cell viability data for each test chemical, including classification) should be reported in tabular form, including data from repeat experiments as appropriate. In addition means \pm SD for each run should be reported. Observed interactions with MTT reagent and coloured test chemicals should be reported for each tested chemical.

Test Report

38. The test report should include the following information:

Test Chemical and Control Chemicals:

- Mono-constituent substance: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc;
- Multi-constituent substance, UVCB and mixture: characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents;
- Physical appearance, water solubility, and any additional relevant physicochemical properties;
- Source, lot number if available;
- Treatment of the test chemical/control chemicals prior to testing, if applicable (e.g. warming, grinding);
- Stability of the test chemical, limit date for use, or date for re-analysis if known;
- Storage conditions.

RhE model and protocol used (and rationale for the choice, if applicable)

Test Conditions:

- RhE model used (including batch number);
- Calibration information for measuring device (e.g. spectrophotometer), wavelength and band pass (if applicable) used for quantifying MTT formazan, and linearity range of measuring device; Description of the method used to quantify MTT formazan;
- Description of the qualification of the HPLC/UPLC-spectrophotometry system, if applicable; Complete supporting information for the specific RhE model used including its performance. This should include, but is not limited to;
 - i) Viability;
 - ii) Barrier function;
 - iii) Morphology;
 - iv) Reproducibility and predictivity;
 - v) Quality controls (QC) of the model;
- Reference to historical data of the model. This should include, but is not limited to acceptability of the QC data with reference to historical batch data.
- Demonstration of proficiency in performing the test method before routine use by testing of the proficiency substances.

Test Procedure:

- Details of the test procedure used (including washing procedures used after exposure period); Dose of test chemical and controls used;
- Duration and temperature of exposure and post-exposure incubation period;
- Indication of controls used for direct MTT-reducers and/or colouring test chemicals, if applicable;
- Number of tissue replicates used per test chemical and controls (PC, negative control, and NSMTT, NSC_{living} and NSC_{killed}, if applicable);
- Description of decision criteria/prediction model applied based on the RhE model used;
- Description of any modifications to the test procedure (including washing procedures).

Run and Test Acceptance Criteria:

- Positive and negative control mean values and acceptance ranges based on historical data; Acceptable variability between tissue replicates for positive and negative controls;
- Acceptable variability between tissue replicates for test chemical.

Results:

- Tabulation of data for individual test chemical for each run and each replicate measurement including OD or MTT formazan peak area, percent tissue viability, mean percent tissue viability and SD;
- If applicable, results of controls used for direct MTT-reducers and/or colouring test chemicals including OD or MTT formazan peak area, %NSMTT, %NSC_{living}, %NSC_{killed}, SD, final correct percent tissue viability;
- Results obtained with the test chemical(s) and controls in relation to the defined run and test acceptance criteria;
- Description of other effects observed;
- The derived classification with reference to the prediction model/decision criteria used.

*Discussion of the results**Conclusions***LITERATURE**

- (1) United Nations (UN) (2013). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Second Revised Edition, UN New York and Geneva, 2013. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html.
- (2) EURL-ECVAM (2009). Statement on the "Performance Under UN GHS of Three *In Vitro* Assays for Skin Irritation Testing and the Adaptation of the Reference Chemicals and Defined Accuracy Values of the ECVAM Skin Irritation Performance Standards", Issued by the ECVAM Scientific Advisory Committee (ESAC31), 9 April 2009. Available at: https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/publication//ESAC31_skin-irritation-statement_20090922.pdf
- (3) OECD (2014). Guidance Document on Integrated Approaches to Testing and Assessment for Skin Irritation/Corrosion. Environment, Health and Safety Publications, Series on Testing and Assessment (No 203), Organisation for Economic Cooperation and Development, Paris.
- (4) Chapter B.4 of this Annex, Acute Dermal Irritation, Corrosion.
- (5) Chapter B.40 of this Annex, *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance (TER).
- (6) Chapter B.40bis of this Annex, *In Vitro* Skin Corrosion: Reconstructed Human Epidermis (RHE) test method.
- (7) Chapter B.65 of this Annex, *In Vitro* Membrane Barrier Test Method.
- (8) OECD (2015). Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Reconstructed Human *Epidermis* (RhE) Test Methods for Skin Irritation in Relation to TG 439. Environment, health and Safety Publications, Series on Testing and Assessment (No 220). Organisation for Economic Cooperation and Development, Paris.
- (9) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environment, Health and Safety Publications, Series on Testing and Assessment (No 34) Organisation for Economic Cooperation and Development, Paris.

- (10) Fentem, J.H., Briggs, D., Chesné, C., Elliot, G.R., Harbell, J.W., Heylings, J.R., Portes, P., Roguet, R., van de Sandt, J.J. M. and Botham, P. (2001). A Prevalidation Study on *In Vitro* Tests for Acute Skin Irritation, Results and Evaluation by the Management Team, *Toxicol. in Vitro* 15, 57-93.
- (11) Portes, P., Grandidier, M.-H., Cohen, C. and Roguet, R. (2002). Refinement of the EPISKIN Protocol for the Assessment of Acute Skin Irritation of Chemicals: Follow-Up to the ECVAM Prevalidation Study, *Toxicol. in Vitro* 16, 765-770.
- (12) Kandárová, H., Liebsch, M., Genschow, E., Gerner, I., Traue, D., Slawik, B. and Spielmann, H. (2004). Optimisation of the EpiDerm Test Protocol for the Upcoming ECVAM Validation Study on *In Vitro* Skin Irritation Tests, *ALTEX* 21, 107-114.
- (13) Kandárová, H., Liebsch, M., Gerner, I., Schmidt, E., Genschow, E., Traue, D. and Spielmann, H. (2005), The EpiDerm Test Protocol for the Upcoming ECVAM Validation Study on *In Vitro* Skin Irritation Tests – An Assessment of the Performance of the Optimised Test, *ATLA* 33, 351-367.
- (14) Cotovio, J., Grandidier, M.-H., Portes, P., Roguet, R. and Rubinsteen, G. (2005). The *In Vitro* Acute Skin Irritation of Chemicals: Optimisation of the EPISKIN Prediction Model Within the Framework of the ECVAM Validation Process, *ATLA* 33, 329-349.
- (15) Zuang, V., Balls, M., Botham, P.A., Coquette, A., Corsini, E., Curren, R.D., Elliot, G.R., Fentem, J.H., Heylings, J.R., Liebsch, M., Medina, J., Roguet, R., van De Sandt, J.J.M., Wiemann, C. and Worth, A. (2002). Follow-Up to the ECVAM Prevalidation Study on *In Vitro* Tests for Acute Skin Irritation, The European Centre for the Validation of Alternative Methods Skin Irritation Task Force report 2, *ATLA* 30, 109-129.
- (16) Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J., Eskes, C., Roguet, R., Cotovio, J., Cole, T., Worth, A., Heylings, J., Jones, P., Robles, C., Kandárová, H., Gamer, A., Remmele, M., Curren, R., Raabe, H., Cockshott, A., Gerner, I. and Zuang, V. (2007). The ECVAM International Validation Study on *In Vitro* Tests for Acute Skin Irritation: Report on the Validity of the EPISKIN and EpiDerm Assays and on the Skin Integrity Function Test, *ATLA* 35, 559-601.
- (17) Hoffmann S. (2006). ECVAM Skin Irritation Validation Study Phase II: Analysis of the Primary Endpoint MTT and the Secondary Endpoint IL1- α .
- (18) Eskes C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I. and Zuang, V. (2007). The ECVAM International Validation Study on *In Vitro* Tests for Acute Skin Irritation: Selection of Test Chemicals, *ATLA* 35, 603-619.
- (19) Cotovio, J., Grandidier, M.-H., Lelièvre, D., Roguet, R., Tinois-Tessonnaud, E. and Leclaire, J. (2007). *In Vitro* Acute Skin Irritancy of Chemicals Using the Validated EPISKIN Model in a Tiered Strategy - Results and Performances with 184 Cosmetic Ingredients, *ALTEX*, 14, 351-358.
- (20) EURL-ECVAM (2007). Statement on the Validity of *In Vitro* Tests for Skin Irritation, Issued by the ECVAM Scientific Advisory Committee (ESAC26), 27 April 2007. Available at: https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/publication/ESAC26_statement_SkinIrritation_20070525_C.pdf
- (21) EURL-ECVAM. (2007). Performance Standards for Applying Human Skin Models to *In Vitro* Skin Irritation Testing. N.B. These are the original PS used for the validation of two test methods. These PS should not be used any longer as an updated version (8) is now available.
- (22) EURL-ECVAM. (2008). Statement on the Scientific Validity of *In Vitro* Tests for Skin Irritation Testing, Issued by the ECVAM Scientific Advisory Committee (ESAC29), 5 November 2008. https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/archive-publications/publication/ESAC_Statement_SkinEthic-EpiDerm-FINAL-0812-01.pdf

- (23) OECD (2010). Explanatory Background Document to the OECD Draft Test Guideline on *In Vitro* Skin Irritation Testing. Environment, Health and Safety Publications. Series on Testing and Assessment, (No 137), Organisation for Economic Cooperation and Development, Paris.
- (24) Katoh, M., Hamajima, F., Ogasawara, T. and Hata K. (2009). Assessment of Human Epidermal Model LabCyte EPI-MODEL for *In Vitro* Skin Irritation Testing According to European Centre for the Validation of Alternative Methods (ECVAM)-Validated Protocol, *J Toxicol Sci*, 34, 327-334
- (25) Katoh, M. and Hata K. (2011). Refinement of LabCyte EPI-MODEL24 Skin Irritation Test Method for Adaptation to the Requirements of OECD Test Guideline 439, *AATEX*, 16, 111-122
- (26) OECD (2011). Validation Report for the Skin Irritation Test Method Using LabCyte EPI-MODEL24. Environment, Health and Safety Publications, Series on Testing and Assessment (No 159), Organisation for Economic Cooperation and Development, Paris.
- (27) OECD (2011). Peer Review Report of Validation of the Skin Irritation Test Using LabCyte EPI-MODEL24. Environment, Health and Safety Publications, Series on Testing and Assessment (No 155), Organisation for Economic Cooperation and Development, Paris.
- (28) Kojima, H., Ando, Y., Idehara, K., Katoh, M., Kosaka, T., Miyaoka, E., Shinoda, S., Suzuki, T., Yamaguchi, Y., Yoshimura, I., Yuasa, A., Watanabe, Y. and Omori, T. (2012). Validation Study of the *In Vitro* Skin Irritation Test with the LabCyte EPI-MODEL24, *Altern Lab Anim*, 40, 33-50.
- (29) Welss, T., Basketter, D.A. and Schröder, K.R. (2004). *In Vitro* Skin Irritation: Fact and Future. State of the Art Review of Mechanisms and Models, *Toxicol. In Vitro* 18, 231-243.
- (30) Eskes, C. *et al.* (2012). Regulatory Assessment of *In Vitro* Skin Corrosion and Irritation Data within the European Framework: Workshop Recommendations. *Regul.Toxicol.Pharmacol.* 62, 393-403).
- (31) Mosmann, T. (1983). Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays, *J. Immunol. Methods* 65, 55-63.
- (32) EpiSkin™ (February 2009). SOP, Version 1, ECVAM Skin Irritation Validation Study: Validation of the EpiSkin™ Test Method 15 min - 42 hours for the Prediction of acute Skin Irritation of Chemicals
- (33) EpiDerm™ (Revised March 2009). SOP, Version 7.0, Protocol for: *In Vitro* EpiDerm™ Skin Irritation Test (EPI-200-SIT), for Use with MatTek Corporation's Reconstructed Human Epidermal Model EpiDerm (EPI-200).
- (34) SkinEthic™ RHE (February 2009) SOP, Version 2.0, SkinEthic Skin Irritation Test-42bis Test Method for the Prediction of Acute Skin Irritation of Chemicals: 42 Minutes Application + 42 Hours Post-Incubation.
- (35) LabCyte (June 2011). EPI-MODEL24 SIT SOP, Version 8.3, Skin Irritation Test Using the Reconstructed Human Model "LabCyte EPI-MODEL24"
- (36) Alépée, N., Barroso, J., De Smedt, A., De Wever, B., Hibatallah, J., Klaric, M., Mewes, K.R., Millet, M., Pfannenbecker, U., Tailhardat, M., Templier, M., and McNamee, P. Use of HPLC/UPLC-Spectrophotometry for Detection of MTT Formazan in *In Vitro* Reconstructed Human Tissue (RhT)-Based Test Methods Employing the MTT Assay to Expand their Applicability to Strongly Coloured Test Chemicals. Manuscript in preparation.
- (37) US FDA (2001). Guidance for Industry: Bioanalytical Method Validation. U.S. Department of Health and Human Services, Food and Drug Administration. May 2001. Available at: [<http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf>].

-
- (38) Harvell, J.D., Lamminstausta, K., and Maibach, H.I. (1995). Irritant Contact Dermatitis, in: Practical Contact Dermatitis, pp 7-18, (Ed. Guin J. D.). Mc Graw-Hill, New York.
- (39) EURL-ECVAM (2009). Performance Standards for *In Vitro* Skin Irritation Test Methods Based on Reconstructed Human Epidermis (RhE). *N.B. This is the current version of the ECVAM PS, updated in 2009 in view of the implementation of UN GHS. These PS should not be used any longer as an updated version (8) is now available related to the present TG.*
- (40) EURL-ECVAM. (2009). ESAC Statement on the Performance Standards (PS) for *In Vitro* Skin Irritation Testing Using Reconstructed Human Epidermis, Issued by the ECVAM Scientific Advisory Committee (ESAC31), 8 July 2009.
- (41) EC (2001). Commission Directive 2001/59/EC of 6 August 2001 Adapting to Technical Progress for the 28th Time Council Directive 67/548/EEC on the Approximation of Laws, Regulations and Administrative Provisions Relating to the Classification, Packaging and Labelling of Dangerous Substances, *Official Journal of the European Union* L225, 1-333.

Appendix 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with “concordance” to mean the proportion of correct outcomes of a test method (9).

Cell viability: Parameter measuring total activity of a cell population *e.g.* as ability of cellular mitochondrial dehydrogenases to reduce the vital dye MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue), which depending on the endpoint measured and the test design used, correlates with the total number and/or vitality of living cells.

Chemical: means a substance or a mixture.

Concordance: This is a measure of performance for test models that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (9).

ET₅₀: Can be estimated by determination of the exposure time required to reduce cell viability by 50 % upon application of the benchmark chemical at a specified, fixed concentration, see also IC₅₀.

GHS (Globally Harmonized System of Classification and Labelling of Chemicals by the United Nations (UN)): A system proposing the classification of chemicals (substances and mixtures) according to standardised types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

HPLC: High Performance Liquid Chromatography.

IATA: Integrated Approach on Testing and Assessment

IC₅₀: Can be estimated by determination of the concentration at which a benchmark chemical reduces the viability of the tissues by 50 % (IC₅₀) after a fixed exposure time, see also ET₅₀.

Infinite dose: Amount of test chemical applied to the *epidermis* exceeding the amount required to completely and uniformly cover the *epidermis* surface.

Mixture: A mixture or a solution composed of two or more substances.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

MTT: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide.

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration $\geq 10\%$ (w/w) and $< 80\%$ (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

NSC_{killed}: Non-Specific Colour in killed tissues.

NSC_{living}: Non-Specific Colour in living tissues.

NSMTT: Non-Specific MTT reduction.

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (9).

PC: Positive Control, a replicate containing all components of a test system and treated with a chemical known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (9).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility (9).

Replacement test: A test which is designed to substitute for a test that is in routine use and accepted for hazard identification and/or risk assessment, and which has been determined to provide equivalent or improved protection of human or animal health or the environment, as applicable, compared to the accepted test, for all possible testing situations and chemicals (9).

Run: A run consists of one or more test chemicals tested concurrently with a negative control and with a PC.

Sensitivity: The proportion of all positive/active test chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (9).

Skin irritation *in vivo*: The production of reversible damage to the skin following the application of a test chemical for up to 4 hours. Skin irritation is a locally arising reaction of the affected skin tissue and appears shortly after stimulation (38). It is caused by a local inflammatory reaction involving the innate (non-specific) immune system of the skin tissue. Its main characteristic is its reversible process involving inflammatory reactions and most of the clinical characteristic signs of irritation (erythema, oedema, itching and pain) related to an inflammatory process.

Specificity: The proportion of all negative/inactive test chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (9).

Substance: A chemical element and its compounds in the natural state or obtained by any production process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Test chemical: Any substance or mixture tested using this test method.

UPLC: Ultra-High Performance Liquid Chromatography.

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

Appendix 2

TEST MODELS INCLUDED IN THIS TEST METHOD

Nr.	Test model name	Validation study type	References
1	EpiSkin™	Full prospective validation study (2003-2007). The components of this model were used to define the essential test method components of the original and updated ECVAM PS (39) (40) (21) (*). Moreover, the method's data relating to identification of non-classified vs classified substances formed the main basis for defining the specificity and sensitivity values of the original PS (*).	(2) (10) (11) (14) (15) (16) (17) (18) (19) (20) (21) (23) (32) (39) (40)
2	EpiDerm™ SIT (EPI-200)	EpiDerm™ (<i>original</i>): Initially the test model underwent full prospective validation together with Nr. 1. from 2003-2007. The components of this model were used to define the essential test methods components of the original and updated ECVAM PS (39) (40) (21) (*). EpiDerm™ SIT (EPI-200) : A modification of the original EpiDerm™ was validated using the original ECVAM PS (21) in 2008 (*)	(2) (10) (12) (13) (15) (16) (17) (18) (20) (21) (23) (33) (39) (40) (2) (21) (22) (23) (33)
3	SkinEthic™ RHE	Validation study based on the original ECVAM Performance Standards (21) in 2008 (*).	(2) (21) (22) (23) (31)
4	LabCyte EPI-MODEL24 SIT	Validation study (2011-2012) based on the Performance Standards (PS) of OECD TG 439 (8) which are based on the updated ECVAM PS (*) (39) (40).	(24) (25) (26) (27) (28) (35) (39) (40) and PS of this TG (8) (*)

(*) The original ECVAM Performance Standards (PS) (21) were developed in 2007 upon completion of the prospective validation study (16) which had assessed the performance of test models Nr 1 and 2 in reference to the classification system as described in the 28th amendment to the EU Dangerous Substances Directive (41). In 2008 the UN GHS (1) and EU CLP were introduced, effectively shifting the cut-off value for distinguishing non-classified from classified substances from an *in vivo* score of 2.0 to 2.3. To adapt to this changed regulatory requirement, the accuracy values and reference chemical list of the ECVAM PS were updated in 2009 (2) (39) (40). As the original PS, also the updated PS were largely based on data from models Nr. 1 and 2 (16), but additionally used data on reference chemicals from model Nr. 3. In 2010, the updated ECVAM PS were used for stipulating the PS related to this TG (8). For the purpose of this test method, EpiSkin™ is considered the VRM, due to the fact that it was used to develop all the criteria of the PS. Detailed information on the validation studies, a compilation of the data generated as well as background to the necessary adaptations of the PS as a consequence of the UN GHS/CLP implementation can be found in the ECVAM/BfR explanatory background document to the corresponding OECD TG 439 (23).

SIT: Skin Irritation Test

RHE: Reconstructed Human Epidermis

Appendix 3

PROTOCOL PARAMETERS SPECIFIC TO EACH OF THE TEST MODELS INCLUDED IN THIS TEST METHOD

The RhE models do show very similar protocols and notably all use a post-incubation period of 42 hours (32) (33) (34) (35). Variations concern mainly three parameters relating to the different barrier functions of the test models and listed here: A) pre-incubation time and volume, B) Application of test chemicals and C) Post-incubation volume.

	EpiSkin™ (SM)	EpiDerm™ SIT (EPI-200)	SkinEthic RHE™	LabCyte EPI-MODEL24 SIT
--	---------------	------------------------	----------------	-------------------------

A) Pre-incubation

Incubation time	18-24 hours	18-24 hours	< 2 hours	15-30 hours
Medium volume	2ml	0,9ml	0,3 or 1ml	0,5ml

B) Test chemical application

For liquids	10µl (26µl/cm ²)	30µl (47µl/cm ²)	16µl (32µl/cm ²)	25µl (83µl/cm ²)
For solids	10mg (26mg/cm ²)+ DW (5µl)	25mg (39mg/cm ²)+ DPBS (25µl)	16mg (32mg/cm ²)+ DW (10µl)	25mg (83mg/cm ²)+ DW (25µl)
Use of nylon mesh	Not used	If necessary	Applied	Not used
Total application time	15 minutes	60 minutes	42 minutes	15 minutes
Application temperature	RT	a) at RT for 25 minutes b) at 37°C for 35 minutes	RT	RT

C) Post-incubation volume

Medium volume	2 ml	0,9ml x 2	2 ml	1 ml
---------------	------	-----------	------	------

D) Maximum acceptable variability

Standard deviation between tissue replicates	SD≤18	SD≤18	SD≤18	SD≤18
--	-------	-------	-------	-------

RT: Room temperature

DW: distilled water

DPBS: Dulbecco's Phosphate Buffer Saline

Appendix 4

KEY PARAMETERS AND ACCEPTANCE CRITERIA FOR QUALIFICATION OF AN HPLC/UPLC-SPECTROPHOTOMETRY SYSTEM FOR MEASUREMENT OF MTT FORMAZAN EXTRACTED FROM RHE TISSUES

Parameter	Protocol Derived from FDA Guidance (36) (37)	Acceptance Criteria
Selectivity	Analysis of isopropanol, living blank (isopropanol extract from living RhE tissues without any treatment), dead blank (isopropanol extract from killed RhE tissues without any treatment)	$\text{Area}_{\text{interference}} \leq 20 \% \text{ of Area}_{\text{LLOQ}}^{(1)}$
Precision	Quality Controls (i.e. MTT formazan at 1,6 µg/ml, 16 µg/ml and 160 µg/ml) in isopropanol (n=5)	$\text{CV} \leq 15 \% \text{ or } \leq 20 \% \text{ for the LLOQ}$
Accuracy	Quality Controls in isopropanol (n=5)	$\% \text{Dev} \leq 15 \% \text{ or } \leq 20 \% \text{ for LLOQ}$
Matrix Effect	Quality Controls in living blank (n=5)	$85 \% \leq \text{Matrix Effect } \% \leq 115 \%$
Carryover	Analysis of isopropanol after an ULOQ ⁽²⁾ standard	$\text{Area}_{\text{interference}} \leq 20 \% \text{ of Area}_{\text{LLOQ}}$
Reproducibility (intra-day)	3 independent calibration curves (based on 6 consecutive 1/3 dilutions of MTT formazan in isopropanol starting at ULOQ, i.e. 200 µg/ml); Quality Controls in isopropanol (n=5)	Calibration Curves: $\% \text{Dev} \leq 15 \% \text{ or } \leq 20 \% \text{ for LLOQ}$
Reproducibility (inter-day)	Day 1: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 2: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 3: 1 calibration curve and Quality Controls in isopropanol (n=3)	Quality Controls: $\% \text{Dev} \leq 15 \% \text{ and } \text{CV} \leq 15 \%$
Short Term Stability of MTT Formazan in RhE Tissue Extract	Quality Controls in living blank (n=3) analysed the day of the preparation and after 24 hours of storage at room temperature	$\% \text{Dev} \leq 15 \%$
Long Term Stability of MTT Formazan in RhE Tissue Extract, if required	Quality Controls in living blank (n=3) analysed the day of the preparation and after several days of storage at a specified temperature (e.g. 4°C, -20°C, -80°C)	$\% \text{Dev} \leq 15 \%$

⁽¹⁾ LLOQ: Lower Limit of Quantification, defined to cover 1-2 % tissue viability, i.e. 0,8 µg/ml.

⁽²⁾ ULOQ: Upper Limit of Quantification, defined to be at least two times higher than the highest expected MTT formazan concentration in isopropanol extracts from negative controls i.e. 200 µg/ml.

(8) In Part B, the following Chapters are added:

"B.63 REPRODUCTION/DEVELOPMENTAL TOXICITY SCREENING TEST

INTRODUCTION

1. This test method is equivalent to OECD test guideline (TG) 421 (2016). OECD guidelines for the testing of chemicals are periodically reviewed in the light of scientific progress. The original screening test guideline 421 was adopted in 1995, based on a protocol for a "Preliminary Reproduction Toxicity Screening Test" discussed in two expert meetings, in London in 1990 (1) and in Tokyo in 1992 (2).
2. This test method has been updated with endocrine disruptor relevant endpoints, as a follow up to the high-priority activity initiated at OECD in 1998 to revise existing test guidelines and to develop new test guidelines for the screening and testing of potential endocrine disruptors (3). OECD TG 407 (Repeated Dose 28-Day Oral Toxicity Study in Rodents, Chapter B.7 of this Annex) for example, was enhanced in 2008 by parameters suitable to detect endocrine activity of test chemicals. The objective in updating TG 421 was to include some endocrine disruptor relevant endpoints in screening TGs where the exposure periods cover some of the sensitive periods during development (pre- or early postnatal periods).
3. The selected additional endocrine disruptor relevant endpoints, also part of TG 443 (Extended One Generation Reproductive Toxicity Study, Chapter B.56 of this Annex), were included in TG 421 based on a feasibility study addressing scientific and technical questions related to their inclusion, as well as possible adaptations of the test design needed for their inclusion (4).
4. This test method is designed to generate limited information concerning the effects of a test chemical on male and female reproductive performance such as gonadal function, mating behaviour, conception, development of the conceptus and parturition. It is not an alternative to, nor does it replace the existing test methods B.31, B.34, B.35 or B.56.

INITIAL CONSIDERATIONS

5. This screening test method can be used to provide initial information on possible effects on reproduction and/or development, either at an early stage of assessing the toxicological properties of chemicals, or on chemicals of concern. It can also be used as part of a set of initial screening tests for existing chemicals for which little or no toxicological information is available, as a dose range finding study for more extensive reproduction/developmental studies, or when otherwise considered relevant. In conducting the study, the guiding principles and considerations outlined in the OECD guidance document no 19 on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluations (5) should be followed.
6. This test method does not provide complete information on all aspects of reproduction and development. In particular, it offers only limited means of detecting post-natal manifestations of pre-natal exposure, or effects that may be induced during post-natal exposure. Due (amongst other reasons) to the relatively small numbers of animals in the dose groups, the selectivity of the end points, and the short duration of the study, this method will not provide evidence for definite claims of no effects. Moreover, in the absence of data from other reproduction/developmental toxicity tests, positive results are useful for initial hazard assessment and contribute to decisions with respect to the necessity and timing of additional testing.
7. The results obtained by the endocrine related parameters should be seen in the context of the "OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals" (6). In this Conceptual Framework, the enhanced OECD TG 421 is contained in level 4 as an *in vivo* assay providing data on adverse effects on endocrine relevant endpoints. An endocrine signal might not however be considered sufficient evidence on its own that the test chemical is an endocrine disruptor.
8. This test method assumes oral administration of the test chemical. Modifications may be required if other routes of exposure are used.

9. Before use of the test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture.
10. Definitions used are given in Appendix 1.

PRINCIPLE OF THE TEST

11. The test chemical is administered in graduated doses to several groups of males and females. Males should be dosed for a minimum of four weeks and up to and including the day before scheduled kill (this includes a minimum of two weeks prior to mating, during the mating period and, approximately, two weeks post-mating). In view of the limited pre-mating dosing period in males, fertility may not be a particular sensitive indicator of testicular toxicity. Therefore, a detailed histological examination of the testes is essential. The combination of a pre-mating dosing period of two weeks and subsequent mating/fertility observations with an overall dosing period of at least four weeks, followed by detailed histopathology of the male gonads, is considered sufficient to enable detection of the majority of effects on male fertility and spermatogenesis.
12. Females should be dosed throughout the study. This includes two weeks prior to mating (with the objective of covering at least two complete oestrous cycles), the variable time to conception, the duration of pregnancy and at least thirteen days after delivery, up to and including the day before scheduled kill.
13. Duration of study, following acclimatisation and pre-dosing oestrous cycle evaluation, is dependent on the female performance and is approximately 63 days, [at least 14 days pre-mating, (up to) 14 days mating, 22 days gestation, 13 days lactation].
14. During the period of administration, the animals are observed closely each day for signs of toxicity. Animals which die or are killed during the test period are necropsied and, at the conclusion of the test, surviving animals are killed and necropsied.

DESCRIPTION OF THE METHOD

Selection of animal species

15. This test method is designed for use with the rat. If the parameters specified within this test method are investigated in another rodent species a detailed justification should be given. In the international validation program for the detection of endocrine disrupters in OECD TG 407 (corresponding to Chapter B.7 of this Annex), the rat was the only species used. Strains with low fecundity or well-known high incidence of developmental defects should not be used. Healthy virgin animals, not subjected to previous experimental procedures, should be used. The test animals should be characterised as to species, strain, sex, weight and age. At the commencement of the study the weight variation of animals used should be minimal and not exceed 20 % of the mean weight of each sex. Where the study is conducted as a preliminary study to a long-term or a full-generation study, it is preferable that animals from the same strain and source are used in both studies.

Housing and feeding

16. All procedures should conform to local standards of laboratory animal care. The temperature in the experimental animal room should be 22 °C (± 3 °). Although the relative humidity should be at least 30 % and preferably not exceed 70 % other than during room cleaning, the aim should be 50-60 %. Lighting should be artificial, the photoperiod being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. The choice of diet may be influenced by the need to ensure a suitable admixture of a test chemical when administered by this method.

17. Animals should be group housed in small groups of the same sex; animals may be housed individually if scientifically justified. For group caging, no more than five animals should be housed per cage. Mating procedures should be carried out in cages suitable for the purpose. Pregnant females should be caged individually and provided with nesting materials. Lactating females will be caged individually with their offspring.
18. The feed should be regularly analysed for contaminants. A sample of the diet should be retained until finalisation of the report.

Preparation of the animals

19. Healthy young adult animals are randomly assigned to the control and treatment groups. Cages should be arranged in such a way that possible effects due to cage placement are minimised. The animals are uniquely identified and kept in their cages for at least five days prior to the start of the study to allow for acclimatisation to the laboratory conditions.

Preparation of doses

20. It is recommended that the test chemical be administered orally unless other routes of administration are considered more appropriate. When the oral route is selected, the test chemical is usually administered by gavage; however, alternatively, test chemicals may be administered via the diet or drinking water.
21. Where necessary, the test chemical is dissolved or suspended in a suitable vehicle. It is recommended that, wherever possible, the use of an aqueous solution/suspension be considered first, followed by consideration of a solution/emulsion in oil (e.g. corn oil) and then by possible solution in other vehicles. For vehicles other than water the toxic characteristics of the vehicle should be known. The stability and homogeneity of the test chemical in the vehicle should be determined.

PROCEDURE

Number and sex of animals

22. It is recommended that each group be started with at least 10 males and 12-13 females. Females will be evaluated pre-exposure for oestrous cyclicity and animals that fail to exhibit typical 4-5 day cycles will not be included in the study; therefore, extra females are recommended in order to yield 10 females per group. Except in the case of marked toxic effects, it is expected that this will provide at least 8 pregnant females per group which normally is the minimum acceptable number of pregnant females per group. The objective is to produce enough pregnancies and offspring to assure a meaningful evaluation of the potential of the test chemical to affect fertility, pregnancy, maternal and suckling behaviour, and growth and development of the F₁ offspring from conception to day 13 post-partum.

Dosage

23. Generally, at least three test groups and a control group should be used. Dose levels may be based on information from acute toxicity tests or on results from repeated dose studies. Except for treatment with the test chemical, animals in the control group should be handled in an identical manner to the test group subjects. If a vehicle is used in administering the test chemical, the control group should receive the vehicle in the highest volume used.
24. Dose levels should be selected taking into account any existing toxicity and (toxico-) kinetic data available. It should also be taken into account that there may be differences in sensitivity between pregnant and non-pregnant animals. The highest dose level should be chosen with the aim of inducing toxic effects but not death or severe suffering. Thereafter, a descending sequence of dose levels should be selected with a view to demonstrating any dosage related response and no-observed-adverse effects (NOAEL) at the lowest dose level. Two to four fold intervals are frequently optimal for setting the descending dose levels and addition of a fourth test group is often preferable to using very large intervals (e.g. more than a factor of 10) between dosages.

25. In the presence of observed general toxicity (e.g. reduced body weight, liver, heart, lung or kidney effects, etc.) or other changes that may not be toxic responses (e.g. reduced food intake, liver enlargement), observed effects on endocrine sensitive endpoints should be interpreted with caution.

Limit test

26. If an oral study at one dose level of at least 1 000 mg/kg body weight/day or, for dietary or drinking water administration, an equivalent percentage in the diet or drinking water, using the procedures described for this study, produces no observable toxic effects and if toxicity would not be expected based upon data from structurally related substances, then a full study using several dose levels may not be considered necessary. The limit test applies except when human exposure indicates the need for a higher oral dose level to be used. For other types of administration, such as inhalation or dermal application, the physical chemical properties of the test chemicals often may dictate the maximum attainable concentration.

Administration of doses

27. The animals are dosed with the test chemical daily for 7 days a week. When the test chemical is administered by gavage, this should be done in a single dose to the animals using a stomach tube or a suitable intubation cannula. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. The volume should not exceed 1 ml/100 g body weight, except in the case of aqueous solutions where 2 ml/100 g body weight may be used. Except for irritating or corrosive test chemicals which will normally reveal exacerbated effects with higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure a constant volume at all dose levels.
28. For test chemical administered via the diet or drinking water, it is important to ensure that the quantities of the test chemical involved do not interfere with normal nutrition or water balance. When the test chemical is administered in the diet either a constant dietary concentration (ppm) or a constant dose level in terms of the animals' body weight may be used; the alternative used should be specified. For a test chemical administered by gavage, the dose should be given at similar times each day, and adjusted at least weekly to maintain a constant dose level in terms of animal body weight.

Experimental schedule

29. Dosing of both sexes should begin at least 2 weeks prior to mating, after they have been acclimatised for at least five days and females have been screened for normal oestrous cycles (in a 2 weeks pre-treatment period). The study should be scheduled in such a way that oestrous cycle evaluation begins soon after the animals have attained full sexual maturity. This may vary slightly for different strains of rats in different laboratories, e.g. Sprague Dawley rats 10 weeks of age, Wistar rats about 12 weeks of age. Dams with offspring should be killed on day 13 post-partum, or shortly thereafter. The day of birth (viz. when parturition is complete) is defined as day 0 post-partum. Females showing no-evidence of copulation are killed 24-26 days after the last day of the mating period. Dosing is continued in both sexes during the mating period. Males should further be dosed after the mating period at least until the minimum total dosing period of 28 days has been completed. They are then killed, or, alternatively, are retained and continued to be dosed for the possible conduction of a second mating if considered appropriate.
30. Daily dosing of the parental females should continue throughout pregnancy and at least up to, and including, day 13 post-partum or the day before sacrifice. For studies where the test chemical is administered by inhalation or by the dermal route, dosing should be continued at least up to, and including, day 19 of gestation, and dosing should be re-initiated as soon as possible and not later than PND 4.
31. A diagram of the experimental schedule is given in Appendix 2.

Mating procedure

32. Normally, 1:1 (one male to one female) matings should be used in this study. Exceptions can arise in the case of occasional deaths of males. The female should be placed with the same male until evidence of copulation is observed or two weeks have elapsed. Each morning the females should be examined for the presence of sperm or a vaginal plug. Day 0 of pregnancy is defined as the day on which mating evidence is confirmed (a vaginal plug or sperm is found). In case pairing is unsuccessful, re-mating of females with proven males of the same group could be considered.

Litter size

33. On day 4 after birth, the size of each litter may be adjusted by eliminating extra pups by random selection to yield, as nearly as possible, four or five pups per sex per litter depending on the normal litter size in the strain of rats used. Blood samples should be collected from two of the surplus pups, pooled, and used for determination of serum T4 levels. Selective elimination of pups, e.g. based upon body weight, or anogenital distance (AGD) is not appropriate. Whenever the number of male or female pups prevents having four or five of each sex per litter, partial adjustment (for example, six males and four females) is acceptable. No pups will be eliminated when litter size will drop below the culling target (8 or 10 pups/litter). If there is only one pup available above the culling target, only one pup will be eliminated and used for blood collection for possible serum T4 assessments.
34. If litter size is not adjusted, two pups per litter are sacrificed on day 4 after birth and blood samples are taken for measurement of serum thyroid hormone concentrations. If possible the two pups per litter should be female pups to reserve male pups for nipple retention evaluations except in the event that removing these pups leaves no remaining females for assessment at termination. No pups will be eliminated when litter size will drop below 8 or 10 pups/litter (depending on the normal litter size in the strain of rats used). If there is only one pup available above the normal litter size, only one pup will be eliminated and used for blood collection for possible serum T4 assessments.

In life observations

Clinical observations

35. Throughout the test period, general clinical observations should be made at least once a day, and more frequently when signs of toxicity are observed. They should be made preferably at the same time(s) each day, considering the peak period of anticipated effects after dosing. Pertinent behavioural changes, signs of difficult or prolonged parturition and all signs of toxicity, including mortality, should be recorded. These records should include time of onset, degree and duration of toxicity signs.

Body weight and food/water consumption

36. Males and females should be weighed on the first day of dosing, at least weekly thereafter, and at termination. During pregnancy, females should be weighed on days 0, 7, 14 and 20 and within 24 hours of parturition (day 0 or 1 post-partum) and at least day 4 and 13 post-partum. These observations should be reported individually for each adult animal.
37. During pre-mating, pregnancy and lactation, food consumption should be measured at least weekly. The measurement of food consumption during mating is optional. Water consumption during these periods should also be measured when the test chemical is administered via drinking water.

Oestrous cycles

38. Oestrous cycles should be monitored before treatment starts to select for the study females with regular cyclicity (see paragraph 22). Vaginal smears should also be monitored daily from the beginning of the treatment period until evidence of mating. If there is concern about acute stress effects that could alter oestrous cycles with the initiation of dosing, laboratories can expose test animals for 2 weeks, then collect vaginal smears daily to monitor oestrous cycle for a minimum of two weeks during the pre-mating period with continued monitoring into the mating period until there is evidence of mating. When obtaining vaginal/cervical cells, care should be taken to avoid disturbance of mucosa, which could induce pseudopregnancy (7) (8).

Offspring parameters

39. The duration of gestation should be recorded and is calculated from day 0 of pregnancy. Each litter should be examined as soon as possible after delivery to establish the number and sex of pups, stillbirths, live births, runts (pups that are significantly smaller than corresponding control pups) and the presence of gross abnormalities.

40. Live pups should be counted and sexed and litters weighed within 24 hours of parturition (day 0 or 1 post-partum) and at least on day 4 and 13 post-partum. In addition to the observations described in paragraph 35, any abnormal behaviour of the offspring should be recorded.
41. The AGD of each pup should be measured on the same postnatal day between PND 0 through PND 4. Pup body weight should be collected on the day the AGD is measured and the AGD should be normalised to a measure of pup size, preferably the cube root of body weight (9). The number of nipples/areolae in male pups should be counted on PND 12 or 13 as recommended in OECD GD 151 (10).

Clinical biochemistry

42. Blood samples from a defined site are taken based on the following schedule:
 - from at least two pups per litter on day 4 after birth, if the number of pups allows (see paragraphs 33-34)
 - from all dams and at least two pups per litter at termination on day 13, and
 - from all adult males, at termination,

All blood samples are stored under appropriate conditions. Blood samples from the day 13 pups and the adult males are assessed for serum levels for thyroid hormones (T4). Further assessment of T4 in blood samples from the dams and day 4 pups is done if relevant. As an option other hormones may be measured if relevant. Pup blood can be pooled by litter for thyroid hormone analyses. Thyroid hormones (T4 and TSH) should preferably be measured as 'total'.

43. The following factors may influence the variability and the absolute concentrations of the hormone determinations:
 - time of sacrifice because of diurnal variation of hormone concentrations
 - method of sacrifice to avoid undue stress to the animals that may affect hormone concentrations
 - test kits for hormone determinations that may differ by their standard curves.
44. Plasma samples specifically intended for hormone determination should be obtained at a comparable time of the day. The numerical values obtained when analysing hormone concentrations differ with various commercial assay kits.

Pathology

Gross necropsy

45. At the time of sacrifice or death during the study, the adult animals should be examined macroscopically for any abnormalities or pathological changes. Special attention should be paid to the organs of the reproductive system. The number of implantation sites should be recorded. Vaginal smears should be examined in the morning on the day of necropsy to determine the stage of the oestrous cycle and allow correlation with histopathology of ovaries.
46. The testes and epididymides as well as prostate and seminal vesicles with coagulating glands as a whole, of all male adult animals should be trimmed of any adherent tissue, as appropriate, and their wet weight taken as soon as possible after dissection to avoid drying. In addition, optional organ weights could include levator ani plus bulbocavernosus muscle complex, Cowper's glands and glans penis in males and paired ovaries (wet weight) and uterus (including cervix) in females; if included, these weights should be collected as soon as possible after dissection.
47. Dead pups and pups killed at day 13 post-partum, or shortly thereafter, should, at least, be carefully examined externally for gross abnormalities. Particular attention should be paid to the external reproductive genitals which should be examined for signs of altered development. At day 13 the thyroid from 1 male and 1 female pup per litter should be preserved.

48. The ovaries, testes, accessory sex organs (uterus and cervix, epididymides, prostate, seminal vesicles plus coagulating glands), thyroid and all organs showing macroscopic lesions of all adult animals should be preserved. Formalin fixation is not recommended for routine examination of testes and epididymides. An acceptable method is the use of Bouin's fixative or modified Davidsons for these tissues (11). The tunica albuginea may be gently and shallowly punctured at the both poles of the organ with a needle to permit rapid penetration of the fixative.

Histopathology

49. Detailed histological examination should be performed on the ovaries, testes and epididymides (with special emphasis on stages of spermatogenesis and histopathology of interstitial testicular cell structure) of the animals of the highest dose group and the control group. The other preserved organs including thyroid from pups and adult animals may be examined when necessary. The thyroid weight could be determined after fixation. Trimming should also be done very carefully and only after fixation to avoid tissue damage. Tissue damage could compromise histopathology analysis. Examinations should be extended to the animals of other dosage groups when changes are seen in the highest dose group. The Guidance on histopathology (11) details extra information on dissection, fixation, sectioning and histopathology of endocrine tissues.

DATA AND REPORTING

Data

50. Individual animal data should be provided. Additionally, all data should be summarised in tabular form, showing for each test group the number of animals at the start of the test, the number of animals found dead during the test or killed for humane reasons, the time of any death or humane kill, the number of fertile animals, the number of pregnant females, the number of animals showing signs of toxicity, a description of the signs of toxicity observed, including time of onset, duration, and severity of any toxic effects, the types of histopathological changes, and all relevant litter data. A tabular summary report format that has proven to be very useful for the evaluation of reproductive/developmental effect is given in Appendix 3.
51. Due to the limited dimensions of the study, statistical analyses in the form of tests for "significance" are of limited value for many endpoints, especially reproductive endpoints. If statistical analyses are used then the method chosen should be appropriate for the distribution of the variable examined, and be selected prior to the start of the study. Statistical analysis of AGD and nipple retention should be based on individual pup data, taking litter effects into account. Where appropriate, the litter is the unit of analysis. Statistical analysis of pup body weight should be based on individual pup data, taking litter size into account. Because of the small group size, the use of historic control data (e.g. for litter size), where available, may also be useful as an aid to the interpretation of the study.

Evaluation of results

52. The findings of this toxicity study should be evaluated in terms of the observed effects, necropsy and microscopic findings. The evaluation will include the relationship between the dose of the test chemical and the presence or absence, incidence and severity of abnormalities, including gross lesions, identified target organs, infertility, clinical abnormalities, affected reproductive and litter performance, body weight changes, effects on mortality and any other toxic effects.
53. Because of the short period of treatment of the male, the histopathology of the testes and epididymides should be considered along with the fertility data, when assessing male reproductive effects. The use of historical control data on reproduction/development (e.g., for litter size, AGD, nipple retention, serum T4 levels), where available, may also be useful as an aid to the interpretation of the study.
54. For quality control it is proposed that historical control data are collected and that for numerical data coefficients of variation are calculated, especially for the parameters linked with endocrine disrupter detection. These data can be used for comparison purposes when actual studies are evaluated.

Test report

55. The test report should include the following information:

Test chemical:

- source, lot number, limit date for use, if available
- stability of the test chemical, if known.

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Vehicle (if appropriate):

- justification for choice of vehicle if other than water.

Test animals:

- species/strain used;
- number, age and sex of animals;
- source, housing conditions, diet, etc.;
- individual weights of animals at the start of the test.
- justification for species if not rat

Test conditions:

- rationale for dose level selection;
- details of test chemical formulation/diet preparation, achieved concentrations, stability and homogeneity of the preparation;
- details of the administration of the test chemical;
- conversion from diet/drinking water test chemical concentration (ppm) to the actual dose (mg/kg body weight/day), if applicable;

- details of food and water quality;
- detailed description of the randomisation procedure to select pups for culling, if culled.

Results:

- body weight/body weight changes;
- food consumption, and water consumption if available;
- toxic response data by sex and dose, including fertility, gestation, and any other signs of toxicity;
- gestation length;
- toxic or other effects on reproduction, offspring, post-natal growth, etc.;
- nature, severity and duration of clinical observations (whether reversible or not);
- number of adult females with normal or abnormal oestrous cycle and cycle duration;
- number of live births and post-implantation loss;
- pup body weight data
- AGD of all pups (and body weight on day of AGD measurement)
- nipple retention in male pups,
- thyroid hormone levels, day 13 pups and adult males (and dams and day 4 pups if measured)
- number pups with grossly visible abnormalities, gross evaluation of external genitalia, number of runts;
- time of death during the study or whether animals survived to termination;
- number of implantations, litter size and litter weights at the time of recording;
- body weight at sacrifice and organ weight data for the parental animals;
- necropsy findings;
- detailed description of histopathological findings;

- absorption data (if available);
- statistical treatment of results, where appropriate.

Discussion of results.

Conclusions.

Interpretation of results

56. The study will provide evaluations of reproduction/developmental toxicity associated with administration of repeated doses (see paragraphs 5 and 6). It could provide an indication of the need to conduct further investigations and provides guidance in the design of subsequent studies. OECD Guidance Document 43 should be consulted for aid in the interpretation of reproduction and developmental results (12). OECD Guidance Document No 106 on Histologic Evaluation of Endocrine and Reproductive Tests in Rodents (11) provides information on the preparation and evaluation of (endocrine) organs and vaginal smears that may be helpful for this TG.

LITERATURE

- (1) OECD (1990). Room Document No 1 for the 14th Joint Meeting of the Chemicals Group and Management Committee. Available upon request at Organisation for Economic and Cooperation and Development, Paris.
- (2) OECD (1992). Chairman's Report of the ad hoc Expert Meeting on Reproductive Toxicity Screening Methods, Tokyo, 27th-29th October, 1992. Available Upon Request at Organisation for Economic Cooperation and Development, Paris.
- (3) OECD (1998). Report of the First Meeting of the OECD Endocrine Disrupter Testing and Assessment (EDTA) Task Force, 10th-11th March 1998. Available Upon Request at Organisation for Economic Cooperation and Development, Paris.
- (4) OECD (2015). Feasibility Study for Minor Enhancements of TG 421/422 with ED Relevant Endpoints. Environment, Health and Safety Publications, Series on Testing and Assessment (No 217), Organisation for Economic Cooperation and Development, Paris.
- (5) OECD (2000). Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluations. Series on Testing and Assessment, (No 19), Organisation for Economic Cooperation and Development, Paris.
- (6) OECD (2011). Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Environment, Health and Safety Publications, Series on Testing and Assessment (No 150), Organisation for Economic Cooperation and Development, Paris.
- (7) Goldman, J.M., Murr A.S., Buckalew A.R., Ferrell J.M. and Cooper R.L. (2007). The Rodent Estrous Cycle: Characterization of Vaginal Cytology and its Utility in Toxicological Studies, Birth Defects Research, Part B, 80 (2), 84-97.
- (8) Sadleir R.M.F.S (1979). Cycles and Seasons, in Auston C.R. and Short R.V. (eds.), Reproduction in Mammals: I. Germ Cells and Fertilization, Cambridge, New York.
- (9) Gallavan R.H. Jr, Holson J.F., Stump D.G., Knapp J.F. and Reynolds V.L. (1999). Interpreting the Toxicologic Significance of Alterations in Anogenital Distance: Potential for Confounding Effects of Progeny Body Weights, Reproductive Toxicology, 13: 383-390.

-
- (10) OECD (2013). Guidance Document in Support of the Test Guideline on the Extended One Generation Reproductive Toxicity Study. Environment, Health and Safety Publications, Series on Testing and Assessment (No 151), Organisation for Economic Cooperation and Development, Paris.
 - (11) OECD (2009). Guidance Document for Histologic Evaluation of Endocrine and Reproductive Tests in Rodents. Environment, Health and Safety Publications, Series on Testing and Assessment (No106), Organisation for Economic Cooperation and Development, Paris.
 - (12) OECD (2008). Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment. Environment, Health and Safety Publications, Series on Testing and Assessment (No 43), Organisation for Economic Cooperation and Development, Paris.

Appendix 1

DEFINITIONS (SEE ALSO OECD GD 150 (6))

Androgenicity is the capability of a chemical to act like a natural androgenic hormone (e.g. testosterone) in a mammalian organism.

Antiandrogenicity is the capability of a chemical to suppress the action of a natural androgenic hormone (e.g. testosterone) in a mammalian organism.

Antioestrogenicity is the capability of a chemical to suppress the action of a natural oestrogenic hormone (e.g. oestradiol 17 β) in a mammalian organism.

Antithyroid activity is the capability of a chemical to suppress the action of a natural thyroid hormone (e.g. T₃) in a mammalian organism.

Chemical is a substance or a mixture.

Developmental toxicity: the manifestation of reproductive toxicity, representing pre-, peri- post-natal, structural, or functional disorders in the progeny.

Dosage is a general term comprising of dose, its frequency and the duration of dosing.

Dose is the amount of test chemical administered. The dose is expressed as weight of test chemical per unit body weight of test animal per day (e.g. mg/kg body weight/day), or as a constant dietary concentration.

Evident toxicity is a general term describing clear signs of toxicity following administration of test chemical. These should be sufficient for hazard assessment and should be such that an increase in the dose administered can be expected to result in the development of severe toxic signs and probable mortality.

Impairment of fertility represents disorders of male or female reproductive functions or capacity.

Maternal toxicity: adverse effects on gravid females, occurring either specifically (direct effect) or not specifically (indirect effect).

NOAEL is the abbreviation for no-observed-adverse effect level. This is the highest dose level where no adverse treatment-related findings are observed due to treatment.

Oestrogenicity is the capability of a chemical to act like a natural oestrogenic hormone (e.g. oestradiol 17 β) in a mammalian organism.

Reproduction toxicity represents harmful effects on the progeny and/or an impairment of male and female reproductive functions or capacity.

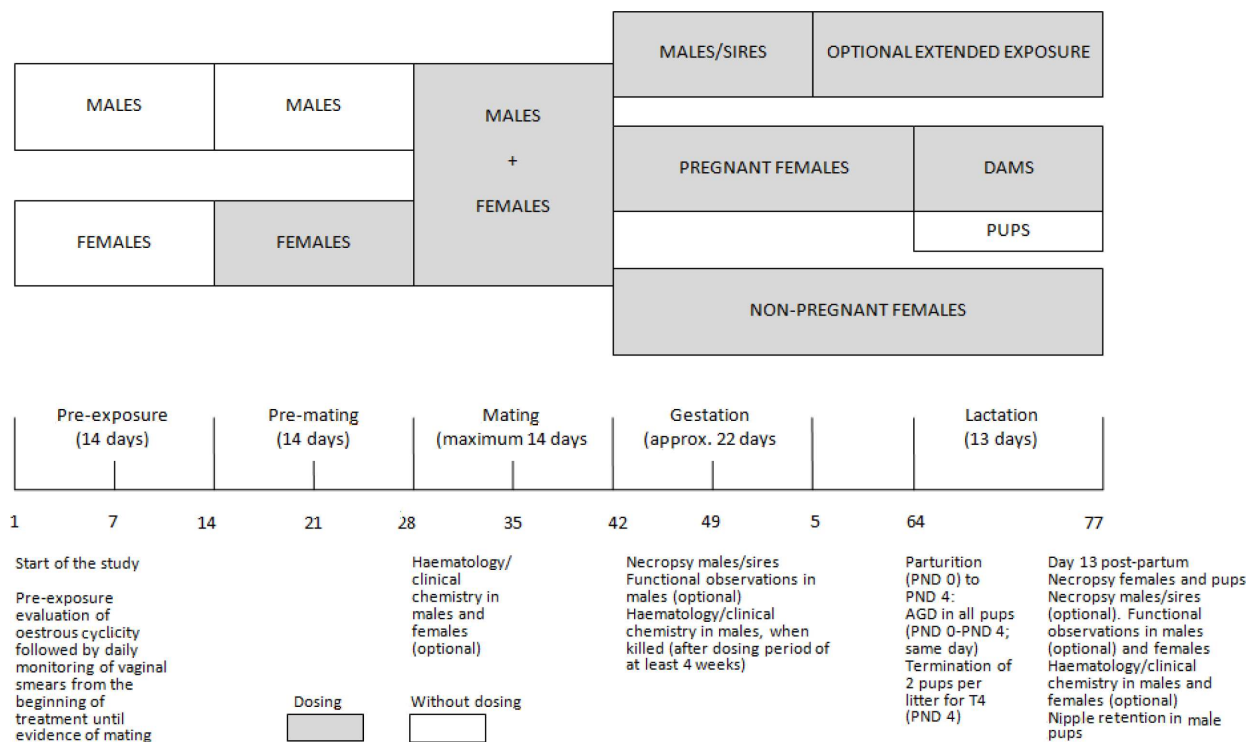
Test chemical is any substance or mixture tested using this test method.

Thyroid activity is the capability of a chemical to act like a natural thyroid hormone (e.g. T₃) in a mammalian organism.

Validation is a scientific process designed to characterise the operational requirements and limitations of a test method and to demonstrate its reliability and relevance for a particular purpose.

Appendix 2

DIAGRAM OF THE EXPERIMENTAL SCHEDULE INDICATING THE MAXIMUM STUDY DURATION, BASED ON A FULL 14-DAY MATING PERIOD



Appendix 3

TABULAR SUMMARY REPORT OF EFFECTS ON REPRODUCTION/DEVELOPMENT

OBSERVATIONS	VALUES				
	0 (control)
Dosage (units)					
Pairs started (N)					
Oestrus cycle (at least mean length and frequency of irregular cycles)					
Females showing evidence of copulation (N)					
Females achieving pregnancy (N)					
Conceiving days 1 - 5 (N)					
Conceiving days 6 -... ⁽¹⁾ (N)					
Pregnancy 21 days (N)					
Pregnancy = 22 days (N)					
Pregnancy 23 days (N)					
Dams with live young born (N)					
Dams with live young at day 4 pp (N)					
Implants/dam (mean)					
Live pups/dam at birth (mean)					
Live pups/dam at day 4 (mean)					
Sex ratio (m/f) at birth (mean)					
Sex ratio (m/f) at day 4 (mean)					
Litter weight at birth (mean)					
Litter weight at day 4 (mean)					
Pup weight at birth (mean)					
Pup weight at the time of AGD measurement (mean males, mean females)					

OBSERVATIONS	VALUES				
Dosage (units)	0 (control)
Pup AGD on the same postnatal day, birth – day 4 (mean males, mean females, note PND)					
Pup weight at day 4 (mean)					
Male pup nipple retention at day 13 (mean)					
Pup weight at day 13 (mean)					
ABNORMAL PUPS					
Dams with 0					
Dams with 1					
Dams with 2					
LOSS OF OFFSPRING					
Pre-natal/post-implantations (implantations minus live births)					
Females with 0					
Females with 1					
Females with 2					
Females with 3					
Post-natal (live births minus alive at post-natal day 13)					
Females with 0					
Females with 1					
Females with 2					
Females with 3					

(¹) last day of the mating period

B.64 COMBINED REPEATED DOSE TOXICITY STUDY WITH THE REPRODUCTION/DEVELOPMENTAL TOXICITY SCREENING TEST

INTRODUCTION

1. This test method is equivalent to OECD test guideline (TG) 422 (2016). OECD guidelines for the Testing of Chemicals are periodically reviewed in the light of scientific progress. The original screening test guideline 422 was adopted in 1996, based on a protocol for a "Combined Repeat Dose and Reproductive/Developmental Screening Test" discussed in two expert meetings, in London in 1990 (1) and in Tokyo in 1992 (2).
2. This test method combines a reproduction/developmental toxicity screening part which is based on experience gained in Member countries from using the original method on existing high production volume chemicals and in exploratory tests with positive control substances (3) (4), and a repeated dose toxicity part, in concordance with OECD test guideline 407 (Repeated Dose 28-Day Oral Toxicity Study in Rodents, corresponding to Chapter B.7 of this Annex).
3. This test method has been updated with endocrine disruptor relevant endpoints, as a follow up to the high-priority activity initiated at OECD in 1998 to revise existing test guidelines and to develop new test guidelines for the screening and testing of potential endocrine disruptors (5). In this context TG 407 (corresponding to Chapter B.7 of this Annex) was enhanced in 2008 by parameters suitable to detect endocrine activity of test chemicals. The objective in updating TG 422 was to include some endocrine disruptor relevant endpoints in screening TGs where the exposure periods cover some of the sensitive periods during development (pre- or early postnatal periods).
4. The selected additional endocrine disruptor relevant endpoints, also part of TG 443 (Extended One Generation Reproductive Toxicity Study, corresponding to Chapter B.56 of this Annex), were included in TG 422 based on a feasibility study addressing scientific and technical questions related to their inclusion, as well as possible adaptations of the test design needed for their inclusion (6).
5. This test method is designed to generate limited information concerning the effects of a test chemical on male and female reproductive performance such as gonadal function, mating behaviour, conception, development of the conceptus and parturition. It is not an alternative to, nor does it replace the existing test methods B.31, B.34, B.35 or B.56.

INITIAL CONSIDERATIONS

6. In the assessment and evaluation of the toxic characteristics of a test chemical the determination of oral toxicity using repeated doses may be carried out after the initial information on toxicity has been obtained by acute testing. This study provides information on the possible health hazards likely to arise from repeated exposure over a relatively limited period of time. The method comprises the basic repeated dose toxicity study that may be used for chemicals on which a 90-day study is not warranted (e.g. when the production volume does not exceed certain limits) or as a preliminary study to a long-term study. In conducting the study, the guiding principles and considerations outlined in the OECD guidance document n° 19 on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluations (7) should be followed.
7. It further comprises a reproduction/developmental toxicity screening test and, therefore, can also be used to provide initial information on possible effects on male and female reproductive performance such as gonadal function, mating behaviour, conception, development of the conceptus and parturition, either at an early stage of assessing the toxicological properties of test chemicals, or on test chemicals of concern. This test method does not provide complete information on all aspects of reproduction and development. In particular, it offers only limited means of detecting postnatal manifestations of prenatal exposure, or effects that may be induced during postnatal exposure. Due (amongst other reasons) to the selectivity of the end points, and the short duration of the study, this method will not provide evidence for definite claims of no reproduction/developmental effects. Moreover, in the absence of data from other reproduction/developmental toxicity tests, positive results are useful for initial hazard assessment and contribute to decisions with respect to the necessity and timing of additional testing.

8. The results obtained by the endocrine related parameters should be seen in the context of the “OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals” (8). In this Conceptual Framework, the enhanced OECD TG 422 is contained in level 4 as an *in vivo* assay providing data on adverse effects on endocrine relevant endpoints. An endocrine signal might not however be considered sufficient evidence on its own that the test chemical is an endocrine disruptor.
9. The test method also places emphasis on neurological effects as a specific endpoint, and the need for careful clinical observations of the animals, so as to obtain as much information as possible, is stressed. The method should identify chemicals with neurotoxic potential, and which may warrant further in-depth investigation of this aspect. In addition, the method may also give a basic indication of immunological effects.
10. In the absence of data from other systemic toxicity, reproduction/developmental toxicity, neurotoxicity and/or immunotoxicity studies, positive results are useful for initial hazard assessment and contribute to decisions with respect to the necessity and timing of additional testing. The test may be particularly useful as part of the OECD Screening Information Data Set (SIDS) for the assessment of existing chemicals for which little or no toxicological information is available and can serve as an alternative to conducting two separate tests for repeated dose toxicity (OCD TG 407, corresponding to Chapter B.7 of this Annex) and reproduction/developmental toxicity (OECD TG 421, corresponding to Chapter B.6.3 of this Annex), respectively. It can also be used as a dose range finding study for more extensive reproduction/developmental studies, or when otherwise considered relevant.
11. Generally, it is assumed that there are differences in sensitivity between pregnant and non-pregnant animals. Consequently, it may be more complicated to determine dose levels in this combined test that are adequate to evaluate both general systemic toxicity and specific reproduction/developmental toxicity, rather than when the individual tests are conducted separately. Moreover, interpretation of the test results with respect to general systemic toxicity may be more difficult than when conducting a separate repeated-dose study, especially when serum and histopathology parameters are not evaluated at the same time in the study. Because of these technical complexities, considerable experience in toxicity testing is required for the performance of this combined screening test. On the other hand, apart from the smaller number of animals involved, the combined test may offer a better means of discriminating direct effects on reproduction/development from those that are secondary to other (systemic) effects.
12. In this test, the dosing period is longer than in a conventional 28-day repeated dose study. However, it uses fewer animals of each sex per group when compared with the situation where a conventional 28-day repeated dose study is conducted in addition to a Reproduction/Developmental Toxicity Screening Test.
13. This test method assumes oral administration of the test chemical. Modifications may be required if other routes of exposure are used.
14. Before use of the test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture.
15. Definitions used are given in Appendix 1.

PRINCIPLE OF THE TEST

16. The test chemical is administered in graduated doses to several groups of males and females. Males should be dosed for a minimum of four weeks, up to and including the day before scheduled kill (this includes a minimum of two weeks prior to mating, during the mating period and, approximately, two weeks post mating). In view of the limited pre-mating dosing period in males, fertility may not be a particularly sensitive indicator of testicular toxicity. Therefore, a detailed histological examination of the testes is essential. The combination of a pre-mating dosing

period of two weeks and subsequent mating/fertility observations with an overall dosing period of at least four weeks, followed by detailed histopathology of the male gonads, is considered sufficient to enable detection of the majority of effects on male fertility and spermatogenesis.

17. Females should be dosed throughout the study. This includes two weeks prior to mating (with the objective of covering at least two complete oestrous cycles), the variable time to conception, the duration of pregnancy and at least thirteen days after delivery, up to and including the day before scheduled kill.
18. Duration of study, following acclimatisation and pre-dosing oestrous cycle evaluation, is dependent on the female performance and is approximately 63 days, [at least 14 days pre-mating, (up to) 14 days mating, 22 days gestation, 13 days lactation].
19. During the period of administration, the animals are observed closely each day for signs of toxicity. Animals which die or are killed during the test are necropsied and, at the conclusion of the test, surviving animals are killed and necropsied.

DESCRIPTION OF THE METHOD

Selection of animal species

20. This test method is designed for use with the rat. If the parameters specified within this TG 422 are investigated in another rodent species a detailed justification should be given. In the international validation program for the detection of endocrine disruptors on TG 407, the rat was the only species used. Strains with low fecundity or well-known high incidence of developmental defects should not be used. Healthy virgin animals, not subjected to previous experimental procedures, should be used. The test animals should be characterised as to species, strain, sex, weight and age. At the commencement of the study the weight variation of animals used should be minimal and not exceed $\pm 20\%$ of the mean weight of each sex. Where the study is conducted as a preliminary study to a long-term or a full-generation study, it is preferable that animals from the same strain and source are used in both studies.

Housing and feeding

21. All procedures should conform to local standards of laboratory animal care. The temperature in the experimental animal room should be $22\text{ }^{\circ}\text{C}$ ($\pm 3\text{ }^{\circ}$). The relative humidity should be at least 30 % and preferably not exceed 70 % other than during room cleaning. Lighting should be artificial, the photoperiod being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water. The choice of diet may be influenced by the need to ensure a suitable admixture of a test chemical when administered by this method.
22. Animals should be group housed in small groups of the same sex; animals may be housed individually if scientifically justified. For group caging, no more than five animals should be housed per cage. Mating procedures should be carried out in cages suitable for the purpose. Pregnant females should be caged individually and provided with nesting materials. Lactating females will be caged individually with their offspring.
23. The feed should be regularly analysed for contaminants. A sample of the diet should be retained until finalisation of the report.

Preparation of the animals

24. Healthy young adult animals are randomised and assigned to the treatment groups and cages. Cages should be arranged in such a way that possible effects due to cage placements are minimised. The animals are uniquely identified and kept in their cages for at least five days prior to the start of the study to allow for acclimatisation to the laboratory conditions.

Preparation of doses

25. It is recommended that the test chemical be administered orally unless other routes of administration are considered more appropriate. When the oral route is selected, the test chemical is usually administered by gavage; however, alternatively, test chemicals may also be administered via the diet or drinking water.

26. Where necessary, the test chemical is dissolved or suspended in a suitable vehicle. It is recommended that, wherever possible, the use of an aqueous solution/suspension be considered first, followed by consideration of a solution/suspension in oil (e.g. corn oil) and then by possible solution in other vehicles. For non-aqueous vehicles the toxic characteristics of the vehicle should be known. The stability and homogeneity of the test chemical in the vehicle should be determined.

PROCEDURE

Number and sex of animals

27. It is recommended that each group be started with at least 10 males and 12-13 females. Females will be evaluated pre-exposure for oestrous cyclicity and animals that fail to exhibit typical 4-5 day cycles will not be included in the study; therefore, extra females are recommended in order to yield 10 females per group. Except in the case of marked toxic effects, it is expected that this will provide at least 8 pregnant females per group which normally is the minimum acceptable number of pregnant females per group. The objective is to produce enough pregnancies and offspring to assure a meaningful evaluation of the potential of the test chemical to affect fertility, pregnancy, maternal and suckling behaviour, and growth and development of the F₁ offspring from conception to day 13 post-partum. If interim kills are planned, the number should be increased by the number of animals scheduled to be killed before the completion of the study. Consideration should be given to an additional satellite group of five animals per sex in the control and the top dose group for observation of reversibility, persistence or delayed occurrence of systemic toxic effects, for at least 14 days post treatment. Animals of the satellite groups will not be mated and, consequently, are not used for the assessment of reproduction/developmental toxicity.

Dosage

28. Generally, at least three test groups and a control group should be used. If there are no suitable general toxicity data available, a range finding study may (animals of the same strain and source) be performed to aid the determination of the doses to be used. Except for treatment with the test chemical, animals in the control group should be handled in an identical manner to the test group subjects. If a vehicle is used in administering the test chemical, the control group should receive the vehicle in the highest volume used.
29. Dose levels should be selected taking into account any existing toxicity and (toxico-) kinetic data available. It should also be taken into account that there may be differences in sensitivity between pregnant and non-pregnant animals. The highest dose level should be chosen with the aim of inducing toxic effects but not death nor obvious suffering. Thereafter, a descending sequence of dose levels should be selected with a view to demonstrating any dosage related response and no adverse effects at the lowest dose level. Two- to four- fold intervals are frequently optimum and addition of a fourth test group is often preferable to using very large intervals (e.g. more than a factor of 10) between dosages.
30. In the presence of observed general toxicity (e.g. reduced body weight, liver, heart, lung or kidney effects, etc.) or other changes that may not be toxic responses (e.g. reduced food intake, liver enlargement), observed effects on endocrine sensitive endpoints should be interpreted with caution.

Limit test

31. If an oral study at one dose level of at least 1 000 mg/kg body weight/day or, for dietary administration, an equivalent percentage in the diet, or drinking water (based upon body weight determinations), using the procedures described for this study, produces no observable toxic effects and if toxicity would not be expected based upon data from structurally related substances, then a full study using several dose levels may not be considered necessary. The limit test applies except when human exposure indicates the need for a higher dose level to be used. For other types of administration, such as inhalation or dermal application, the physical chemical properties of the test chemicals often may dictate the maximum attainable exposure.

Administration of doses

32. The animals are dosed with the test chemical daily for 7 days a week. When the test chemical is administered by gavage, this should be done in a single dose to the animals using a stomach tube or a suitable intubation cannula. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. The volume should not exceed 1 ml/100 g body weight, except in the case of aqueous solutions where 2 ml/100 g body

weight may be used. Except for irritating or corrosive test chemicals which will normally reveal exacerbated effects with higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure a constant volume at all dose levels.

33. For test chemicals administered via the diet or drinking water, it is important to ensure that the quantities of the test chemical involved do not interfere with normal nutrition or water balance. When the test chemical is administered in the diet either a constant dietary concentration (ppm) or a constant dose level in terms of the animals' body weight may be used; the alternative used should be specified. For a test chemical administered by gavage, the dose should be given at similar times each day, and adjusted at least weekly to maintain a constant dose level in terms of animal body weight. Where the combined study is used as a preliminary to a long term or a full reproduction toxicity study, a similar diet should be used in both studies.

Experimental schedule

34. Dosing of both sexes should begin 2 weeks prior to mating, after they have been acclimatised for at least five days and females have been screened for normal oestrous cycles (in a 2 weeks pre-treatment period). The study should be scheduled in such a way that oestrous cycle evaluation begins soon after the animals have attained full sexual maturity. This may vary slightly for different strains of rats in different laboratories, e.g. Sprague Dawley rats 10 weeks of age, Wistar rats about 12 weeks of age. Dams with offspring should be killed on day 13 post-partum, or shortly thereafter. In order to allow for overnight fasting of dams prior to blood collection (if this option is preferred), dams and their offspring need not necessarily be killed on the same day. The day of birth (viz. when parturition is complete) is defined as day 0 post-partum. Females showing no-evidence of copulation are killed 24-26 days after the last day of the mating period. Dosing is continued in both sexes during the mating period. Males should further be dosed after the mating period at least until the minimum total dosing period of 28 days has been completed. They are then killed, or, alternatively, are retained and continued to be dosed for the possible conduction of a second mating if considered appropriate.
35. Daily dosing of the parental females should continue throughout pregnancy and at least up to, and including, day 13 post-partum or the day before sacrifice. For studies where the test chemical is administered by inhalation or by the dermal route, dosing should be continued at least up to, and including, day 19 of gestation, and dosing should be re-initiated as soon as possible and not later than postnatal day (PND) 4.
36. Animals in a satellite group scheduled for follow-up observations, if included, are not mated. They should be kept at least for a further 14 days after the first scheduled kill of dams, without treatment to detect delayed occurrence, or persistence of, or recovery from toxic effects.
37. A diagram of the experimental schedule is given in Appendix 2.

Oestrous cycles

38. Oestrous cycles should be monitored before treatment starts to select for the study females with regular cyclicity (see paragraph 27). Vaginal smears should also be monitored daily from the beginning of the treatment period until evidence of mating. If there is concern about acute stress effects that could alter estrous cycles with the initiation of dosing, laboratories can expose test animals for 2 weeks, then collect vaginal smears daily to monitor estrous cycle for a minimum of two weeks during the pre-mating period with continued monitoring into the mating period until there is evidence of mating. When obtaining vaginal/cervical cells, care should be taken to avoid disturbance of mucosa, which could induce pseudopregnancy (8) (9).

Mating procedure

39. Normally, 1:1 (one male to one female) matings should be used in this study. Exceptions can arise in the case of occasional deaths of males. The female should be placed with the same male until evidence of copulation is observed or two weeks have elapsed. Each morning the females should be examined for the presence of sperm or a vaginal plug. Day 0 of pregnancy is defined as the day on which mating evidence is confirmed (a vaginal plug or sperm is found). In case pairing was unsuccessful, re-mating of females with proven males of the same group could be considered.

Litter size

40. On day 4 after birth, the size of each litter may be adjusted by eliminating extra pups by random selection to yield, as nearly as possible, four or five pups per sex per litter depending on the normal litter size in the strain of rats used. Blood samples should be collected from two of the surplus pups, pooled, and used for determination of serum T4 levels. Selective elimination of pups, e.g. based upon body weight, or anogenital distance (AGD) is not appropriate. Whenever the number of male or female pups prevents having four or five of each sex per litter, partial adjustment (for example, six males and four females) is acceptable. No pups will be eliminated when litter size will drop below the culling target (8 or 10 pups/litter). If there is only one pup available above the culling target, only one pup will be eliminated and used for blood collection for possible serum T4 assessments.
41. If litter size is not adjusted, two pups per litter are sacrificed on day 4 after birth and blood samples are taken for measurement of serum thyroid hormone concentrations. If possible the two pups per litter should be female pups to reserve male pups for nipple retention evaluations, except in the event that removing these pups leaves no remaining females for assessment at termination. No pups will be eliminated when litter size will drop below 8 or 10 pups/litter (depending on the normal litter size in the strain of rats used). If there is only one pup available above the normal litter size, only one pup will be eliminated and used for blood collection for possible serum T4 assessments.

Observations

42. General clinical observations should be made at least once a day, preferably at the same time(s) each day and considering the peak period of anticipated effects after dosing. The health condition of the animals should be recorded. At least twice daily all animals are observed for morbidity and mortality.
43. Once before the first exposure (to allow for within-subject comparisons), and at least once a week thereafter, detailed clinical observations should be made in all parental animals. These observations should be made outside the home cage in a standard arena and preferably at the same time, each day. They should be carefully recorded; preferably using scoring systems, explicitly defined by the testing laboratory. Effort should be made to ensure that variations in the test conditions are minimal and that observations are preferably conducted by observers unaware of the treatment. Signs noted should include, but not be limited to, changes in skin, fur, eyes, mucous membranes, occurrence of secretions and excretions and autonomic activity (e.g. lacrimation, piloerection, pupil size, unusual respiratory pattern). Changes in gait, posture and response to handling as well as the presence of clonic or tonic movements, stereotypies (e.g. excessive grooming, repetitive circling), difficult or prolonged parturition or bizarre behaviour (e.g. self-mutilation, walking backwards) should also be recorded (10).
44. At one time during the study, sensory reactivity to stimuli of different modalities (e.g. auditory, visual and proprioceptive stimuli) (8) (9) (11), assessment of grip strength (12) and motor activity assessment (13) should be conducted in five males and five females, randomly selected from each group. Further details of the procedures that could be followed are given in the respective references. However, alternative procedures than those referenced could also be used. In males, these functional observations should be made towards the end of their dosing period, shortly before scheduled kill but before blood sampling for haematology or clinical chemistry (see paragraphs 53-56, including footnote 1). Females should be in a physiologically similar state during these functional tests and should preferably be tested once during the last week of lactation (e.g., LD 6-13), shortly before scheduled kill. To the extent possible, minimise dams and pups separation times.
45. Functional observations made once towards the end of the study may be omitted when the study is conducted as a preliminary study to a subsequent subchronic (90-day) or long-term study. In that case, the functional observations should be included in this follow-up study. On the other hand, the availability of data on functional observations from this repeated dose study may enhance the ability to select dose levels for a subsequent subchronic or long-term study.
46. As an exception, functional observations may also be omitted for groups that otherwise reveal signs of toxicity to an extent that would significantly interfere with the functional test performance.
47. The duration of gestation should be recorded and is calculated from day 0 of pregnancy. Each litter should be examined as soon as possible after delivery to establish the number and sex of pups, stillbirths, live births, runts (pups that are significantly smaller than corresponding control pups), and the presence of gross abnormalities.
48. Live pups should be counted and sexed and litters weighed within 24 hours of parturition (day 0 or 1 post-partum) and at least on day 4 and day 13 post-partum. In addition to the observations on parent animals (see paragraphs 43 and 44), any abnormal behaviour of the offspring should be recorded.

49. The AGD of each pup should be measured on the same postnatal day between PND 0 through PND 4. Pup body weight should be collected on the day the AGD is measured and the AGD should be normalised to a measure of pup size, preferably the cube root of body weight (14). The number of nipples/areolae in male pups should be counted on PND 12 or 13 as recommended in OECD GD 151 (15).

Body weight and food/water consumption

50. Males and females should be weighed on the first day of dosing, at least weekly thereafter, and at termination. During pregnancy, females should be weighed on days 0, 7, 14 and 20 and within 24 hours of parturition (day 0 or 1 post-partum), and at least day 4 and day 13 post-partum. These observations should be reported individually for each adult animal.
51. During pre-mating, pregnancy and lactation, food consumption should be measured at least weekly. The measurement of food consumption during mating is optional. Water consumption during these periods should also be measured, when the test chemical is administered by that medium.

Haematology

52. Once during the study, the following haematological examinations should be made in five males and five females randomly selected from each group: haematocrit, haemoglobin concentrations, erythrocyte count, reticulocytes, total and differential leucocyte count, platelet count and a measure of blood clotting time/potential. Other determinations that should be carried out, if the test chemical or its putative metabolites have or are suspected to have oxidising properties include methaemoglobin concentration and Heinz bodies.
53. Blood samples should be taken from a named site. Females should be in a physiologically similar state during sampling. In order to avoid practical difficulties related to the variability in the onset of gestation, blood collection in females may be done at the end of the pre-mating period as an alternative to sampling just prior to, or as part of, the procedure for euthanasia of the animals. Blood samples of males should preferably be taken just prior to, or as part of, the procedure for euthanasia of the animals. Alternatively, blood collection in males may also be done at the end of the pre-mating period when this time point was preferred for females.
54. Blood samples should be stored under appropriate conditions.

Clinical biochemistry

55. Clinical biochemistry determinations to investigate major toxic effects in tissues and, specifically, effects on kidney and liver, should be performed on blood samples obtained from the selected five males and five females of each group. Overnight fasting of the animals prior to blood sampling is recommended⁽¹⁾. Investigations of plasma or serum should include sodium, potassium, glucose, total cholesterol, urea, creatinine, total protein and albumin, at least two enzymes indicative of hepatocellular effects (such as alanin aminotransferase, aspartate aminotransferase and sorbitol dehydrogenase) and bile acids. Measurements of additional enzymes (of hepatic or other origin) and bilirubin may provide useful information under certain circumstances.
56. Blood samples from a defined site are taken based on the following schedule:
- from at least two pups per litter on day 4 after birth, if the number of pups allows (see paragraphs 40-41)
 - from all dams and at least two pups per litter at termination on day 13, and
 - from all adult males, at termination

All blood samples are stored under appropriate conditions. Blood samples from the day 13 pups and the adult males are assessed for serum levels for thyroid hormones (T4). Further assessment of T4 in blood samples from the dams and day 4 pups is done if relevant. As an option, other hormones may be measured if relevant. Pup blood can be pooled by litter for thyroid hormone analyses. Thyroid hormones (T4 and TSH) should preferably be measured as 'total'.

⁽¹⁾ For a number of measurements in serum and plasma, most notably for glucose, overnight fasting would be preferable. The major reason for this preference is that the increased variability which would inevitably result from non-fasting, would tend to mask more subtle effects and make interpretation difficult. On the other hand, however, overnight fasting may interfere with the general metabolism of the (pregnant) animals, disturbs lactation and nursing behaviour, and, particularly in feeding studies, may disturb the daily exposure to the test chemical. If overnight fasting is adopted, clinical biochemical determinations should be performed after the conduct of functional observations in week 4 of the study for the males. The dams should be retained for an additional day after the pups are removed on e.g. PND 13). Dams should be fasted overnight from lactation day 13-14 and terminal blood used for clinical chemistry parameters.

57. Optionally, the following urinalysis determinations could be performed in five randomly selected males of each group during the last week of the study using timed urine volume collection; appearance, volume, osmolality or specific gravity, pH, protein, glucose and blood/blood cells.
58. In addition, studies to investigate serum markers of general tissue damage should be considered. Other determinations that should be carried out if the known properties of the test chemical may, or are suspected to, affect related metabolic profiles include calcium, phosphate, fasting triglycerides and fasting glucose, specific hormones, methaemoglobin and cholinesterase. These need to be identified on a case-by-case basis.
59. The following factors may influence the variability and the absolute concentrations of the hormone determinations:
- time of sacrifice because of diurnal variation of hormone concentrations
 - method of sacrifice to avoid undue stress to the animals that may affect hormone concentrations
 - test kits for hormone determinations that may differ by their standard curves.
60. Plasma samples specifically intended for hormone determination should be obtained at a comparable time of the day. The numerical values obtained when analysing hormone concentrations differ with various commercial assay kits.
61. If historical baseline data are inadequate, consideration should be given to determination of haematological and clinical biochemistry variables before dosing commences or preferably in a set of animals not included in the experimental groups. For females, the data have to be from lactating animals.

PATHOLOGY

Gross necropsy

62. All adult animals in the study should be subjected to a full, detailed gross necropsy which includes careful examination of the external surface of the body, all orifices, and the cranial, thoracic and abdominal cavities and their contents. Special attention should be paid to the organs of the reproductive system. The number of implantation sites should be recorded. Vaginal smears should be examined on the day of necropsy to determine the stage of the oestrous cycle and allow correlation with histopathology of female reproductive organs.
63. The testes and epididymides as well as prostate and seminal vesicles with coagulating glands as a whole of all male adult animals should be trimmed of any adherent tissue, as appropriate, and their wet weight taken as soon as possible after dissection to avoid drying. In addition, optional organ weights could include levator ani plus bulbo-cavernosus muscle complex, Cowper's glands and glans penis in males and paired ovaries (wet weight) and uterus (including cervix) in females; if included, these weights should be collected as soon as possible after dissection. The ovaries, testes, epididymides, accessory sex organs, and all organs showing macroscopic lesions of all adult animals, should be preserved.
64. From all adult males and females and one male and female day 13 pup from each litter thyroid glands should be preserved in the most appropriate fixation medium for the intended subsequent histopathological examination. The thyroid weight could be determined after fixation. Trimming should also be done very carefully and only after fixation to avoid tissue damage. Tissue damage could compromise histopathology analysis. Blood samples should be taken from a named site just prior to or as part of the procedure for euthanasia of the animals, and stored under appropriate conditions (see paragraph 56).
65. In addition, for a least five adult males and females, randomly selected from each group (apart from those found moribund and/or euthanised prior to the termination of the study), the liver, kidneys, adrenals, thymus, spleen, brain and heart should be trimmed of any adherent tissue, as appropriate and their wet weight taken as soon as possible after dissection to avoid drying. The following tissues should be preserved in the most appropriate fixation medium for both the type of tissue and the intended subsequent histopathological examination: all gross lesions, brain (representative regions including cerebrum, cerebellum and pons), spinal cord, eye, stomach, small and large intestines (including Peyer's patches), liver, kidneys, adrenals, spleen, heart, thymus, trachea and lungs (preserved by inflation with fixative and then immersion), gonads (testis and ovaries), accessory sex organs (uterus and cervix),

epididymides, prostate, seminal vesicles plus coagulating glands), vagina, urinary bladder, lymph nodes (besides the most proximal draining node, another lymph node should be taken according to the laboratory's experience (16)), peripheral nerve (sciatic or tibial) preferably in close proximity to the muscle, skeletal muscle and bone, with bone marrow (section or, alternatively, a fresh mounted bone marrow aspirate). It is recommended that testes be fixed by immersion in Bouin's or modified Davidson's fixative (16) (17) (18); formalin fixation is not recommended for these tissues. The tunica albuginea may be gently and shallowly punctured at the both poles of the organ with a needle to permit rapid penetration of the fixative. The clinical and other findings may suggest the need to examine additional tissues. Also any organs considered likely to be target organs based on the known properties of the test chemical should be preserved.

66. The following tissues may give valuable indication for endocrine-related effects: Gonads (ovaries and testes), accessory sex organs (uterus including cervix, epididymides, seminal vesicles with coagulation glands, dorsolateral and ventral prostate), vagina, pituitary, male mammary gland and adrenal gland. Changes in male mammary glands have not been sufficiently documented but this parameter may be very sensitive to substances with estrogenic action. Observation of organs/tissues that are not listed in paragraph 65 is optional.
67. Dead pups and pups killed at day 13 post-partum, or shortly thereafter, should, at least, be carefully examined externally for gross abnormalities. Particular attention should be paid to the external reproductive genitals which should be examined for signs of altered development.

Histopathology

68. Full histopathology should be carried out on the preserved organs and tissues of the selected animals in the control and high dose groups (with special emphasis on stages of spermatogenesis in the male gonads and histopathology of interstitial testicular cell structure). The thyroid gland from pups and from the remaining adult animals may be examined when necessary. These examinations should be extended to animals of other dosage groups, if treatment-related changes are observed in the high dose group. The Guidance on histopathology (10) details extra information on dissection, fixation, sectioning and histopathology of endocrine tissues.
69. All gross lesions should be examined. To aid in the elucidation of NOAELs, target organs in other dose groups should be examined, particularly in groups claimed to show a NOAEL.
70. When a satellite group is used, histopathology should be performed on tissues and organs identified as showing effects in the treated groups.

DATA AND REPORTING

Data

71. Individual animal data should be provided. Additionally, all data should be summarised in tabular form, showing for each test group the number of animals at the start of the test, the number of animals found dead during the test or euthanised for humane reasons, the time of any death or euthanasia, the number of fertile animals, the number of pregnant females, the number of animals showing signs of toxicity, a description of the signs of toxicity observed, including time of onset, duration, and severity of any toxic effects, the types of histopathological changes, and all relevant litter data. A tabular summary report format, which has proven to be very useful for the evaluation of reproductive/developmental effects, is given in Appendix 3.
72. When possible, numerical results should be evaluated by an appropriate and general acceptable statistical method. Comparisons of the effect along a dose range should avoid the use of multiple t-tests. The statistical methods should be selected during the design of the study. Statistical analysis of AGD and nipple retention should be based on individual pup data, taking litter effects into account. Where appropriate, the litter is the unit of analysis. Statistical analysis of pup body weight should be based on individual pup data, taking litter size into account. Due to the limited dimensions of the study, statistical analyses in the form of tests for "significance" are of limited value for many endpoints, especially reproductive endpoints. Some of the most widely used methods, especially parametric tests for measures of central tendency, are inappropriate. If statistical analyses are used then the method chosen should be appropriate for the distribution of the variable examined and be selected prior to the start of the study.

Evaluation of results

73. The findings of this toxicity study should be evaluated in terms of the observed effects, necropsy and microscopic findings. The evaluation will include the relationship between the dose of the test chemical and the presence or absence, incidence and severity of abnormalities, including gross lesions, identified target organs, infertility, clinical abnormalities, affected reproductive and litter performance, body weight changes, effects on mortality and any other toxic effects.
74. Because of the short period of treatment of the male, the histopathology of the testes and epididymides should be considered along with the fertility data, when assessing male reproduction effects. The use of historic control data on reproduction/development (e.g. for litter size, AGD, nipple retention, serum T4 levels), where available, may also be useful as an aid to the interpretation of the study.
75. For quality control it is proposed that historical control data are collected and that for numerical data coefficients of variation are calculated, especially for the parameters linked with endocrine disrupter detection. These data can be used for comparison purposes when actual studies are evaluated.

Test report

76. The test report should include the following information:

Test chemical:

- source, lot number, limit date for use, if available
- stability of the test chemical, if known.

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Vehicle (if appropriate):

- justification for choice of vehicle, if other than water.

Test animals:

- species/strain used;
- number, age and sex of animals;
- source, housing conditions, diet, etc.;
- individual weights of animals at the start of the test.

- justification for species if not rat

Test conditions:

- rationale for dose level selection;
- details of test chemical formulation/diet preparation, achieved concentration, stability and homogeneity of the preparation;
- details of the administration of the test chemical;
- conversion from diet/drinking water test chemical concentration (ppm) to the actual dose (mg/kg body weight/day), if applicable;
- details of food and water quality;
- detailed description of the randomisation procedure to select pups for culling, if culled.

Results:

- body weight/body weight changes;
- food consumption and water consumption, if applicable;
- toxic response data by sex and dose, including fertility, gestation, and any other signs of toxicity;
- gestation length;
- toxic or other effects on reproduction, offspring, postnatal growth, etc.;
- nature, severity and duration of clinical observations (whether reversible or not);
- sensory activity, grip strength and motor activity assessments;
- haematological tests with relevant baseline values;
- clinical biochemistry tests with relevant base-line values;
- number of adult females with normal or abnormal oestrous cycle and cycle duration;
- number of live births and post implantation loss;
- number of pups with grossly visible abnormalities; gross evaluation of external genitalia, number of runts;
- time of death during the study or whether animals survived to termination;

- number of implantations, litter size and litter weights at the time of recording;
- pup body weight data
- AGD of all pups (and body weight on day of AGD measurement)
- nipple retention in male pups,
- thyroid hormone levels, day 13 pups and adult males (and dams and day 4 pups if measured)
- body weight at sacrifice and organ weight data for the parental animals;
- necropsy findings;
- a detailed description of histopathological findings;
- absorption data (if available);
- statistical treatment of results, where appropriate.

Discussion of results.

Conclusions.

Interpretation of Results

77. The study will provide evaluations of reproduction/developmental toxicity associated with administration of repeated doses. In particular, since emphasis is placed on both general toxicity and reproduction/developmental toxicity endpoints, the results of the study will allow for the discrimination between reproduction/developmental effects occurring in the absence of general toxicity and those which are only expressed at levels that are also toxic to parent animals (see paragraphs 7-11). It could provide an indication of the need to conduct further investigations and could provide guidance in the design of subsequent studies. OECD Guidance Document 43 should be consulted for aid in the interpretation of reproduction and developmental results (19). OECD Guidance Document 106 on Histologic Evaluation of Endocrine and Reproductive Tests in Rodents (16) provides information on the preparation and evaluation of (endocrine) organs and vaginal smears that may be helpful for this test method.

LITERATURE

- (1) OECD (1990). Room Document No 1 for the 14th Joint Meeting of the Chemicals Group and Management Committee. Available upon request at Organisation for Economic Cooperation and Development, Paris
- (2) OECD (1992). Chairman's Report of the ad hoc Expert Meeting on Reproductive Toxicity Screening Methods, Tokyo, 27th-29th October, 1992. Available upon request at Organisation for Economic Cooperation and Development, Paris
- (3) Mitsumori K., Kodama Y., Uchida O., Takada K., Saito M. Naito K., Tanaka S., Kurokawa Y., Usami, M., Kawashima K., Yasuhara K., Toyoda K., Onodera H., Furukawa F., Takahashi M. and Hayashi Y. (1994). Confirmation Study, Using Nitro-Benzene, of the Combined Repeat Dose and Reproductive/ Developmental Toxicity Test Protocol Proposed by the Organization for Economic Cooperation and Development (OECD). *J. Toxicol, Sci.*, 19, 141-149.
- (4) Tanaka S., Kawashima K., Naito K., Usami M., Nakadate M., Imaida K., Takahashi M., Hayashi Y., Kurokawa Y. and Tobe M. (1992). Combined Repeat Dose and Reproductive/Developmental Toxicity Screening Test (OECD): Familiarization Using Cyclophosphamide. *Fundam. Appl. Toxicol.*, 18, 89-95.

- (5) OECD (1998). Report of the First Meeting of the OECD Endocrine Disrupter Testing and Assessment (EDTA) Task Force, 10th-11th March 1998, Available upon request at Organisation for Economic Cooperation and Development, Paris
- (6) OECD (2015). Feasibility Study for Minor Enhancements of TG 421/422 with ED Relevant Endpoints. Environment, Health and Safety Publications, Series on Testing and Assessment (No 217), Organisation for Economic Cooperation and Development, Paris.
- (7) OECD (2000). Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluations, Environment, Health and Safety Publications, Series on Testing and Assessment, (No 19), Organisation for Economic Cooperation and Development, Paris.
- (8) Goldman J.M., Murr A.S., Buckalew A.R., Ferrell J.M. and Cooper R.L. (2007). The Rodent Estrous Cycle: Characterization of Vaginal Cytology and its Utility in Toxicological Studies, *Birth Defects Research, Part B*, 80 (2), 84-97.
- (9) Sadleir R.M.F.S. (1979). Cycles and Seasons, in Auston C.R. and Short R.V. (Eds.), *Reproduction in Mammals: I. Germ Cells and Fertilization*, Cambridge, New York.
- (10) IPCS (1986). Principles and Methods for the Assessment of Neurotoxicity Associated with Exposure to Chemicals. Environmental Health Criteria Document (No 60).
- (11) Moser V.C., McDaniel K.M. and Phillips P.M. (1991). Rat Strain and Stock Comparisons Using a Functional Observational Battery: Baseline Values and Effects of Amitraz. *Toxicol. Appl. Pharmacol.*, 108, 267-283.
- (12) Meyer O.A., Tilson H.A., Byrd W.C. and Riley M.T. (1979). A Method for the Routine Assessment of Fore- and Hindlimb Grip Strength of Rats and Mice. *Neurobehav. Toxicol.*, 1, 233-236.
- (13) Crofton K.M., Howard J.L., Moser V.C., Gill M.W., Reiter L.W., Tilson H.A., MacPhail R.C. (1991). Interlaboratory Comparison of Motor Activity Experiments: Implication for Neurotoxicological Assessments. *Neurotoxicol. Teratol.* 13, 599-609.
- (14) Gallavan R.H. Jr, J.F. Holson, D.G. Stump, J.F. Knapp and V.L. Reynolds. (1999). "Interpreting the Toxicologic Significance of Alterations in Anogenital Distance: Potential for Confounding Effects of Progeny Body Weights", *Reproductive Toxicology*, 13: 383-390.
- (15) OECD (2013). Guidance Document in Support of the Test Guideline on the Extended One Generation Reproductive Toxicity Study. Environment, Health and Safety Publications, Series on Testing and Assessment (No 151). Organisation for Economic Cooperation and Development, Paris.
- (15) OECD (2009). Guidance Document for Histologic Evaluation of Endocrine and Reproductive Tests in Rodents. Environment, Health and Safety Publications, Series on Testing and Assessment (No. 106) Organisation for Economic Cooperation and Development, Paris.
- (17) Hess RA and Moore BJ. (1993). Histological Methods for the Evaluation of the Testis. In: *Methods in Reproductive Toxicology*, Chapin RE and Heindel JJ (Eds.). Academic Press: San Diego, CA, pp. 52-85.
- (18) Latendresse JR, Warbritton AR, Jonassen H, Creasy DM. (2002). Fixation of Testes and Eyes Using a Modified Davidson's Fluid: Comparison with Bouin's Fluid and Conventional Davidson's fluid. *Toxicol. Pathol.* 30, 524-533.
- (19) OECD (2008). Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment. Environment, Health and Safety Publications, Series on Testing and Assessment (No 43), Organisation for Economic Cooperation and Development, Paris.
- (20) OECD (2011), Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption (No 150), Organisation for Economic Cooperation and Development, Paris.

Appendix 1

DEFINITIONS (SEE ALSO (20) OECD GD 150)

Androgenicity is the capability of a chemical to act like a natural androgenic hormone (e.g. testosterone) in a mammalian organism.

Antiandrogenicity is the capability of a chemical to suppress the action of a natural androgenic hormone (e.g. testosterone) in a mammalian organism.

Antioestrogenicity is the capability of a chemical to suppress the action of a natural oestrogenic hormone (e.g. oestradiol 17 β) in a mammalian organism.

Antithyroid activity is the capability of a chemical to suppress the action of a natural thyroid hormone (e.g. T₃) in a mammalian organism.

Chemical is a substance or a mixture.

Developmental toxicity: the manifestation of reproductive toxicity, representing pre-, peri- post-natal, structural, or functional disorders in the progeny.

Dose is the amount of test chemical administered. The dose is expressed as weight of test chemical per unit body weight of test animal per day (e.g. mg/kg body weight/day), or as a constant dietary concentration.

Dosage is a general term comprising dose, its frequency and the duration of dosing.

Evident toxicity is a general term describing clear signs of toxicity following administration of test chemical. These should be sufficient for hazard assessment and should be such that an increase in the dose administered can be expected to result in the development of severe toxic signs and probable mortality.

Impairment of fertility represents disorders of male or female reproductive functions or capacity.

Maternal toxicity: adverse effects on gravid females, occurring either specifically (direct effect) or not specifically (indirect effect) and being related to the gravid state.

NOAEL is the abbreviation for no-observed-adverse-effect level. This is the highest dose level where no adverse treatment-related findings are observed due to treatment.

Oestrogenicity is the capability of a chemical to act like a natural oestrogenic hormone (e.g. oestradiol 17 β) in a mammalian organism.

Reproduction toxicity represents harmful effects on the progeny and/or an impairment of male and female reproductive functions or capacity.

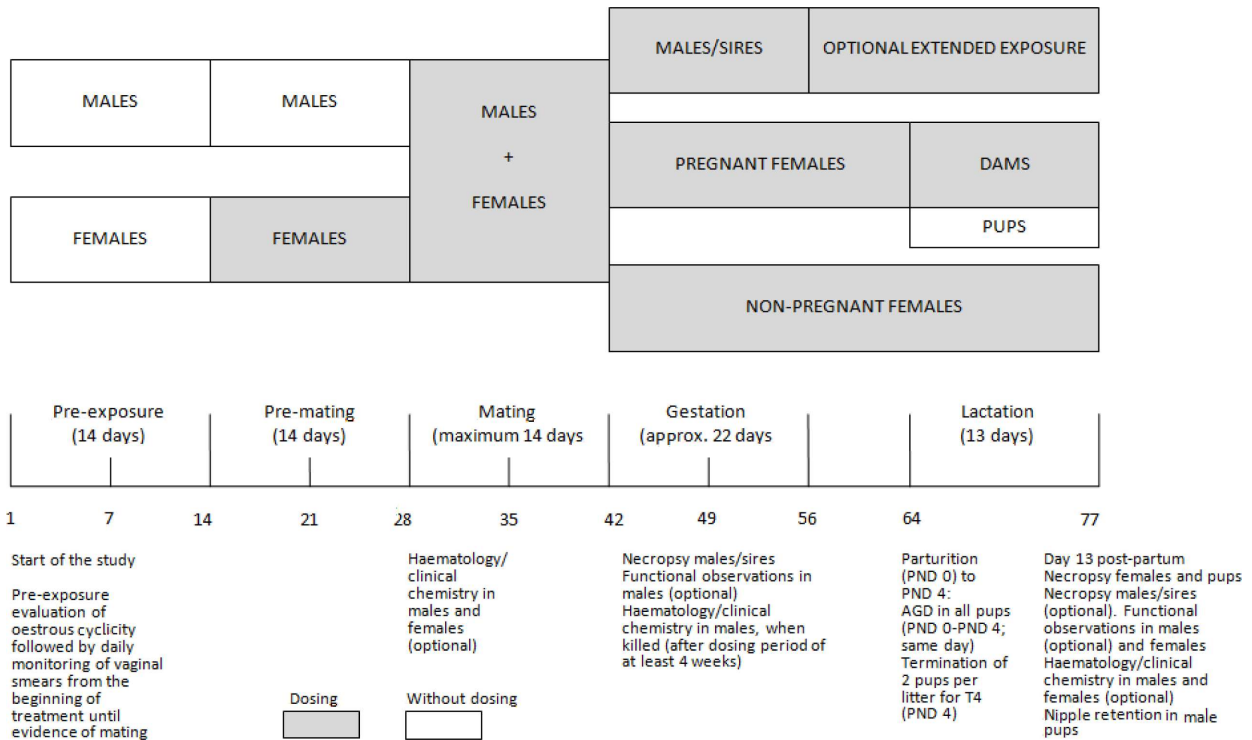
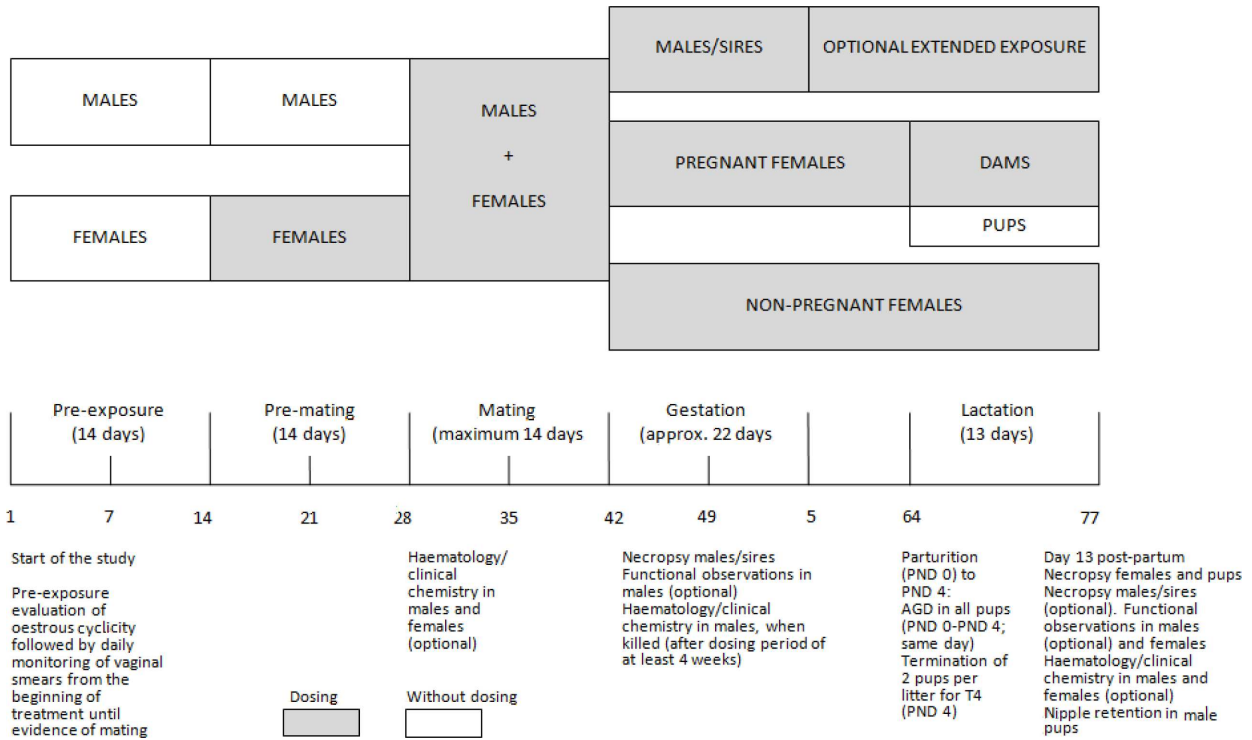
Test chemical is any substance or mixture tested using this test method.

Thyroid activity is the capability of a chemical to act like a natural thyroid hormone (e.g. T₃) in a mammalian organism.

Validation is a scientific process designed to characterise the operational requirements and limitations of a test method and to demonstrate its reliability and relevance for a particular purpose.

Appendix 2

DIAGRAM OF THE EXPERIMENTAL SCHEDULE, INDICATING THE MAXIMUM STUDY DURATION, BASED ON A FULL 14-DAY MATING PERIOD



Appendix 3

TABULAR SUMMARY REPORT OF EFFECTS ON REPRODUCTION/DEVELOPMENT

OBSERVATIONS	VALUES				
	0 (control)
Dosage (units).....	0 (control)
Pairs started (N)					
Oestrus cycle (at least mean length and frequency of irregular cycles)					
Females showing evidence of copulation (N)					
Females achieving pregnancy (N)					
Conceiving days 1 - 5 (N)					
Conceiving days 6 -... ⁽¹⁾ (N)					
Pregnancy ≤ 21 days (N)					
Pregnancy = 22 days (N)					
Pregnancy ≥ 23 days (N)					
Dams with live young born (N)					
Dams with live young at day 4 pp (N)					
Implants/dam (mean)					
Live pups/dam at birth (mean)					
Live pups/dam at day 4 (mean)					
Sex ratio (m/f) at birth (mean)					
Sex ratio (m/f) at day 4 (mean)					
Litter weight at birth (mean)					
Litter weight at day 4 (mean)					
Pup weight at birth (mean)					
Pup weight at the time of AGD measurement(mean males, mean females)					
Pup AGD on the same postnatal day, birth- day 4 (mean males, mean females, note PND)					

OBSERVATIONS	VALUES				
Pup weight at day 4 (mean)					
Pup weight at day 13 (mean)					
Male pup nipple retention at day 13 (mean)					
ABNORMAL PUPS					
Dams with 0					
Dams with 1					
Dams with ≥ 2					
LOSS OF OFFSPRING					
Pre-natal (implantations minus live births)					
Females with 0					
Females with 1					
Females with 2					
Females with ≥ 3					
Post-natal (live births minus alive at post natal day 13)					
Females with 0					
Females with 1					
Females with 2					
Females with ≥ 3					
(!) last day of the mating period					

B.65 IN VITRO MEMBRANE BARRIER TEST METHOD FOR SKIN CORROSION

INTRODUCTION

1. This test method is equivalent to OECD test guideline (TG) 435 (2015). Skin corrosion refers to the production of irreversible damage to the skin, manifested as visible necrosis through the epidermis and into the dermis, following the application of a test chemical as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (1) and the European Union (EU) Regulation 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) (1). This test method, equivalent to the updated OECD test guideline 435 provides an *in vitro* membrane barrier test method that can be used to identify corrosive chemicals. The test method utilises an artificial membrane designed to respond to corrosive chemicals in a manner similar to animal skin *in situ*.
2. Skin corrosivity has traditionally been assessed by applying the test chemical to the skin of living animals and assessing the extent of tissue damage after a fixed period of time (2). Besides the present test method, a number of other *in vitro* test methods have been adopted as alternatives (3)(4) to the standard *in vivo* rabbit skin procedure (Chapter B.4 of this Annex, equivalent to OECD TG 404) used to identify corrosive chemicals (2). The UN GHS tiered testing and evaluation strategy for the assessment and classification of skin corrosivity and the OECD Guidance Document on Integrated Approaches to Testing and Assessment (IATA) for Skin Irritation/Corrosion recommend the use of validated and accepted *in vitro* test methods under modules 3 and 4 (1)(5). The IATA describes several modules which group information sources and analysis tools and (i) provides guidance on how to integrate and use existing test and non-test data for the assessment of the skin irritation and skin corrosion potentials of chemicals and (ii) proposes an approach when further testing is needed, including when negative results are found (5). In this modular approach, positive results from *in vitro* test methods can be used to classify a chemical as corrosive without the need for animal testing, thus reducing and refining the use of animals in and avoiding the pain and distress that might occur if animals were used for this purpose.
3. Validation studies have been completed for the *in vitro* membrane barrier model commercially available as Corrositex® (6)(7)(8), showing an overall accuracy to predict skin corrosivity of 79 % (128/163), a sensitivity of 85 % (76/89), and a specificity of 70 % (52/74) for a database of 163 substances and mixtures (7). Based on its acknowledged validity, this validated reference method (VRM) has been recommended for use as part of a tiered testing strategy for assessing the dermal corrosion hazard potential of chemicals (5)(7). Before an *in vitro* membrane barrier model for skin corrosion can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined to ensure that it is similar to that of the VRM (9), in accordance with the pre-defined performance standards (PS) (10). The OECD Mutual Acceptance of Data will only be guaranteed after any proposed new or updated method following the PS have been reviewed and included in the equivalent OECD test guideline. Currently, only one *in vitro* method is covered by OECD test guideline 435 and this test method, the commercially available Corrositex® model.
4. Other test methods for skin corrosivity testing are based on the use of reconstituted human skin (OECD TG 431) (3) and isolated rat skin (OECD TG 430) (4). This Test Guideline also provides for subcategorisation of corrosive chemicals into the three UN GHS Sub-categories of corrosivity and the three UN Transport Packing Groups for corrosivity hazard. This Test Guideline was originally adopted in 2006 and updated in 2015 to refer to the IATA guidance document and update the list of proficiency substances.

DEFINITIONS

5. Definitions used are provided in the Appendix.

INITIAL CONSIDERATIONS AND LIMITATIONS

6. The test described in this test method allows the identification of corrosive test chemicals and allows the sub-categorisation of corrosive test chemicals according to UN GHS/CLP (Table 1). In addition, such a test method may be used to make decisions on the corrosivity and non-corrosivity of specific classes of chemicals, e.g. organic and inorganic acids, acid derivatives (2), and bases for certain transport testing purposes (7)(11)(12). This test method describes a generic procedure similar to the validated reference test method (7). While this test method does not provide adequate information on skin irritation, it should be noted that TM B.46 (equivalent to OECD TG439)

(1) Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006, OJ L 353/1, 31.12.2008

(2) "Acid derivative" is a non-specific class designation and is broadly defined as a chemical produced from an acid either directly or by modification or partial substitution. This class includes anhydrides, halo acids, salts, and other types of chemicals.

specifically addresses the health effect skin irritation *in vitro* (13). For a full evaluation of local skin effects after a single dermal exposure, the OECD Guidance Document on Integrated Approaches for Testing Assessment should be consulted (5).

Table 1

The UN GHS Skin Corrosive Category and Subcategories ⁽¹⁾

Corrosive Category (category 1) (for authorities not using subcategories)	Potential Corrosive Subcategories ⁽¹⁾ (for authorities using subcategories, including the CLP Regulation)	Corrosive in ≥ 1 of 3 animals	
		Exposure	Observation
Corrosive	Corrosive subcategory 1A	≤ 3 minutes	≤ 1 hour
	Corrosive subcategory 1B	> 3 minutes / ≤ 1 hour	≤ 14 days
	Corrosive subcategory 1C	> 1 hour / ≤ 4 hours	≤ 14 days

⁽¹⁾ For the EU, the CLP Regulation applies the three skin corrosion subcategories 1A, 1B and 1C.

7. A limitation of the validated reference method (7) is that many non-corrosive chemicals and some corrosive chemicals may not qualify for testing, based on the results of the initial compatibility test (see paragraph 13). Aqueous chemicals with a pH in the range of 4.5 to 8.5 often do not qualify for testing; however, 85 % of chemicals tested in this pH range were non-corrosive in animal tests (7). The *in vitro* membrane barrier method may be used to test solids (soluble or insoluble in water), liquids (aqueous or non-aqueous), and emulsions. However, test chemicals not causing a detectable change in the compatibility test (*i.e.* colour change in the Chemical Detection System (CDS) of the validated reference test method) cannot be tested with the membrane barrier method and should be tested using other test methods.

PRINCIPLE OF THE TEST

8. The test system comprises two components: a synthetic macromolecular bio-barrier and a chemical detection system (CDS); this test method detects via the CDS membrane barrier damage caused by corrosive test chemicals after the application of the test chemical to the surface of the synthetic macromolecular membrane barrier (7), presumably by the same mechanism(s) of corrosion that operate on living skin.
9. Penetration of the membrane barrier (or breakthrough) might be measured by a number of procedures or CDS, including a change in the colour of a pH indicator dye or in some other property of the indicator solution below the barrier.
10. The membrane barrier should be determined to be valid, *i.e.* relevant and reliable, for its intended use. This includes ensuring that different preparations are consistent in regard to barrier properties, *e.g.* capable of maintaining a barrier to non-corrosive chemicals, able to categorise the corrosive properties of chemicals across the various UN GHS Subcategories of corrosivity (1). The classification assigned is based on the time it takes a chemical to penetrate through the membrane barrier to the indicator solution.

DEMONSTRATION OF PROFICIENCY

11. Prior to routine use of the *in vitro* membrane barrier method, adhering to this test method, laboratories should demonstrate technical proficiency by correctly classifying the twelve Proficiency Substances recommended in Table 2. In situations where a listed substance is unavailable or where justifiable, another substance for which adequate *in vivo* and *in vitro* reference data are available may be used (*e.g.* from the list of reference chemicals (10)) provided that the same selection criteria as described in Table 1 is applied.

Table 2

Proficiency Substances ⁽¹⁾

Substance ⁽²⁾	CASRN	Chemical Class	<i>In Vivo</i> UN GHS Sub-category ⁽³⁾	<i>In Vitro</i> UN GHS Sub-category ⁽³⁾
Boron trifluoride dihydrate	13 319-75-0	Inorganic acids	1A	1A
Nitric acid	7 697-37-2	Inorganic acids	1A	1A
Phosphorus pentachloride	10 026-13-8	Precursors of inorganic acids	1A	1A
Valeryl chloride	638-29-9	Acid chlorides	1B	1B
Sodium Hydroxide	1 310-73-2	Inorganic bases	1B	1B
1-(2-Aminoethyl) piperazine	140-31-8	Aliphatic amines	1B	1B
Benzenesulfonyl chloride	98-09-9	Acid chlorides	1C	1C
N,N-Dimethyl benzylamine	103-83-3	Anilines	1C	1C
Tetraethylenepentamine	112-57-2	Aliphatic amines	1C	1C
Eugenol	97-53-0	Phenols	NC	NC
Nonyl acrylate	2 664-55-3	Acrylates/methacrylates	NC	NC
Sodium bicarbonate	144-55-8	Inorganic salts	NC	NC

⁽¹⁾ The twelve substances listed above contain three substances from each of the three UN GHS subcategories for corrosive substances and three non-corrosive substances, are readily available from commercial suppliers, and the UN GHS subcategory is based on the results of high-quality *in vivo* testing. These substances are taken from the list of 40 reference substances that are included in the minimum list of chemicals identified for demonstrating the accuracy and reliability of test methods that are structurally and functionally similar to the validated reference test method, and were selected from the 163 reference chemicals that were originally used to validate the reference test method (Corrositex[®]) (7) (10) (14). The goal of this selection process was to include, to the extent possible, chemicals that: were representative of the range of corrosivity responses (e.g. non-corrosives; UN Packing Groups I, II, and III corrosives) that the validated reference test method is capable of measuring or predicting; were representative of the chemical classes used during the validation process; have chemical structures that were well-defined; induced reproducible results in the validated reference test method; induced definitive results in the *in vivo* reference test; were commercially available; and were not associated with prohibitive disposal costs (14).

⁽²⁾ Substances tested neat or with purity 90 %

⁽³⁾ The corresponding UN Packing groups are I, II and III, respectively, for the UN GHS Sub-categories 1A, 1B and 1C. NC; Non-corrosive.

PROCEDURE

- The following paragraphs describe the components and procedures of an artificial membrane barrier test method for corrosivity assessment (7)(15), based on the current VRM, i.e. the commercially available Corrositex[®]. The membrane barrier and the compatibility/indicator and categorisation solutions can be constructed, prepared or obtained commercially such as in the case of the VRM Corrositex[®]. A sample test method protocol for the validated reference test method is available (7). Testing should be performed at ambient temperature (17-25°C) and the components should comply with the following conditions.

Test Chemical Compatibility Test

- Prior to performing the membrane barrier test, a compatibility test is performed to determine if the test chemical is detectable by the CDS. If the CDS does not detect the test chemical, the membrane barrier test method is not suitable for evaluating the potential corrosivity of that particular test chemical and a different test method should be used. The CDS and the exposure conditions used for the compatibility test should reflect the exposure in the subsequent membrane barrier test.

Test Chemical Timescale Category Test

14. If appropriate for the test method, a test chemical that has been qualified by the compatibility test should be subjected to a timescale category test, *i.e.* a screening test to distinguish between weak and strong acids or bases. For example, in the validated reference test method a timescale categorisation test is used to indicate which of two timescales should be used based on whether significant acid or alkaline reserve is detected. Two different break-through timescales should be used for determining corrosivity and UN GHS skin corrosivity Sub-category, based on the acid or alkali reserve of the test chemical.

MEMBRANE BARRIER TEST METHOD COMPONENTS

Membrane Barrier

15. The membrane barrier consists of two components: a proteinaceous macromolecular aqueous gel and a permeable supporting membrane. The proteinaceous gel should be impervious to liquids and solids but can be corroded and made permeable. The fully constructed membrane barrier should be stored under pre-determined conditions shown to preclude deterioration of the gel, *e.g.* drying, microbial growth, shifting, cracking, which would degrade its performance. The acceptable storage period should be determined and membrane barrier preparations not used after that period.
16. The permeable supporting membrane provides mechanical support to the proteinaceous gel during the gelling process and exposure to the test chemical. The supporting membrane should prevent sagging or shifting of the gel and be readily permeable to all test chemicals.
17. The proteinaceous gel, composed of protein, *e.g.* keratin, collagen, or mixtures of proteins, forming a gel matrix, serves as the target for the test chemical. The proteinaceous material is placed on the surface of the supporting membrane and allowed to gel prior to placing the membrane barrier over the indicator solution. The proteinaceous gel should be of equal thickness and density throughout, and with no air bubbles or defects that could affect its functional integrity.

Chemical Detection System (CDS)

18. The indicator solution, which is the same solution used for the compatibility test, should respond to the presence of a test chemical. A pH indicator dye or combination of dyes, *e.g.* cresol red and methyl orange that will show a colour change, in response to the presence of the test chemical, should be used. The measurement system can be visual or electronic.
19. Detection systems that are developed for detecting the passage of the test chemical through the barrier membrane should be assessed for their relevance and reliability in order to demonstrate the range of chemicals that can be detected and the quantitative limits of detection.

TEST PERFORMANCE

Assembly of the Test Method Components

20. The membrane barrier is positioned in a vial (or tube) containing the indicator solution so that the supporting membrane is in full contact with the indicator solution and with no air bubbles present. Care should be taken to ensure that barrier integrity is maintained.

Application of the Test Chemical

21. A suitable amount of the test chemical, *e.g.* 500 µl of a liquid or 500 mg of a finely powdered solid (7), is carefully layered onto the upper surface of the membrane barrier and evenly distributed. An appropriate number of replicates, *e.g.* four (7), is prepared for each test chemical and its corresponding controls (see paragraphs 23 to 25). The time of applying the test chemical to the membrane barrier is recorded. To ensure that short corrosion times are accurately recorded, the application times of the test chemical to the replicate vials are staggered.

Measurement of Membrane Barrier Penetrations

22. Each vial is appropriately monitored and the time of the first change in the indicator solution, *i.e.* barrier penetration, is recorded, and the elapsed time between application and penetration of the membrane barrier determined.

Controls

23. In tests that involve the use of a vehicle or solvent with the test chemical, the vehicle or solvent should be compatible with the membrane barrier system, *i.e.* not alter the integrity of the membrane barrier system, and should not alter the corrosivity of the test chemical. When applicable, solvent (or vehicle) control should be tested concurrently with the test chemical to demonstrate the compatibility of the solvent with the membrane barrier system.
24. A positive (corrosive) control with intermediate corrosivity activity, *e.g.* 110 ± 15 mg sodium hydroxide (UN GHS Corrosive Sub-category 1B) (7), should be tested concurrently with the test chemical to assess if the test system is performing in an acceptable manner. A second positive control that is of the same chemical class as the test chemical may be useful for evaluating the relative corrosivity potential of a corrosive test chemical. Positive control(s) should be selected that are intermediate in their corrosivity (*e.g.* UN GHS Sub-category 1B) in order to detect changes in the penetration time that may be unacceptably longer or shorter than the established reference value, thereby indicating that the test system is not functioning properly. For this purpose, extremely corrosive (UN GHS Sub-category 1A) or non-corrosive chemicals are of limited utility. A corrosive UN GHS Sub-category 1B chemical would allow detection of a too rapid or too slow breakthrough time. A weakly corrosive (UN GHS Sub-category 1C) might be employed as a positive control to measure the ability of the test method to consistently distinguish between weakly corrosive and non-corrosive chemicals. Regardless of the approach used, an acceptable positive control response range should be developed based on the historical range of breakthrough times for the positive control(s) employed, such as the mean ± 2 -3 standard deviations. In each study, the exact breakthrough time should be determined for the positive control so that deviations outside the acceptable range can be detected.
25. A negative (non-corrosive) control, *e.g.* 10 % citric acid, 6 % propionic acid (7), should also be tested concurrently with the test chemical as another quality control measure to demonstrate the functional integrity of the membrane barrier.

Study Acceptability Criteria

26. According to the established time parameters for each of the UN GHS corrosivity Sub-categories, the time (in minutes) elapsed between application of a test chemical to the membrane barrier and barrier penetration is used to predict the corrosivity of the test chemical. For a study to be considered acceptable, the concurrent positive control should give the expected penetration response time (*e.g.* 8-16 min breakthrough time for sodium hydroxide if used as a positive control), the concurrent negative control should not be corrosive, and, when included, the concurrent solvent control should neither be corrosive nor should it alter the corrosivity potential of the test chemical. Prior to routine use of a method that adheres to this test method, laboratories should demonstrate technical proficiency, using the twelve substances recommended in Table 2. For new "me-too" methods developed under this test method that are structurally and functionally similar to the validated reference method (14) the pre-defined performance standards should be used to demonstrate the reliability and accuracy of the new method prior to its use for regulatory testing (10).

Interpretation of Results and Corrosivity Classification of Test Chemicals

27. The time (in minutes) elapsed between application of the test chemical to the membrane barrier and barrier penetration is used to classify the test chemical in terms of UN GHS corrosive Sub-categories (1) and, if applicable, UN Packing Group (16). Cut-off time values for each of the three corrosive subcategories are established for each proposed test method. Final decisions on cut-off times should consider the need to minimise under-classification of corrosive hazard (*i.e.* false negatives). In the present test guideline, the cut-off times of Corrositex® as described in table 3 should be used as it represents the only test method currently falling within the test guideline (7).

Table 3
Corrositex[®] prediction model

Mean breakthrough time (min.)		UN GHS prediction ⁽³⁾
Category 1 test chemicals ⁽¹⁾ (determined by the method's categorisation test)	Category 2 test chemicals ⁽²⁾ (determined by the method's categorisation test)	
0-3 min.	0-3 min.	Corrosiveoptional Sub-category 1A
> 3 to 60 min.	> 3 to 30 min.	Corrosiveoptional Sub-category 1B
> 60 to 240 min.	> 30 to 60 min.	Corrosiveoptional Sub-category 1C
> 240 min.	> 60 min.	Non-corrosive

⁽¹⁾ Test chemicals with high acid/alkaline reserve (6)

⁽²⁾ Test chemicals with low acid/alkaline reserve (6)

⁽³⁾ UN GHS Subcategories 1A, 1B and 1C correspond to UN packing groups I, II and III respectively

DATA AND REPORTING

Data

28. The time (in minutes) elapsed between application and barrier penetration for the test chemical and the positive control(s) should be reported in tabular form as individual replicate data, as well as means \pm the standard deviation for each trial.

Test Report

29. The test report should include the following information:

Test Chemical and Control Substances:

- Mono-constituent substance: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc;
- Multi-constituent substance, UVCB and mixture: characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents;
- Physical appearance, water solubility, and additional relevant physicochemical properties;
- Source, lot number if available;
- Treatment of the test chemical/control substance prior to testing, if applicable (e.g. warming, grinding);
- Stability of the test chemical, limit date for use, or date for re-analysis if known;
- Storage conditions.

Vehicle:

- Identification, concentration (where appropriate), volume used;
- Justification for choice of vehicle.

In vitro membrane barrier model and protocol used, including demonstrated accuracy and reliability

Test Conditions:

- Description of the apparatus and preparation procedures used;
- Source and composition of the *in vitro* membrane barrier used;
- Composition and properties of the indicator solution;
- Method of detection;
- Test chemical and control substance amounts;
- Number of replicates;
- Description and justification for the timescale categorisation test;
- Method of application;
- Observation times.
- Description of the evaluation and classification criteria applied;
- Demonstration of proficiency in performing the test method before routine use by testing of the proficiency chemicals.

Results:

- Tabulation of individual raw data from individual test and control samples for each replicate;
- Descriptions of other effects observed;
- The derived classification with reference to the prediction model/decision criteria used.

*Discussion of the results**Conclusions***LITERATURE**

- (1) United Nations (UN) (2013). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Fifth Revised Edition, UN New York and Geneva, 2013. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html
- (2) Chapter B.4 of this Annex, Acute Dermal Irritation, Corrosion.
- (3) Chapter B.40bis of this Annex, *In vitro* skin corrosion: reconstructed human epidermis (RHE) test method.

- (4) Chapter B.40 of this Annex, *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance (TER).
- (5) OECD (2015). Guidance Document on Integrated Approaches to Testing and Assessment of Skin Irritation/Corrosion. Environment, Health and Safety Publications, Series on Testing and Assessment, (No 203). Organisation for Economic Cooperation and Development, Paris.
- (6) Fentem, J.H., Archer, G.E.B., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., Holzhutter, H.-G. and Liebsch, M. (1998). The ECVAM International Validation Study on *In Vitro* Tests for Skin Corrosivity. 2. Results and Evaluation by the Management Team. *Toxicology In Vitro* 12, 483-524.
- (7) ICCVAM (1999). Corrositex®. An *In Vitro* Test Method for Assessing Dermal Corrosivity Potential of Chemicals. The Results of an Independent Peer Review Evaluation Coordinated by ICCVAM, NTP and NICEATM. NIEHS, NIH Publication (No 99-4495).
- (8) Gordon V.C., Harvell J.D. and Maibach H.I. (1994). Dermal Corrosion, the Corrositex® System: A DOT Accepted Method to Predict Corrosivity Potential of Test Materials. *In vitro* Skin Toxicology-Irritation, Photo-toxicity, Sensitization. *Alternative Methods in Toxicology* 10, 37-45.
- (9) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environmental, Health and Safety Publications. Series on testing and Assessment (No 34).
- (10) OECD (2014). Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Membrane Barrier Test Method for Skin Corrosion in Relation to TG 435. Organisation for Economic Cooperation and Development, Paris. Available at: <http://www.oecd.org/chemicalsafety/testing/PerfStand-TG430-June14.pdf>.
- (11) ECVAM (2001). Statement on the Application of the CORROSITEX Assay for Skin Corrosivity Testing. 15th Meeting of ECVAM Scientific Advisory Committee (ESAC), Ispra, Italy. ATLA 29, 96-97.
- (12) U.S. DOT (2002). Exemption DOT-E-10904 (Fifth Revision). (September 20, 2002). Washington, D.C., U.S. DOT.
- (13) Chapter B.46 of this Annex, *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method. ICCVAM (2004). ICCVAM Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion. NIEHS, NIH Publication No 04-4510. Available at: http://www.ntp.niehs.nih.gov/iccvam/docs/dermal_docs/ps/ps044510.pdf.
- (14) U.S. EPA (1996). Method 1120, Dermal Corrosion. Available at: <http://www.epa.gov/osw/hazard/testmethods/sw846/pdfs/1120.pdf>.
- (15) United Nations (UN) (2013). UN Recommendations on the Transport of Dangerous Goods, Model Regulations, 18th Revised Edition (Part, Chapter 2.8), UN, 2013. Available at: http://www.unece.org/fileadmin/DAM/trans/danger/publi/unrec/rev18/English/Rev18_Volume1_Part2.pdf.

Appendix

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with “concordance” to mean the proportion of correct outcomes of a test method (9).

Chemical: A substance or a mixture.

Chemical Detection System (CDS): A visual or electronic measurement system with an indicator solution that responds to the presence of a test chemical, *e.g.* by a change in a pH indicator dye, or combination of dyes, that will show a colour change in response to the presence of the test chemical or by other types of chemical or electrochemical reactions.

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (9).

GHS (Globally Harmonized System of Classification and Labelling of Chemicals): A system proposing the classification of chemicals (substances and mixtures) according to standardised types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

IATA: Integrated Approach on Testing and Assessment.

Mixture: A mixture or solution composed of two or more substances.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration ≥ 10 % (w/w) and < 80 % (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

NC: Non corrosive.

Performance standards: Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of reliability and accuracy, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (9).

Relevance: Description of relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test method correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (9).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility (9).

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (9).

Skin corrosion *in vivo*: The production of irreversible damage of the skin; namely, visible necrosis through the *epidermis* and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (9).

Substance: A chemical element and its compounds in the natural state or obtained by any production process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Test chemical: Any substance or mixture tested using this test method.

UVCB: Substances of unknown or variable composition, complex reaction products or biological materials.

B.66 STABLY TRANSFECTED TRANSACTIVATION IN VITRO ASSAYS TO DETECT ESTROGEN RECEPTOR AGONISTS AND ANTAGONISTS

GENERAL INTRODUCTION

OECD Performance-Based Test Guideline

1. This test method is equivalent to OECD test guideline (TG) 455 (2016). TG 455 is a performance-based test guideline (PBTG), describing the methodology of Stably Transfected Transactivation *In Vitro* Assays to detect Estrogen Receptor Agonists and Antagonists (ER TA assays). It comprises several mechanistically and functionally similar test methods for the identification of estrogen receptor (i.e. ER α , and/or ER α) agonists and antagonists and should facilitate the development of new similar or modified test methods in accordance with the principles for validation set forth in the OECD Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (1). The fully validated reference test methods (Appendix 2 and Appendix 3) that provide the basis for this PBTG are:
 - The Stably Transfected TA (STTA) assay (2) using the (h) ER α -HeLa-9903 cell line; and
 - The VM7Luc ER TA assay (3) using the VM7Luc4E2 cell line (1) which predominately expresses hER α with some contribution from hER (4)(5).

For the development and validation of similar assays for the same hazard endpoint, performance standards (PS) (6) (7) are available and should be used. They allow for timely amendment of PBTG 455 so that new similar assays can be added to an updated PBTG; however, similar assays will only be added after review and agreement by OECD that performance standards are met. The assays included in TG 455 can be used indiscriminately to address OECD member countries' requirements for test results on estrogen receptor transactivation while benefiting from the OECD Mutual Acceptance of Data.

Background and principles of the assays included in this test method

2. The OECD initiated a high-priority activity in 1998 to revise existing, and to develop new test guidelines for the screening and testing of potential endocrine disrupting chemicals. The OECD conceptual framework (CF) for testing and assessment of potential endocrine disrupting chemicals was revised in 2012. The original and revised CFs are included as Annexes in the OECD Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption (8). The CF comprises five levels, each level corresponding to a different level of biological complexity. The ER Transactivation (TA) assays described in this test method are level 2, which includes *in vitro* assays providing data about selected endocrine mechanism(s)/pathway(s). This test method is for *in vitro* Transactivation (TA) assays designed to identify estrogen receptor (ER) agonists and antagonists.
3. The interaction of estrogens with ERs can affect transcription of estrogen-controlled genes, which can lead to the induction or inhibition of cellular processes, including those necessary for cell proliferation, normal fetal development, and reproductive function (9)(10)(11). Perturbation of normal estrogenic systems may have the potential to trigger adverse effects on normal development (ontogenesis), reproductive health and the integrity of the reproductive system.
4. *In vitro* TA assays are based on a direct or indirect interaction of the substances with a specific receptor that regulates the transcription of a reporter gene product. Such assays have been used extensively to evaluate gene expression regulated by specific nuclear receptors, such as ERs (12) (13) (14) (15) (16). They have been proposed for the detection of estrogenic transactivation regulated by the ER (17) (18) (19). There are at least two major subtypes of nuclear ERs, α and β , which are encoded by distinct genes. The respective proteins have different biological functions as well as different tissue distributions and ligand binding affinities (20)(21)(22)(23)(24)(25)(26). Nuclear ER α mediates the classic estrogenic response (27)(28)(29)(30), and therefore most models currently being developed to measure ER activation or inhibition are specific to ER α . The assays are used to identify chemicals that activate (or inhibit) the ER following ligand binding, after which the receptor-ligand complex binds to specific DNA response

elements and transactivates a reporter gene, resulting in increased cellular expression of a marker protein. Different reporter responses can be used in these assays. In luciferase based systems, the luciferase enzyme transforms the luciferin substrate to a bioluminescent product that can be quantitatively measured with a luminometer. Other examples of common reporters are fluorescent protein and the *LacZ* gene, which encodes β -galactosidase, an enzyme that can transform the colourless substrate X-gal (5-bromo-4-chloro-indolyl-galactopyranoside) into a blue product that can be quantified with a spectrophotometer. These reporters can be evaluated quickly and inexpensively with commercially available test kits.

- Validation studies of the STTA and the VM7Luc TA assays have demonstrated their relevance and reliability for their intended purpose (3)(4)(5)(30). Performance standards for luminescence-based ER TA assays using breast cells lines are included in ICCVAM Test Method Evaluation Report on the LUMI-CELL® ER (VM7Luc ER TA) Test Method: An *In Vitro* Assay for Identifying Human Estrogen Receptor Agonist and Antagonist Activity of Chemicals (3). These performance standards have been modified to be applicable to both the STTA and VM7Luc TA assays (2).
- Definitions and abbreviations used in this test method are described in Appendix 1.

Scope and limitations related to the TA assays

- These assays are being proposed for screening and prioritisation purposes, but can also provide mechanistic information that can be used in a weight of evidence approach. They address TA induced by chemical binding to the ERs in an *in vitro* system. Thus, results should not be directly extrapolated to the complex signalling and regulation of the intact endocrine system *in vivo*.
- TA mediated by the ERs is considered one of the key mechanisms of endocrine disruption (ED), although there are other mechanisms through which ED can occur, including (i) interactions with other receptors and enzymatic systems within the endocrine system, (ii) hormone synthesis, (iii) metabolic activation and/or inactivation of hormones, (iv) distribution of hormones to target tissues, and (v) clearance of hormones from the body. None of the assays under this test method addresses these modes of action.
- This test method addresses the ability of chemicals to activate (i.e. act as agonists) and also to suppress (i.e. act as antagonists) ER-dependent transcription. Some chemicals may, in a cell type-dependent manner, display both agonist and antagonist activity and are known as selective estrogen receptor modulators (SERMs). Chemicals that are negative in these assays could be evaluated in an ER binding assay before concluding that the chemical does not bind to the receptor. In addition, the assays are only likely to inform on the activity of the parent molecule bearing in mind the limited metabolising capacities of the *in vitro* cell systems. Considering that only single substances were used during the validation, the applicability to test mixtures has not been addressed. The test method is nevertheless theoretically applicable to the testing of multi-constituent substances, UVCBs and mixtures. Before use of the test method on a multi-constituent substance, UVCB or mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture.
- For informational purposes, Table 1 provides the agonist test results for the 34 substances that were tested in both of the fully validated reference test methods described in this test method. Of these substances, 26 are classified as definitive ER agonists and 8 negatives based upon published reports, including *in vitro* assays for ER binding and TA, and/or the uterotrophic assay (2)(3)(18)(31)(32)(33)(34). Table 2 provides the antagonist test results for the 15 substances that were tested in both of the fully validated reference test methods described in this test method. Of these substances, 4 are classified as definitive/presumed ER antagonists and 10 negatives based upon published reports, including *in vitro* assays for ER binding and TA (2)(3)(18)(31). In reference to the data summarised in Table 1 and Table 2, there was 100 % agreement between the two reference test methods on the classifications of all the substances except for one substance (Mifepristone) for antagonist assay, and each substance was correctly classified as an ER agonist/antagonist or negative. Supplementary information on this group of chemicals as well as additional chemicals tested in the STTA and VM7Luc ER TA assays during the validation studies is provided in the Performance Standards for the ERTA (6)(7), Appendix 2 (Tables 1, 2 and 3).

Table 1
Overview of the Results from STTA and VM7Luc ER TA Assays for Substances Tested in Both Agonist Assays and Classified as ER Agonists (POS) or Negatives (NEG)

	Substance	CASRN	STTA Assay ⁽¹⁾			VM7Luc ER TA Assay ⁽²⁾		Data Source For Classification ⁽⁴⁾		
			ER TA Activity	PC ₁₀ Value (M)	PC ₅₀ Value ^(b) (M)	ER TA Activity	EC ₅₀ Value ^{(b), (3)} (M)	Other ER TAs ^(c)	ER Binding	Uterotrophic
1	17β-estradiol ^(a)	50-28-2	POS	<1,00 × 10 ⁻¹¹	<1,00 × 10 ⁻¹¹	POS	5,63 × 10 ⁻¹²	POS (22/7/22/7)	POS	POS
2	17α-estradiol ^(a)	57-91-0	POS	7,24 × 10 ⁻¹¹	6,44 × 10 ⁻¹⁰	POS	1,40 × 10 ⁻⁹	POS(11/11)	POS	POS
3	17α-ethinyl estradiol ^(a)	57-63-6	POS	<1,00 × 10 ⁻¹¹	<1,00 × 10 ⁻¹¹	POS	7,31 × 10 ⁻¹²	POS(22/22)	POS	POS
4	17β-trenbolone	10161-33-8	POS	1,78 × 10 ⁻⁸	2,73 × 10 ⁻⁷	POS	4,20 × 10 ⁻⁸	POS (2/2)	NT	NT
5	19-nortestosterone ^(a)	434-22-0	POS	9,64 × 10 ⁻⁹	2,71 × 10 ⁻⁷	POS	1,80 × 10 ⁻⁶	POS(4/4)	POS	POS
6	4-cumylphenol ^(a)	599-64-4	POS	1,49 × 10 ⁻⁷	1,60 × 10 ⁻⁶	POS	3,20 × 10 ⁻⁷	POS(5/5)	POS	NT
7	4-tert-octylphenol ^(a)	140-66-9	POS	1,85 × 10 ⁻⁹	7,37 × 10 ⁻⁸	POS	3,19 × 10 ⁻⁸	POS(21/24)	POS	POS
8	Apigenin ^(a)	520-36-5	POS	1,31 × 10 ⁻⁷	5,71 × 10 ⁻⁷	POS	1,60 × 10 ⁻⁶	POS(26/26)	POS	NT

	Substance	CASRN	STTA Assay ⁽¹⁾			VM7Luc ER TA Assay ⁽²⁾		Data Source For Classification ⁽⁴⁾		
			ER TA Activity	PC ₁₀ Value (M)	PC ₅₀ Value ⁽⁶⁾ (M)	ER TA Activity	EC ₅₀ Value ⁽⁶⁾ , ⁽⁷⁾ (M)	Other ER TAs ⁽⁶⁾	ER Binding	Uterotrophic
9	Atrazine ^(a)	1912-24-9	NEG	—	—	NEG	—	NEG (30/30)	NEG	NT
10	Bisphenol A ^(a)	80-05-7	POS	2,02 × 10 ⁻⁸	2,94 × 10 ⁻⁷	POS	5,33 × 10 ⁻⁷	POS(65/65)	POS	POS
11	Bisphenol B ^(a)	77-40-7	POS	2,36 × 10 ⁻⁸	2,11 × 10 ⁻⁷	POS	1,95 × 10 ⁻⁷	POS(6/6)	POS	POS
12	Butylbenzyl phthalate ^(a)	85-68-7	POS	1,14 × 10 ⁻⁶	4,11 × 10 ⁻⁶	POS	1,98 × 10 ⁻⁶	POS(12/14)	POS	NEG
13	Corticosterone ^(a)	50-22-6	NEG	—	—	NEG	—	NEG(6/6)	NEG	NT
14	Coumestrol ^(a)	479-13-0	POS	1,23 × 10 ⁻⁹	2,00 × 10 ⁻⁸	POS	1,32 × 10 ⁻⁷	POS(30/30)	POS	NT
15	Daidzein ^(a)	486-66-8	POS	1,76 × 10 ⁻⁸	1,51 × 10 ⁻⁷	POS	7,95 × 10 ⁻⁷	POS(39/39)	POS	POS
16	Diethylstilbestrol ^(a)	56-53-1	POS	<1,00 × 10 ⁻¹¹	2,04 × 10 ⁻¹¹	POS	3,34 × 10 ⁻¹¹	POS(42/42)	POS	NT
17	Di-n-butyl phthalate	84-74-2	POS	4,09 × 10 ⁻⁶	—	POS	4,09 × 10 ⁻⁶	POS(6/11)	POS	NEG
18	Ethyl paraben	120-47-8	POS	5,00 × 10 ⁻⁶	(no PC ₅₀)	POS	2,48 × 10 ⁻⁵	POS	—	NT
19	Estrone ^(a)	53-16-7	POS	3,02 × 10 ⁻¹¹	5,88 × 10 ⁻¹⁰	POS	2,34 × 10 ⁻¹⁰	POS(26/28)	POS	POS
20	Genistein ^(a)	446-72-0	POS	2,24 × 10 ⁻⁹	2,45 × 10 ⁻⁸	POS	2,71 × 10 ⁻⁷	POS(100/102)	POS	POS

	Substance	CASRN	STTA Assay (1)			VM7Luc ER TA Assay (2)		Data Source For Classification (4)		
			ER TA Activity	PC ₁₀ Value (M)	PC ₅₀ Value (6)	ER TA Activity	EC ₅₀ Value (6), (2)	Other ER TAs (6)	ER Binding	Uterotrophic
21	Haloperidol	52-86-8	NEG	—	—	NEG	—	NEG (2/2)	NEG	NT
22	Kaempferol (6)	520-18-3	POS	$1,36 \times 10^{-7}$	$1,21 \times 10^{-6}$	POS	$3,99 \times 10^{-6}$	POS(23/23)	POS	NT
23	Kepone (4)	143-50-0	POS	$7,11 \times 10^{-7}$	$7,68 \times 10^{-6}$	POS	$4,91 \times 10^{-7}$	POS(14/18)	POS	NT
24	Ketoconazole	65277-42-1	NEG	—	—	NEG	—	NEG (2/2)	NEG	NT
25	Linuron (4)	330-55-2	NEG	—	—	NEG	—	NEG (8/8)	NEG	NT
26	meso-Hexestrol (4)	84-16-2	POS	$<1,00 \times 10^{-11}$	$2,75 \times 10^{-11}$	POS	$1,65 \times 10^{-11}$	POS(4/4)	POS	NT
27	Methyl testosterone (4)	58-18-4	POS	$1,73 \times 10^{-7}$	$4,11 \times 10^{-6}$	POS	$2,68 \times 10^{-6}$	POS(5/6)	POS	NT
28	Morin	480-16-0	POS	$5,43 \times 10^{-7}$	$4,16 \times 10^{-6}$	POS	$2,37 \times 10^{-6}$	POS(2/2)	POS	NT
29	Norethynodrel (4)	68-23-5	POS	$1,11 \times 10^{-11}$	$1,50 \times 10^{-9}$	POS	$9,39 \times 10^{-10}$	POS(5/5)	POS	NT
30	p-p'-Methoxychlor (4)	72-43-5	POS	$1,23 \times 10^{-6}$	(no PC ₅₀) (6)	POS	$1,92 \times 10^{-6}$	POS(24/27)	POS	POS
31	Phenobarbital (4)	57-30-7	NEG	—	—	NEG	—	NEG(2/2)	NEG	NT

	Substance	CASRN	STTA Assay (1)			VM7Luc ER TA Assay (2)		Data Source For Classification (4)		
			ER TA Activity	PC ₁₀ Value (M)	PC ₅₀ Value (6)	ER TA Activity	EC ₅₀ Value (6), (3) (M)	Other ER TAs (6)	ER Binding	Uterotrophic
32	Reserpine	50-55-5	NEG	—	—	NEG	—	NEG(4/4)	NEG	NT
33	Spironolactone (4)	52-01-7	NEG	—	—	NEG	—	NEG(4/4)	NEG	NT
34	Testosterone	58-22-0	POS	$2,82 \times 10^{-8}$	$9,78 \times 10^{-6}$	POS	$1,75 \times 10^{-5}$	POS(5/10)	POS	NT

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; M = molar; EC₅₀ = half maximal effective concentration of test substance; NEG = negative; POS = positive; NT = Not tested; PC₁₀ (and PC₅₀) = the concentration of a test substance at which the response is 10 % (or 50 % for PC₅₀) of the response induced by the positive control (E2, 1nM) in each plate.

(6) Common substances tested in the STTA and VM7Luc ER TA assays that were designated as ER agonists or negatives and used to evaluate accuracy in the VM7Luc ER TA validation study (ICCVAM VM7Luc ER TA Evaluation Report, Table 4-1 (3)).

(6) Maximum concentration tested in the absence of limitations due to cytotoxicity or insolubility was 1×10^{-5} M (STTA Assay) and 1×10^{-3} M (VM7Luc ER TA Assay).

(6) Number in parenthesis represents the test results classified as positive (POS) or negative (NEG) over the total number of referenced studies.

(1) Values reported in Draft Report of Pre-validation and Inter-laboratory Validation For Stably Transfected Transcriptional Activation (TA) Assay to Detect Estrogenic Activity - The Human Estrogen Receptor Alpha Mediated Reporter Gene Assay Using hER-HeLa-9903 Cell Line (2)

(2) ICCVAM Test Method Evaluation Report on the LUMI-CELL® ER (VM7Luc ER TA) Test Method: An *In Vitro* Method for Identifying ER Agonists and Antagonists (3)

(3) Mean EC₅₀ values were calculated with values reported by the laboratories of the VM7Luc ER TA validation study (XDS, ECVAM, and Hiyoshi) (3).

(4) Classification as an ER agonist or negative was based upon information in the ICCVAM Background Review Documents (BRD) for ER Binding and TA test methods (31) as well as information obtained from publications published and reviewed after the completion of the ICCVAM BRDs (2) (3) (18) (31) (33) (34).

Notes: Each assay within this test method does not have the same measurements. In some situations the EC₅₀ cannot be calculated because a full dose response curve is not generated. Whilst with the STTA assay, the PC₁₀ value is a key measurement, there may also be further examples where a PCx will provide useful information.

Table 2
Comparison of Results from STTA and VM7Luc ER TA Assays for Substances Tested in Both Antagonist Assays and Classified as ER Antagonists (POS) or Negatives (NEG)

	Substance ^(a)	CASRN	ER STTA assay ⁽¹⁾		VM7Luc ER TA assay ⁽²⁾		ER STTA candidate effects ⁽⁴⁾	ICCVAM ⁽⁵⁾ Consensus Classification	MeSH ⁽⁶⁾ Chemical Class	Product Class ⁽⁷⁾
			ER TAActivity	IC ₅₀ Value ^(b) (M)	ER TAActivity	IC ₅₀ Value ^(b) , ^(c) (M)				
1	4-hydroxytamoxifen	68047-06-3	POS	$3,97 \times 10^{-9}$	POS	$2,08 \times 10^{-7}$	moderate POS	POS	Hydrocarbon (Cyclic)	Pharmaceutical
2	Dibenzo[a,h]anthracene	53-70-3	POS	No IC ₅₀	POS	No IC ₅₀	POS	PP	Polycyclic Compound	Laboratory Chemical, Natural Product
3	Mifepristone	84371-65-3	POS	$5,61 \times 10^{-6}$	NEG	—	mild POS	NEG	Steroid	Pharmaceutical
4	Raloxifene HCl	82640-04-8	POS	$7,86 \times 10^{-10}$	POS	$1,19 \times 10^{-9}$	moderate POS	POS	Hydrocarbon (Cyclic)	Pharmaceutical
5	Tamoxifen	10540-29-1	POS	$4,91 \times 10^{-7}$	POS	$8,17 \times 10^{-7}$	POS	POS	Hydrocarbon (Cyclic)	Pharmaceutical
6	17β-estradiol	50-28-2	NEG	—	NEG	—	PN	PN	Steroid	Pharmaceutical, Veterinary Agent
7	Apigenin	520-36-5	NEG	—	NEG	—	NEG	NEG	Heterocyclic Compound	Dye, Natural Product, Pharmaceutical Intermediate

	Substance ^(a)	CASRN	ER STTA assay ⁽¹⁾		VM7Luc ER TA assay ⁽²⁾		ER STT Acandida- teeffects ⁽⁴⁾	ICCVAM ⁽⁵⁾ Consensus Clas- sification	MeSH ⁽⁶⁾ Chemical Class	Product Class ⁽⁷⁾
			ER TAAactivity	IC ₅₀ Value ^(b) (M)	ER TAAactivity	IC ₅₀ Value ^(b) , ^(c) (M)				
8	Atrazine	1912-24-9	NEG	—	NEG	—	NEG	PN	Heterocyclic Compound	Herbicide
9	Di-n-butyl phthalate	84-74-2	NEG	—	NEG	—	NEG	NEG	Ester, Phthalic Acid	Cosmetic Ingredient, Industrial Chemical, Plas- ticiser
10	Fenarimol	60168-88-9	NEG	—	NEG	—	not tested	PN	Heterocyclic Compound, Pyrimidine	Fungicide
11	Flavone	525-82-6	NEG	—	NEG	—	PN	PN	Flavonoid, Heterocyclic Compound	Natural Product, Phar- maceutical
12	Flutamide	13311-84-7	NEG	—	NEG	—	NEG	PN	Amide	Pharmaceutical, Veterinary Agent
13	Genistein	446-72-0	NEG	—	NEG	—	PN	NEG	Flavonoid, Heterocyclic Compound	Natural Product, Phar- maceutical
14	p-n-nonylphenol	104-40-5	NEG	—	NEG	—	not tested	NEG	Phenol	Chemical Inter- mediate

Substance ⁽⁴⁾	CASRN	ER STTA assay ⁽¹⁾		VM7Luc ER TA assay ⁽²⁾		ER STTAcandida- teeffects ⁽⁴⁾	ICCVAM ⁽⁵⁾ Consensus Clas- sification	MeSH ⁽⁶⁾ Chemical Class	Product Class ⁽⁷⁾
		ER TAAactivity	IC ₅₀ Value ⁽⁶⁾ (M)	ER TAAactivity	IC ₅₀ Value ⁽⁶⁾ , ⁽⁷⁾ (M)				
15	Resveratrol	501-36-0	NEG	NEG	—	PN	NEG	Hydrocarbon (Cyclic)	Natural Product

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; M = molar; IC₅₀ = half maximal inhibitory concentration of test substance; NEG = negative; PN = presumed negative; POS = positive; PP = presumed positive.

⁽⁴⁾ Common substances tested in the STTA and VM7Luc ER TA assays that were designated as ER antagonists or negatives and used to evaluate accuracy in the VM7Luc ER TA validation study (2) (3).
⁽⁵⁾ Maximum concentration tested in the absence of limitations due to cytotoxicity or insolubility was 1×10^{-5} M (STTA Assay) and 1×10^{-5} M (VM7Luc ER TA Assay).
⁽⁶⁾ The Validation Report of the Stably transfected Transcriptional Activation Assay to Detect ER mediated activity, Part B (2).
⁽⁷⁾ ICCVAM Test Method Evaluation Report on the LUMI-CELL@ER (VM7Luc ER TA) Test Method: An *In Vitro* Method for Identifying ER Agonists and Antagonists (3).
⁽⁸⁾ Mean IC₅₀ values were calculated with values reported by the laboratories of the VM7Luc ER TA validation study (XDS, ECVAM, and Hiyoshi) (3).
⁽⁹⁾ ER STTA activity assumed from their reported effects known from the CERi historical data of ER receptor binding assay, the uterotrophic assay and information collated from the open literature (2).
⁽¹⁰⁾ Classification as an ER antagonist or negative was based upon information in the ICCVAM Background Review Documents (BRD) for ER Binding and TA assays (31) as well as information obtained from publications published and reviewed after the completion of the ICCVAM BRDs (2) (3) (18) (31).
⁽¹¹⁾ Substances were assigned to one or more chemical classes using the U.S. National Library of Medicine's Medical Subject Headings (MeSH), an internationally recognised standardised classification scheme (available at <http://www.nlm.nih.gov/mesh>).
⁽¹²⁾ Substances were assigned to one or more product classes using the U.S. National Library of Medicine's Hazardous Substances Data Bank (available at <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDDB>).

ER TA ASSAY COMPONENTS

Essential Assay Components

11. This test method applies to assays using a stably transfected or endogenous ER α receptor and stably transfected reporter gene construct under the control of one or more estrogen response elements; however, other receptors such as ER β may be present. These are essential assay components.

Controls

12. The basis for the proposed concurrent reference standards for each of agonist and antagonist assay should be described. Concurrent controls (negative, solvent, and positive), as appropriate, serve as an indication that the assay is operative under the test conditions and provide a basis for experiment-to-experiment comparisons; they are usually part of the acceptability criteria for a given experiment (1).

Standard Quality Control Procedures

13. Standard quality control procedures should be performed as described for each assay to ensure the cell line remains stable through multiple passages, remains mycoplasma-free (i.e. free of bacterial contamination), and retains the ability to provide the expected ER-mediated responses over time. Cell lines should be further checked for their correct identity as well as for other contaminants (e.g. fungi, yeast and viruses).

Demonstration of Laboratory Proficiency

14. Prior to testing unknown chemicals with any of the assays under this test method, each laboratory should demonstrate proficiency in using the assay. To demonstrate proficiency, each laboratory should test the 14 proficiency substances listed in Table 3 for the agonist assay and 10 proficiency substances in Table 4 for the antagonist assay. This proficiency testing will also confirm the responsiveness of the test system. The list of proficiency substances is a subset of the reference substances provided in the Performance Standards for the ER TA assays (6). These substances are commercially available, represent the classes of chemicals commonly associated with ER agonist or antagonist activity, exhibit a suitable range of potency expected for ER agonists/antagonists (i.e. strong to weak) and include negatives. Testing of the proficiency substances should be replicated at least twice, on different days. Proficiency is demonstrated by correct classification (positive/negative) of each proficiency substance. Proficiency testing should be repeated by each technician when learning the assays. Dependent on cell type, some of these proficiency substances may behave as SERMs and display activity as both agonists and antagonists. However, the proficiency substances are classified in Tables 3 and 4 by their known predominant activity which should be used for proficiency evaluation.
15. To demonstrate performance and for quality control purposes each laboratory should compile agonist and antagonist historical databases with reference standard (e.g. 17 β -estradiol and tamoxifen), positive and negative control chemicals and solvent control (e.g. DMSO) data. As a start, the database should be generated from at least 10 independent agonist (e.g. 17 β -estradiol) and 10 independent antagonist (e.g. tamoxifen) runs. Results from future analyses of these reference standards and solvent controls should be added to enlarge the database to ensure consistency and performance of the bioassay by the laboratory over time.

Table 3
List of (14) Proficiency Substances for agonist assay ⁽⁸⁾

N° ⁽⁷⁾	Substance	CASRN	Expected Response ⁽¹⁾	STTA Assay			VM7Luc ER TA Assay		MeSH Chemical Class ⁽⁵⁾	Product Class ⁽⁶⁾
				PC ₁₀ Value (M) ⁽²⁾	PC ₅₀ Value (M) ⁽²⁾	Test Conc. Range (M)	VM7Luc EC ₅₀ Value (M) ⁽³⁾	Highest Conc. for Range Finder (M) ⁽⁴⁾		
14	Diethylstilbestrol	56-53-1	POS	<1,00 × 10 ⁻¹¹	2,04 × 10 ⁻¹¹	10 ⁻¹⁴ – 10 ⁻⁸	3,34 × 10 ⁻¹¹	3,73 × 10 ⁻⁴	Hydrocarbon (Cyclic)	Pharmaceutical, Veterinary Agent
12	17α-estradiol	57-91-0	POS	4,27 × 10 ⁻¹¹	6,44 × 10 ⁻¹⁰	10 ⁻¹¹ – 10 ⁻⁵	1,40 × 10 ⁻⁹	3,67 × 10 ⁻³	Steroid	Pharmaceutical, Veterinary Agent
15	meso-Hexestrol	84-16-2	POS	<1,00 × 10 ⁻¹¹	2,75 × 10 ⁻¹¹	10 ⁻¹¹ – 10 ⁻⁵	1,65 × 10 ⁻¹¹	3,70 × 10 ⁻³	Hydrocarbon (Cyclic), Phenol	Pharmaceutical, Veterinary Agent
11	4-tert-Octylphenol	140-66-9	POS	1,85 × 10 ⁻⁹	7,37 × 10 ⁻⁸	10 ⁻¹¹ – 10 ⁻⁵	3,19 × 10 ⁻⁸	4,85 × 10 ⁻³	Phenol	Chemical Intermediate
9	Genistein	446-72-0	POS	2,24 × 10 ⁻⁹	2,45 × 10 ⁻⁸	10 ⁻¹¹ – 10 ⁻⁵	2,71 × 10 ⁻⁷	3,70 × 10 ⁻⁴	Flavonoid, Heterocyclic Compound	Natural Product, Pharmaceutical
6	Bisphenol A	80-05-7	POS	2,02 × 10 ⁻⁸	2,94 × 10 ⁻⁷	10 ⁻¹¹ – 10 ⁻⁵	5,33 × 10 ⁻⁷	4,38 × 10 ⁻³	Phenol	Chemical Intermediate

N° (7)	Substance	CASRN	Expected Response (1)	STTA Assay			VM7Luc ER TA Assay		MeSH Chemical Class (5)	Product Class (6)
				PC ₁₀ Value (M) (2)	PC ₅₀ Value (M) (2)	Test Conc. Range (M)	VM7Luc EC ₅₀ Value (M) (3)	Highest Conc. for Range Finder (M) (4)		
2	Kaempferol	520-18-3	POS	$1,36 \times 10^{-7}$	$1,21 \times 10^{-6}$	$10^{-11} - 10^{-5}$	$3,99 \times 10^{-6}$	$3,49 \times 10^{-3}$	Flavonoid, Heterocyclic Compound	Natural Product
3	Butylbenzyl phthalate	85-68-7	POS	$1,14 \times 10^{-6}$	$4,11 \times 10^{-6}$	$10^{-11} - 10^{-5}$	$1,98 \times 10^{-6}$	$3,20 \times 10^{-4}$	Carboxylic Acid, Ester, Phthalic Acid	Plasticiser, Industrial Chemical
4	<i>p,p'</i> -Methoxychlor	72-43-5	POS	$1,23 \times 10^{-6}$	—	$10^{-11} - 10^{-5}$	$1,92 \times 10^{-6}$	$2,89 \times 10^{-3}$	Hydrocarbon (Halogenated)	Pesticide, Veterinary Agent
1	Ethyl paraben	120-47-8	POS	$5,00 \times 10^{-6}$	—	$10^{-11} - 10^{-5}$	$2,48 \times 10^{-5}$	$6,02 \times 10^{-3}$	Carboxylic Acid, Phenol	Pharmaceutical, Preservative
17	Atrazine	1912-24-9	NEG	—	—	$10^{-10} - 10^{-4}$	—	$4,64 \times 10^{-4}$	Heterocyclic Compound	Herbicide
20	Spirolactone	52-01-7	NEG	—	—	$10^{-11} - 10^{-5}$	—	$2,40 \times 10^{-3}$	Lactone, Steroid	Pharmaceutical

N° (7)	Substance	CASRN	Expected Response (1)	STTA Assay			VM7Luc ER TA Assay		MeSH Chemical Class (5)	Product Class (6)
				PC ₁₀ Value (M) (2)	PC ₅₀ Value (M) (2)	Test Conc. Range (M)	VM7Luc EC ₅₀ Value (M) (3)	Highest Conc. for Range Finder (M) (4)		
21	Ketoconazole	65277-42-1	NEG	—	—	10 ⁻¹¹ – 10 ⁻⁵	—	9,41 × 10 ⁻⁵	Heterocyclic Compound	Pharmaceutical
22	Reserpine	50-55-5	NEG	—	—	10 ⁻¹¹ – 10 ⁻⁵	—	1,64 × 10 ⁻³	Heterocyclic Compound, Indole	Pharmaceutical, Veterinary Agent

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; EC₅₀ = half maximal effective concentration of test substance; NEG = negative; POS = positive; PC₁₀ (and PC₅₀) = the concentration of a test substance at which the response is 10 % (or 50 % for PC₅₀) of the response induced by the positive control (E2, 1nM) in each plate.

(1) Classification as positive or negative for ER agonist activity was based upon the ICCVAM Background Review Documents (BRD) for ER Binding and TA assays (31) as well as empirical data and other information obtained from referenced studies published and reviewed after the completion of the ICCVAM BRDs(2) (3) (18) (31) (32) (33) (34).

(2) Values reported in Draft Report of Pre-validation and Inter-laboratory Validation For Stably Transfected Transcriptional Activation (TA) Assay to Detect Estrogenic Activity - The Human Estrogen Receptor Alpha Mediated Reporter Gene Assay Using hER-HeLa-9903 Cell Line (30).

(3) Mean EC₅₀ values were calculated with values reported by the laboratories of the VM7Luc ER TA validation study (XDS, ECVAM, and Hiyoshi) (3).

(4) Concentrations reported were the highest concentrations tested (range finder) during the validation of the VM7Luc ER TA Assay. If concentrations differed between the laboratories, the highest concentration is reported. See table 4-10 of ICCVAM Test Method Evaluation Report: The LUMI-Cell@ER (VM7Luc ER TA) Test Method: An *In Vitro* Assay for Identifying Human Estrogen Receptor Agonist and Antagonist Activity of Chemicals (3).

(5) Substances were assigned into one or more chemical classes using the U.S. National Library of Medicine's Medical Subject Headings (MeSH), an internationally recognised standardised classification scheme (available at: <http://www.nlm.nih.gov/mesh>).

(6) Substances were assigned into one or more product classes using the U.S. National Library of Medicine's Hazardous Substances Database (available at: <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDDB>) From Table 1 (List of Reference Chemicals (22) for Evaluation of ER Agonist Accuracy) of the Performance Standards (6)

(7) If a proficiency substance is no longer commercially available, a substance with the same classification and, comparable potency, mode of action and chemical class can be used.

Table 4
List of (10) Proficiency Substances for antagonist assay

	Substance ⁽⁴⁾	CASRN	ER STTA assay ⁽¹⁾			VM7Luc ER TA assay ⁽²⁾			ER STTA ⁽¹⁾ Candidate Effects	ICCVAM ⁽⁵⁾ Consensus Classification	MeSH ⁽⁶⁾ Chemical Class	Product Class ⁽⁷⁾
			ER TA Activity	IC ₅₀ (M)	Test Conc. range (M)	ER TA Activity	IC ₅₀ ⁽³⁾ (M)	Highest Conc. for Range Finder (M) ⁽⁴⁾				
1	4-hydroxy-tamoxifen	68047-06-3	POS	$3,97 \times 10^{-9}$	$10^{-12} - 10^{-7}$	POS	$2,08 \times 10^{-7}$	$2,58 \times 10^{-4}$	moderate POS	POS	Hydrocarbon (Cyclic)	Pharmaceutical
2	Raloxifene HCl	82640-04-8	POS	$7,86 \times 10^{-10}$	$10^{-12} - 10^{-7}$	POS	$1,19 \times 10^{-9}$	$1,96 \times 10^{-4}$	moderate POS	POS	Hydrocarbon (Cyclic)	Pharmaceutical
3	Tamoxifen	10540-29-1	POS	$4,91 \times 10^{-7}$	$10^{-10} - 10^{-5}$	POS	$8,17 \times 10^{-7}$	$2,69 \times 10^{-4}$	POS	POS	Hydrocarbon (Cyclic)	Pharmaceutical
4	17 β -estradiol	50-28-2	NEG	—	$10^{-9} - 10^{-4}$	NEG	—	$3,67 \times 10^{-3}$	to be negative (*)	PN	Steroid	Pharmaceutical, Veterinary Agent
5	Apigenin	520-36-5	NEG	—	$10^{-9} - 10^{-4}$	NEG	—	$3,70 \times 10^{-4}$	NEG	NEG	Heterocyclic Compound	Dye, Natural Product, Pharmaceutical Intermediate
6	Di-n-butyl phthalate	84-74-2	NEG	—	$10^{-8} - 10^{-3}$	NEG	—	$3,59 \times 10^{-3}$	NEG	NEG	Ester, Phthalic Acid	Cosmetic Ingredient, Industrial Chemical, Plasticiser

	Substance ⁽⁴⁾	CASRN	ER STTA assay ⁽¹⁾			VM7Luc ER TA assay ⁽²⁾			ER STTA ⁽¹⁾ Candidate Effects	ICCVAM ⁽⁵⁾ Consensus Classification	MeSH ⁽⁶⁾ Chemical Class	Product Class ⁽⁷⁾
			ER TA Activity	IC ₅₀ (M)	Test Conc. range (M)	ER TA Activity	IC ₅₀ ⁽³⁾ (M)	Highest Conc. for Range Finder (M) ⁽⁴⁾				
7	Flavone	525-82-6	NEG	—	10 ⁻⁸ – 10 ⁻³	NEG	—	4,50 × 10 ⁻⁴	to be negative (*)	PN	Flavonoid, Heterocyclic Compound	Natural Product, Pharmaceutical
8	Genistein	446-72-0	NEG	—	10 ⁻⁹ – 10 ⁻⁴	NEG	—	3,70 × 10 ⁻⁴	to be negative (*)	NEG	Flavonoid, Heterocyclic Compound	Natural Product, Pharmaceutical
9	p-n-nonylphenol	104-40-5	NEG	—	10 ⁻⁹ – 10 ⁻⁴	NEG	—	4,54 × 10 ⁻⁴	not tested	NEG	Phenol	Chemical Intermediate
10	Resveratrol	501-36-0	NEG	—	10 ⁻⁸ – 10 ⁻³	NEG	—	4,38 × 10 ⁻⁴	to be negative (*)	NEG	Hydrocarbon (Cyclic)	Natural Product

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; M = molar; IC₅₀ = half maximal inhibitory concentration of test substance; NEG = negative; PN = presumed negative; POS = positive. (*) classified negative according to literature review (2).

(4) Common substances tested in the STTA and VM7Luc ER TA assays that were designated as ER antagonists or negatives and used to evaluate accuracy in the VM7Luc ER TA validation study (2) (3).

(1) The Validation Report of the Stably transfected Transcriptional Activation Assay to Detect ER mediated activity, Part B (2)

(2) ICCVAM Test Method Evaluation Report on the LUMI-CELL ER (VM7Luc ER TA) Test Method: An *In Vitro* Method for Identifying ER Agonists and Antagonists (3).

(3) Mean IC₅₀ values were calculated with values reported by the laboratories of the VM7Luc ER TA validation study (XDS, ECVAM, and Hiyoshi) (3).

(4) Concentrations reported were the highest concentrations tested (range finder) during the validation of the VM7Luc ER TA Assay. If concentrations differed between the laboratories, the highest concentration is reported. See table 4-11 of ICCVAM Test Method Evaluation Report: The LUMI-Cell@ER (VM7Luc ER TA) Test Method: An *In Vitro* Assay for Identifying Human Estrogen Receptor Agonist and Antagonist Activity of Chemicals (3).

(5) Classification as an ER antagonist or negative was based upon information in the ICCVAM Background Review Documents (BRD) for ER Binding and TA test methods (31) as well as information obtained from publications published and reviewed after the completion of the ICCVAM BRDs (2) (3) (18) (31).

(6) Substances were assigned to one or more chemical classes using the U.S. National Library of Medicine's Medical Subject Headings (MeSH), an internationally recognised standardised classification scheme (available at <http://www.nlm.nih.gov/mesh>).

(7) Substances were assigned to one or more product classes using the U.S. National Library of Medicine's Hazardous Substances Data Bank (available at <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDDB>).

Test Run Acceptability Criteria

16. Acceptance or rejection of a test run is based on the evaluation of results obtained for the reference standards and controls used for each experiment. Values for the PC_{50} (EC_{50}) or IC_{50} for the reference standards should meet the acceptability criteria as provided for the selected assay (for STTA see Appendix 2, for VM7Luc ER TA see Appendix 3), and all positive/negative controls should be correctly classified for each accepted experiment. The ability to consistently conduct the assay should be demonstrated by the development and maintenance of a historical database for the reference standards and controls (see paragraph 15). Standard deviations (SD) or coefficients of variation (CV) for the means of reference standards curve fitting parameters from multiple experiments may be used as a measure of within-laboratory reproducibility. In addition, the following principles regarding acceptability criteria should be met:
- Data should be sufficient for a quantitative assessment of ER activation (for agonist assay) or suppression (for antagonist assay) (i.e. efficacy and potency).
 - The mean reporter activity for the reference concentration of reference estrogen should be at least the minimum specified in the assays relative to that of the vehicle (solvent) control to ensure adequate sensitivity. For the STTA and VM7Luc ER TA assays, this is four times that of the mean vehicle control on each plate.
 - The concentrations tested should remain within the solubility range of the test chemicals and not demonstrate cytotoxicity.

Analysis of data

17. The defined data interpretation procedure for each assay should be used for classifying a positive and negative response.
18. Meeting the acceptability criteria (paragraph 16) indicates the assay is operating properly, but it does not ensure that any particular test run will produce accurate data. Replicating the results of the first run is the best indication that accurate data were produced. If two runs give reproducible results (e.g. both test run results indicate a test chemical is positive), it is not necessary to conduct a third run.
19. If two runs do not give reproducible results (e.g. a test chemical is positive in one run and negative in the other run), or if a higher degree of certainty is required regarding the outcome of this assay, at least three independent runs should be conducted. In this case the classification is based on the two concordant results out of the three.

General Data Interpretation Criteria

20. There is currently no universally agreed method for interpreting ER TA data. However, both qualitative (e.g. positive/negative) and/or quantitative (e.g. EC_{50} , PC_{50} , IC_{50}) assessments of ER-mediated activity should be based on empirical data and sound scientific judgment. Where possible, positive results should be characterised by both the magnitude of the effect as compared to the vehicle (solvent) control or reference estrogen and the concentration at which the effect occurs (e.g. an EC_{50} , PC_{50} , RPC_{Max} , IC_{50} , etc.).

Test Report

21. The test report should include the following information:

Assay:

- assay used;
- control/Reference standard/Test chemical;
- source, lot number, limit date for use, if available;

- stability of the test chemical itself, if known;
- solubility and stability of the test chemical in solvent, if known;
- measurement of pH, osmolality and precipitate in the culture medium to which the test chemical was added, as appropriate.

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Solvent/Vehicle:

- characterisation (nature, supplier and lot);
- justification for choice of solvent/vehicle;
- solubility and stability of the test chemical in solvent/vehicle, if known.

Cells:

- type and source of cells:
 - Is ER endogenously expressed? If not, which receptor(s) were transfected?
 - Reporter construct(s) used (including source species);
 - Transfection method;
 - Selection method for maintenance of stable transfection (where applicable);
 - Is the transfection method relevant for stable lines?
- number of cell passages (from thawing);
- passage number of cells at thawing;
- methods for maintenance of cell cultures.

Test conditions:

- solubility limitations;
- description of the methods of assessing viability applied;
- composition of media, CO₂ concentration;
- concentrations of test chemical;
- volume of vehicle and test chemical added;
- incubation temperature and humidity;
- duration of treatment;
- cell density at the start of - and during treatment;
- positive and negative reference standards;
- reporter reagents (product name, supplier and lot);
- criteria for considering test runs as positive, negative or equivocal.

Acceptability check:

- fold inductions for each assay plate and whether they meet the minimum required by the assay based on historical controls;
- actual values for acceptability criteria, e.g. $\log_{10}EC_{50}$, $\log_{10}PC_{50}$, $\log IC_{50}$ and Hillslope values, for concurrent positive controls/reference standards.

Results:

- raw and normalised data;
- the maximum fold induction level;
- cytotoxicity data;
- if it exists, the lowest effective concentration (LEC);
- RPC_{Max} , PC_{Max} , PC_{50} , IC_{50} and/or EC_{50} values, as appropriate;
- concentration-response relationship, where possible;

- statistical analyses, if any, together with a measure of error and confidence (e.g. SEM, SD, CV or 95 % CI) and a description of how these values were obtained.

Discussion of the results

Conclusion

LITERATURE

- (1) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environment, Health and Safety Publications, Series on Testing and Assessment (No 34.), Organisation for Economic Cooperation and Development, Paris.
- (2) OECD (2015). Report of the Inter-Laboratory Validation for Stably Transfected Transactivation Assay to detect Estrogenic and Anti-estrogenic Activity. Environment, Health and Safety Publications, Series on Testing and Assessment (No 225), Organisation for Economic Cooperation and Development, Paris.
- (3) ICCVAM (2011). ICCVAM Test Method Evaluation Report on the LUMI-CELL® ER (BG1Luc ER TA) Test Method, an *In Vitro* Method for Identifying ER Agonists and Antagonists, National Institute of Environmental Health Sciences: Research Triangle Park, NC.
- (4) Pujol P. *et al.* (1998). Differential Expression of Estrogen Receptor-Alpha and -Beta Messenger RNAs as a Potential Marker of Ovarian Carcinogenesis, *Cancer. Res.*, 58(23): p. 5367-73.
- (5) Rogers J.M. and Denison M.S. (2000). Recombinant Cell Bioassays for Endocrine Disruptors: Development of a Stably Transfected Human Ovarian Cell Line for the Detection of Estrogenic and Anti-Estrogenic Chemicals, *In Vitro and Molecular Toxicology: Journal of Basic and Applied Research*, 13(1): p. 67-82.
- (6) OECD (2012). Performance Standards For Stably Transfected Transactivation *In Vitro* Assay to Detect Estrogen Receptor Agonists (for TG 455). Environment, Health and Safety Publications, Series on Testing and Assessment (No 173.), Organisation for Economic Cooperation and Development, Paris.
- (7) OECD (2015). Performance Standards For Stably Transfected Transactivation *In Vitro* Assay to Detect Estrogen Receptor Antagonists. Environment, Health and Safety Publications, Series on Testing and Assessment (No 174.), Organisation for Economic Cooperation and Development, Paris.
- (8) OECD (2012). Guidance Document on Standardized Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Environment, Health and Safety Publications, Series on Testing and Assessment (No 150.), Organisation for Economic Cooperation and Development, Paris.
- (9) Cavailles V. (2002). Estrogens and Receptors: an Evolving Concept. *Climacteric*, 5 Suppl 2: p. 20- 6.
- (10) Welboren W.J. *et al.* (2009). Genomic Actions of Estrogen Receptor Alpha: What are the Targets and how are they Regulated? *Endocr. Relat. Cancer*, 16(4): p. 1073-89.
- (11) Younes M. and Honma N. (2011). Estrogen Receptor Beta, *Arch. Pathol. Lab. Med.*, 135(1): p. 63- 6.
- (12) Jefferson W.N., *et al.* (2002). Assessing Estrogenic Activity of Phytochemicals Using Transcriptional Activation and Immature Mouse Uterotrophic Responses, *Journal of Chromatography B*, 777(1-2): p. 179-189.

- (13) Sonneveld E. *et al.* (2006). Comparison of *In Vitro* and *In Vivo* Screening Models for Androgenic and Estrogenic Activities, *Toxicol. Sci.*, 89(1): p. 173-187.
- (14) Takeyoshi M. *et al.* (2002). The Efficacy of Endocrine Disruptor Screening Tests in Detecting Anti- Estrogenic Effects Downstream of Receptor-Ligand Interactions, *Toxicology Letters*, 126(2): p. 91- 98.
- (15) Combes R.D. (2000). Endocrine Disruptors: a Critical Review of *In Vitro* and *In Vivo* Testing Strategies for Assessing their Toxic Hazard to Humans, *ATLA Alternatives to Laboratory Animals*, 28(1): p. 81-118.
- (16) Escande A. *et al.* (2006). Evaluation of Ligand Selectivity Using Reporter Cell Lines Stably Expressing Estrogen Receptor Alpha or Beta, *Biochem. Pharmacol.*, 71(10): p. 1459-69.
- (17) Gray L.E. Jr. (1998). Tiered Screening and Testing Strategy for Xenoestrogens and Antiandrogens, *Toxicol. Lett.*, 102-103, 677-680.
- (18) EDSTAC (1998). Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) Final Report.
- (19) ICCVAM (2003). ICCVAM Evaluation of *In Vitro* Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays.
- (20) Gustafsson J.Ö. (1999). Estrogen Receptor β - A New Dimension in Estrogen Mechanism of Action, *Journal of Endocrinology*, 163(3): p. 379-383.
- (21) Ogawa S. *et al.* (1998). The Complete Primary Structure of Human Estrogen Receptor β (hER β) and its Heterodimerization with ER *In Vivo* and *In Vitro*, *Biochemical and Biophysical Research Communications*, 243(1): p. 122-126.
- (22) Enmark E. *et al.* (1997). Human Estrogen Receptor β -Gene Structure, Chromosomal Localization, and Expression Pattern, *Journal of Clinical Endocrinology and Metabolism*, 82(12): p. 4258-4265.
- (23) Ball L.J. *et al.* (2009). Cell Type- and Estrogen Receptor-Subtype Specific Regulation of Selective Estrogen Receptor Modulator Regulatory Elements, *Molecular and Cellular Endocrinology*, 299(2): p. 204-211.
- (24) Barkhem T. *et al.* (1998). Differential Response of Estrogen Receptor Alpha and Estrogen Receptor Beta to Partial Estrogen Agonists/Antagonists, *Mol. Pharmacol.*, 54(1): p. 105-12.
- (25) Deroo B.J. and Buensuceso A.V. (2010). Minireview: Estrogen Receptor- β : Mechanistic Insights from Recent Studies, *Molecular Endocrinology*, 24(9): p. 1703-1714.
- (26) Harris D.M. *et al.* (2005). Phytoestrogens Induce Differential Estrogen Receptor Alpha- or Beta- Mediated Responses in Transfected Breast Cancer Cells, *Experimental Biology and Medicine*, 230(8): p. 558-568.
- (27) Anderson J.N. Clark J.H. and Peck E.J.Jr. (1972). The Relationship Between Nuclear Receptor- Estrogen Binding and Uterotrophic Responses, *Biochemical and Biophysical Research Communications*, 48(6): p. 1460-1468.
- (28) Toft D. (1972). The Interaction of Uterine Estrogen Receptors with DNA, *Journal of Steroid Biochemistry*, 3(3): p. 515-522.
- (29) Gorski J. *et al.* (1968), Hormone Receptors: Studies on the Interaction of Estrogen with the Uterus, *Recent Progress in Hormone Research*, 24: p. 45-80.

-
- (30) Jensen E.V. *et al.* (1967), Estrogen-Receptor Interactions in Target Tissues, *Archives d'Anatomie Microscopique et de Morphologie Experimentale*, 56(3):p. 547-569.
- (31) ICCVAM (2002). Background Review Document: Estrogen Receptor Transcriptional Activation (TA) Assay. Appendix D, Substances Tested in the ER TA Assay, NIH Publication Report (No 03-4505).
- (32) Kanno J. *et al.* (2001). The OECD Program to Validate the Rat Uterotrophic Bioassay to Screen Compounds for *In Vivo* Estrogenic Responses: Phase 1, *Environ. Health Persp.*, 109:785-94.
- (33) Kanno J. *et al.* (2003). The OECD Program to Validate the Rat Uterotrophic Bioassay: Phase Two Dose-Response Studies, *Environ. Health Persp.*, 111:1530-1549.
- (34) Kanno J. *et al.* (2003), The OECD Program to Validate the Rat Uterotrophic Bioassay: Phase Two – Coded Single-Dose Studies, *Environ. Health Persp.*, 111:1550-1558.
- (35) Geisinger *et al.* (1989) Characterization of a human ovarian carcinoma cell line with estrogen and progesterone receptors, *Cancer* 63, 280-288.
- (36) Baldwin *et al.* (1998) BG-1 ovarian cell line: an alternative model for examining estrogen-dependent growth *in vitro*, *In Vitro Cell. Dev. Biol. – Animal*, 34, 649-654.
- (37) Li, Y. *et al.* (2014) Research resource: STR DNA profile and gene expression comparisons of human BG-1 cells and a BG-1/MCF-7 clonal variant, *Mol. Endo.* 28, 2072-2081.
- (38) Rogers, J.M. and Denison, M.S. (2000) Recombinant cell bioassays for endocrine disruptors: development of a stably transfected human ovarian cell line for the detection of estrogenic and anti-estrogenic chemicals, *In Vitro & Molec. Toxicol.* 13, 67-82.

Appendix 1

DEFINITIONS AND ABBREVIATIONS

Acceptability criteria: Minimum standards for the performance of experimental controls and reference standards. All acceptability criteria should be met for an experiment to be considered valid.

Accuracy (concordance): The closeness of agreement between assay results and an accepted reference values. It is a measure of assay performance and one aspect of relevance. The term is often used interchangeably with “concordance” to mean the proportion of correct outcomes of a assay (1).

Agonist: A substance that produces a response, e.g. transcription, when it binds to a specific receptor.

Antagonist: A type of receptor ligand or chemical that does not provoke a biological response itself upon binding to a receptor, but blocks or dampens agonist-mediated responses.

Anti-estrogenic activity: the capability of a chemical to suppress the action of 17 β -estradiol mediated through estrogen receptors.

Cell morphology: The shape and appearance of cells grown in a monolayer in a single well of a tissue culture plate. Cells that are dying often exhibit abnormal cell morphology.

CF: The OECD Conceptual Framework for the Testing and Evaluation of Endocrine Disrupters.

Charcoal/dextran treatment: Treatment of serum used in cell culture. Treatment with charcoal/dextran (often referred to as “stripping”) removes endogenous hormones and hormone-binding proteins.

Chemical: A substance or a mixture.

Cytotoxicity: Harmful effects to cell structure or function that can ultimately cause cell death and can be reflected by a reduction in the number of cells present in the well at the end of the exposure period or a reduction of the capacity for a measure of cellular function when compared to the concurrent vehicle control.

CV: Coefficient of variation

DCC-FBS: Dextran-coated charcoal treated fetal bovine serum.

DMEM: Dulbecco's Modification of Eagle's Medium

DMSO: Dimethyl sulfoxide

E2: 17 β -estradiol

EC50: The half maximal effective concentration of a test chemical.

ED: Endocrine disruption

hER α : Human estrogen receptor alpha

hER β : Human estrogen receptor beta

EFM: Estrogen-free medium. Dulbecco's Modification of Eagle's Medium (DMEM) supplemented with 4.5 % charcoal/dextran-treated FBS, 1,9 % L-glutamine, and 0,9 % Pen-Strep.

ER: Estrogen receptor

ERE: Estrogen response element

Estrogenic activity: The capability of a chemical to mimic 17 β -estradiol in its ability to bind to and activate estrogen receptors. hER α -mediated estrogenic activity can be detected with this test method.

ERTA: Estrogen Receptor Trans Activation

FBS: Fetal bovine serum

HeLa: An immortal human cervical cell line

HeLa9903: A HeLa cell subclone into which hER α and a luciferase reporter gene have been stably transfected

IC50: The half maximal effective concentration of an inhibitory test chemical.

ICCVAM: The Interagency Coordinating Committee on the Validation of Alternative Methods.

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes, and indicates the extent to which an assay can be successfully transferred between laboratories, also referred to as between-laboratory reproducibility (1).

Intra-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Also referred to as "within-laboratory reproducibility" (1).

LEC: Lowest effective concentration is the lowest concentration of test chemical that produces a response (i.e. the lowest test chemical concentration at which the fold induction is statistically different from the concurrent vehicle control).

Me-too test: A colloquial expression for an assay that is structurally and functionally similar to a validated and accepted reference test method. Interchangeably used with similar test method.

MT: Metallothionein

MMTV: Mouse Mammary Tumor Virus

OHT: 4-Hydroxytamoxifen

PBTG: Performance-Based Test Guideline

PC (Positive control): a strongly active substance, preferably 17 β -estradiol that is included in all tests to help ensure proper functioning of the assay.

PC10: the concentration of a test chemical at which the measured activity in an agonist assay is 10 % of the maximum activity induced by the PC (E2 at 1nM for the STTA assay) in each plate.

PC50: the concentration of a test chemical at which the measured activity in an agonist assay is 50 % of the maximum activity induced by the PC (E2 at the reference concentration specified in the test method) in each plate.

PCMax: the concentration of a test chemical inducing the RPCMax

Performance standards: Standards, based on a validated assay, that provide a basis for evaluating the comparability of a proposed assay that is mechanistically and functionally similar. Included are (1) essential assay components; (2) a minimum list of reference chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (3) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed assay should demonstrate when evaluated using the minimum list of reference chemicals (1).

Proficiency substances: A subset of the reference substances included in the Performance Standards that can be used by laboratories to demonstrate technical competence with a standardised test method. Selection criteria for these substances typically include that they represent the range of responses, are commercially available, and have high quality reference data available.

Proficiency: The demonstrated ability to properly conduct an assay prior to testing unknown substances.

Reference estrogen (Positive control, PC): 17 β -estradiol (E2, CAS 50-28-2).

Reference standard: a reference substance used to demonstrate the adequacy of a assay. 17 β -estradiol is the reference standard for the STTA and VM7Luc ER TA assays.

Reference test methods: The assays upon which PBTG 455 is based.

Relevance: Description of relationship of an assay to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the assay correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of an assay (1).

Reliability: Measure of the extent that an assay can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility.

RLU: Relative Light Units

RNA: Ribonucleic Acid

RPCMax: maximum level of response induced by a test chemical, expressed as a percentage of the response induced by 1 nM E2 on the same plate

RPMI: RPMI 1640 medium supplemented with 0,9 % Pen-Strep and 8.0 % fetal bovine serum (FBS)

Run: An individual experiment that evaluates chemical action on the biological outcome of the assay. Each run is a complete experiment performed on replicate wells of cells plated from a common pool of cells at the same time.

Independent run: A separate, independent experiment that evaluates chemical action on the biological outcome of the assay, using cells from a different pool, freshly diluted chemicals, conducted on different days or on the same day by different staff.

SD: Standard deviation.

Sensitivity: The proportion of all positive/active substances that are correctly classified by the assay. It is a measure of accuracy for an assay that produces categorical results, and is an important consideration in assessing the relevance of an assay (1).

Specificity: The proportion of all negative/inactive substances that are correctly classified by the test. It is a measure of accuracy for an assay that produces categorical results, and is an important consideration in assessing the relevance of an assay (1).

Stable transfection: When DNA is transfected into cultured cells in such a way that it is stably integrated into the cells genome, resulting in the stable expression of transfected genes. Clones of stably transfected cells are selected by stable markers (e.g. resistance to G418).

STTA Assay: Stably Transfected Transactivation Assay, the ER α transcriptional activation assay using the HeLa 9903 Cell Line.

Study: The full range of experimental work performed to evaluate a single, specific substance using a specific assay. A study comprises all steps including tests of dilution of test substance in the test media, preliminary range finding runs, all necessary comprehensive runs, data analyses, quality assurance, cytotoxicity assessments, etc. Completion of a study allows the classification of the test chemical activity on the toxicity target (i.e. active, inactive or inconclusive) that is evaluated by the assay used and an estimate of potency relative to the positive reference chemical.

Substance: Under REACH ⁽¹⁾, a substance is defined as a chemical element and its compounds in the natural state or obtained by any manufacturing process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition. A very similar definition is used in the context of the UN GHS (1).

TA (Transactivation): The initiation of mRNA synthesis in response to a specific chemical signal, such as a binding of an estrogen to the estrogen receptor

Assay: Within the context of this test method, an assay is one of the methodologies accepted as valid in meeting the outlined performance criteria. Components of assay include, for example, the specific cell line with associated growth conditions, specific media in which the test is conducted, plate set up conditions, arrangement and dilutions of test chemicals along with any other required quality control measures and associated data evaluation steps.

Test chemical: Any substance or mixture tested using this test method.

Transcription: mRNA synthesis

UVCB: Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials

Validated test method: An assay for which validation studies have been completed to determine the relevance (including accuracy) and reliability for a specific purpose. It is important to note that a validated test method may not have sufficient performance in terms of accuracy and reliability to be found acceptable for the proposed purpose (1).

Validation: The process by which the reliability and relevance of a particular approach, method, assay, process or assessment is established for a defined purpose (1).

VC (Vehicle control): The solvent that is used to dissolve test and control chemicals is tested solely as vehicle without dissolved chemical.

VM7: An immortalised adenocarcinoma cell that endogenously express estrogen receptor.

VM7Luc4E2: The VM7Luc4E2 cell line was derived from VM7 immortalised human-derived adenocarcinoma cells that endogenously express both forms of the estrogen receptor (ER α and ER β) and have been stably transfected with the plasmid pGudLuc7.ERE. This plasmid contains four copies of a synthetic oligonucleotide containing the estrogen response element upstream of the mouse mammary tumor viral (MMTV) promoter and the firefly luciferase gene.

Weak positive control: A weakly active substance selected from the reference chemicals list that is included in all tests to help ensure proper functioning of the assay.

⁽¹⁾ Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (Reach), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC (OJ L 304, 22.11.2007, p. 1).

*Appendix 2*STABLY TRANSFECTED HUMAN ESTROGEN RECEPTOR- α TRANSACTIVATION ASSAY FOR DETECTION OF ESTROGENIC AGONIST AND ANTAGONIST ACTIVITY OF CHEMICALS USING THE hER α -HELA-9903 CELL LINE

INITIAL CONSIDERATIONS AND LIMITATIONS (SEE ALSO GENERAL INTRODUCTION)

1. This transactivation (TA) assay uses the hER α -HeLa-9903 cell line to detect estrogenic agonist activity mediated through human estrogen receptor alpha (hER α). The validation study of the Stably Transfected Transactivation (STTA) Assay by the Japanese Chemicals Evaluation and Research Institute (CERI) using the hER α -HeLa-9903 cell line to detect estrogenic agonist and antagonist activity mediated through human estrogen receptor alpha (hER α) demonstrated the relevance and reliability of the assay for its intended purpose (1).
2. This assay is specifically designed to detect hER α -mediated TA by measuring chemiluminescence as the endpoint. However, non-receptor-mediated luminescence signals have been reported at phytoestrogen concentrations higher than 1 μ M due to the over-activation of the luciferase reporter gene (2) (3). While the dose-response curve indicates that true activation of the ER system occurs at lower concentrations, luciferase expression obtained at high concentrations of phytoestrogens or similar compounds suspected of producing phytoestrogen-like over-activation of the luciferase reporter gene needs to be examined carefully in stably transfected ER TA assay systems (Appendix 1).
3. The sections "GENERAL INTRODUCTION" and "ER TA ASSAY COMPONENTS" should be read before using this assay for regulatory purposes. Definitions and abbreviations used in this TG are described in Appendix 2.1.

PRINCIPLE OF THE ASSAY (SEE ALSO GENERAL INTRODUCTION)

4. The assay is used to signal binding of the estrogen receptor with a ligand. Following ligand binding, the receptor-ligand complex translocates to the nucleus where it binds specific DNA response elements and transactivates a firefly luciferase reporter gene, resulting in increased cellular expression of luciferase enzyme. Luciferin is a substrate that is transformed by the luciferase enzyme to a bioluminescence product that can be quantitatively measured with a luminometer. Luciferase activity can be evaluated quickly and inexpensively with a number of commercially available test kits.
5. The test system utilises the hER α -HeLa-9903 cell line, which is derived from a human cervical tumor, with two stably inserted constructs: (i) the hER α expression construct (encoding the full-length human receptor), and (ii) a firefly luciferase reporter construct bearing five tandem repeats of a vitellogenin Estrogen-Responsive Element (ERE) driven by a mouse metallothionein (MT) promoter TATA element. The mouse MT TATA gene construct has been shown to have the best performance, and so is commonly used. Consequently this hER α -HeLa-9903 cell line can measure the ability of a test chemical to induce hER α -mediated transactivation of luciferase gene expression.
6. In case of ER agonist assay, data interpretation is based upon whether or not the maximum response level induced by a test chemical equals or exceeds an agonist response equal to 10 % of that induced by a maximally inducing (1 nM) concentration of the positive control (PC) 17 β -estradiol (E2) (i.e. the PC10). In case of ER antagonist assay, data interpretation is based upon whether or not the response shows at least a 30 % reduction in activity from the response induced by the spike in control (25 pM of E2) without cytotoxicity. Data analysis and interpretation are discussed in detail in paragraphs 34 - 48.

PROCEDURE

Cell Lines

7. The stably transfected hER α -HeLa-9903 cell line should be used for the assay. The cell line can be obtained from the Japanese Collection of Research Bioresources (JCRB) Cell Bank (¹), upon signing a Material Transfer Agreement (MTA).
8. Only cells characterised as mycoplasma-free should be used in testing. RT-PCR (Real Time Polymerase Chain Reaction) is the method of choice for a sensitive detection of mycoplasma infection (4) (5) (6).

Stability of the cell line

9. To monitor the stability of the cell line, E2, 17 α -estradiol, 17 α -methyltestosterone and corticosterone should be used as the reference standards for agonist assay and a complete concentration-response curve in the test concentration range provided in Table 1 should be measured at least once each time the assay is performed, and the results should be in agreement with the results provided in Table 1.
10. In case of antagonist assay, complete concentration curves for two reference standards, tamoxifen and flutamide, should be measured simultaneously with each run. Correct qualitative classification as positive or negative for the two chemicals should be monitored.

Cell Culture and Plating Conditions

11. Cells should be maintained in Eagle's Minimum Essential Medium (EMEM) without phenol red, supplemented with 60 mg/l of antibiotic kanamycine and 10 % dextran-coated-charcoal-treated fetal bovine serum (DCC-FBS), in a CO₂ incubator (5 % CO₂) at 37 \pm 1°C. Upon reaching 75 -90 % confluency, cells can be subcultured at 10 ml of 0,4 x 10⁵ - 1 x 10⁵ cells/ml for 100 mm cell culture dish. Cells should be suspended with 10 % FBS-EMEM (which is the same as EMEM with DCC-FBS) and then plated into wells of a microplate at a density of 1 x 10⁴ cells/(100 μ l x well). Next, the cells should be pre-incubated in a 5 % CO₂ incubator at 37 \pm 1°C for 3 hours before the chemical exposure. The plastic-ware should be free of estrogenic activity.
12. To maintain the integrity of the response, the cells should be grown for more than one passage from the frozen stock in the conditioned media and should not be cultured for more than 40 passages. For the hER α -HeLa-9903 cell line, this will be less than three months. However the performance of cells may be reduced if they are grown in inappropriate culture conditions.
13. The DCC-FBS can be prepared as described in Appendix 2.2, or obtained from commercial sources.

Acceptability criteria

Positive and negative reference standards for ER agonist assay

14. Prior to and during the study, the responsiveness of the test system should be verified using the appropriate concentrations of a strong estrogen: E2, a weak estrogen (17 α -estradiol), a very weak agonist (17 α -methyltestosterone), and a negative substance (corticosterone). Acceptable range values derived from the validation study (1) are given in Table 1. These 4 concurrent reference standards should be included with each experiment and the results should fall within the given acceptable limits. If this is not the case, the cause for the failure to meet the acceptability criteria should be determined (e.g. cell handling, and serum and antibiotics for quality and concentration) and the

(¹) JCRB Cell Bank: National Institute of Biomedical Innovation, 7-6-8 Asagi Saito, Ibaraki-shi, Osaka 567-0085, Japan Fax: +81-72-641-9812

assay repeated. Once the acceptability criteria have been achieved, to ensure minimum variability of EC₅₀, PC₅₀ and PC₁₀ values, consistent use of materials for cell culturing is essential. The four concurrent reference standards, which should be included in each experiment (conducted under the same conditions including the materials, passage level of cells and technicians), can ensure the sensitivity of the assay because the PC₁₀s of the three positive reference standards should fall within the acceptable range, as should the PC₅₀s and EC₅₀s where they can be calculated (see Table 1).

Table 1

Acceptable range values of the four reference standards for the ER agonist assay

Name	logPC ₅₀	logPC ₁₀	logEC ₅₀	Hill slope	Test range
17β-estradiol (E2) CAS No: 50-28-2	-11,4~-10,1	<-11	-11,3~-10,1	0,7~1,5	10 ⁻¹⁴ ~10 ⁻⁸ M
17α-estradiol CAS No: 57-91-0	-9,6~-8,1	-10,7~-9,3	-9,6~-8,4	0,9~2,0	10 ⁻¹² ~10 ⁻⁶ M
Corticosterone CAS No: 50-22-6	—	—	—	—	10 ⁻¹⁰ ~10 ⁻⁴ M
17α-methyltestosterone CAS No: 58-18-4	-6,0~-5,1	-8,0~-6,2	—	—	10 ⁻¹¹ ~10 ⁻⁵ M

Positive and negative reference standards for ER antagonist assay

15. Prior to and during the study, the responsiveness of the test system should be verified using the appropriate concentrations of a positive substance (Tamoxifen), and a negative substance (Flutamide). Acceptable range values derived from the validation study (1) are given in Table 2. These two concurrent reference standards should be included with each experiment and the results should be judged correct as shown in the criteria. If this is not the case, the cause for the failure to meet the criteria should be determined (e.g. cell handling, and serum and antibiotics for quality and concentration) and the assay repeated. In addition, IC₅₀ values for a positive substance (Tamoxifen) should be calculated and the results should fall within the given acceptable limits. Once the acceptability criteria have been achieved, to ensure minimum variability of IC₅₀ values, consistent use of materials for cell culturing is essential. The two concurrent reference standards, which should be included in each experiment (conducted under the same conditions including the materials, passage level of cells and technicians), can ensure the sensitivity of the assay (see Table 2).

Table 2

Criteria and acceptable range values of the two reference standards for the ER antagonist assay

Name	Criteria	LogIC ₅₀	Test range
Tamoxifen CAS No: 10540-29-1	Positive: IC ₅₀ should be calculated	-5,942 ~ -7,596	10 ⁻¹⁰ ~ 10 ⁻⁵ M
Flutamide CAS No: 13311-84-7	Negative: IC ₃₀ should not be calculated	—	10 ⁻¹⁰ ~ 10 ⁻⁵ M

Positive and Vehicle Controls

16. The positive control (PC) for ER agonist assay (1 nM of E2) and for ER antagonist assay (10 µM TAM) should be tested at least in triplicate in each plate. The vehicle that is used to dissolve a test chemical should be tested as a vehicle control (VC) at least in triplicate in each plate. In addition to this VC, if the PC uses a different vehicle than the test chemical, another VC should be tested at least in triplicate on the same plate with the PC.

Quality criteria for ER agonist assay

17. The mean luciferase activity of the positive control (1 nM E2) should be at least 4-fold that of the mean VC on each plate. This criterion is established based on the reliability of the endpoint values from the validation study (historically between four- and 30-fold).
18. With respect to the quality control of the assay, the fold-induction corresponding to the PC10 value of the concurrent PC (1 nM E2) should be greater than 1+2SD of the fold-induction value (=1) of the concurrent VC. For prioritisation purposes, the PC10 value can be useful to simplify the data analysis required compared to a statistical analysis. Although a statistical analysis provides information on significance, such an analysis is not a quantitative parameter with respect to concentration-based potential, and so is less useful for prioritisation purposes.

Quality criteria for ER antagonist assay

19. The mean luciferase activity of the spike in control (25 pM E2) should be at least 4-fold that of the mean VC on each plate. This criterion is established based on the reliability of the endpoint values from the validation study.
20. With respect to the quality control of the assay, relative transcriptional activation (RTA) of 1 nM E2 should be greater than 100 %, RTA of 1 µM 4-Hydroxytamoxifen (OHT) should be less than 40,6 % and RTA of 100 µM Digonin (Dig) should be less than 0 %.

Demonstration of Laboratory Proficiency (see paragraph 14 and Tables 3 and 4 in « ER TA ASSAY COMPONENTS » of this test method).

Vehicle

21. Dimethyl sulfoxide (DMSO), or appropriate solvent, at the same concentration used for the different positive and negative controls and the test chemicals should be used as the concurrent VC. Test chemicals should be dissolved in a solvent that solubilises that test chemical and is miscible with the cell medium. Water, ethanol (95 % to 100 % purity) and DMSO are suitable vehicles. If DMSO is used, the level should not exceed 0,1 % (v/v). For any vehicle, it should be demonstrated that the maximum volume used is not cytotoxic and does not interfere with assay performance.

Preparation of Test Chemicals

22. Generally, the test chemicals should be dissolved in DMSO or other suitable solvent, and serially diluted with the same solvent at a common ratio of 1:10 in order to prepare solutions for dilution with media.

Solubility and Cytotoxicity: Considerations for Range Finding.

23. A preliminary test should be carried out to determine the appropriate concentration range of chemical to be tested, and to ascertain whether the test chemical may have any solubility and cytotoxicity problems. Initially, chemicals are tested up to the maximum concentration of 1 µl/ml, 1 mg/ml, or 1 mM, whichever is the lowest. Based on the extent of cytotoxicity or lack of solubility observed in the preliminary test, the first definite run should test the chemical at log-serial dilutions starting at the maximum acceptable concentration (e.g. 1 mM, 100 µM, 10 µM, etc.) and the presence of cloudiness or precipitate or cytotoxicity noted. Concentrations in the second, and if necessary third run should be adjusted as appropriate to better characterise the concentration-response curve and to avoid concentrations which are found to be insoluble or to induce excessive cytotoxicity.

24. For ER agonists and antagonists, the presence of increasing levels of cytotoxicity can significantly alter or eliminate the typical sigmoidal response and should be considered when interpreting the data. Cytotoxicity testing methods that can provide information regarding 80 % cell viability should be used, utilising an appropriate assay based upon laboratory experience.
25. Should the results of the cytotoxicity test show that the concentration of the test chemical has reduced the cell number by 20 % or more, this concentration should be regarded as cytotoxic, and the concentrations at or above the cytotoxic concentration should be excluded from the evaluation.

Chemical Exposure and Assay Plate Organisation

26. The procedure for chemical dilutions (Steps-1 and 2) and exposure to cells (Step-3) can be conducted as follows:

Step-1: Each test chemical should be serially diluted in DMSO, or appropriate solvent, and added to the wells of a microtitre plate to achieve final serial concentrations as determined by the preliminary range finding test (typically in a series of, for example 1 mM, 100 μ M, 10 μ M, 1 μ M, 100 nM, 10 nM, 1 nM, 100 pM, and 10 pM (10^{-3} - 10^{-11} M)) for triplicate testing.

Step-2: Chemical dilution: First dilute 1,5 μ l of the test chemical in the solvent to a volume of 500 μ l of media.

Step-3: Chemical exposure of the cells: Add 50 μ l of dilution with media (prepared in Step-2) to an assay well containing 10^4 cells/100 μ l/well.

The recommended final volume of media required for each well is 150 μ l. Test samples and reference standards can be assigned as shown in Table 3 and Table 4.

Table 3

Example of plate concentration assignment of the reference standards in the assay plate in ER agonist assay

Row	17 α -methyltestosterone			Corticosterone			17 α -estradiol			E2		
	1	2	3	4	5	6	7	8	9	10	11	12
A	conc 1 (10 μ M)	→	→	100 μ M	→	→	1 μ M	→	→	10 nM	→	→
B	conc 2 (1 μ M)	→	→	10 μ M	→	→	100 nM	→	→	1 nM	→	→
C	conc 3 (100 nM)	→	→	1 μ M	→	→	10 nM	→	→	100 pM	→	→
D	conc 4 (10 nM)	→	→	100 nM	→	→	1 nM	→	→	10 pM	→	→
E	conc 5 (1 nM)	→	→	10 nM	→	→	100 pM	→	→	1 pM	→	→
F	conc 6 (100 pM)	→	→	1 nM	→	→	10 pM	→	→	0,1 pM	→	→
G	conc 7 (10 pM)	→	→	100 pM	→	→	1 pM	→	→	0,01 pM	→	→
H	VC	→	→	→	→	→	PC	→	→	→	→	→

VC: Vehicle control (0.1 % DMSO); PC: Positive control (1 nM E2)

27. The reference standards (E2, 17 α -estradiol, 17 α -methyl testosterone and corticosterone) should be tested in every run (Table 3). PC wells treated with 1 nM of E2 that can produce maximum induction of E2 and VC wells treated with DMSO (or appropriate solvent) alone should be included in each test assay plate (Table 4). If cells from different sources (e.g. different passage number, different lot, etc.) are used in the same experiment, the reference standards should be tested for each cell source.

Table 4

Example of plate concentration assignment of test and plate control chemicals in the assay plate in ER agonist assay

Row	Test Chemical 1			Test Chemical 2			Test Chemical 3			Test Chemical 4		
	1	2	3	4	5	6	7	8	9	10	11	12
A	conc 1 (10 μ M)	→	→	1 mM	→	→	1 μ M	→	→	10 nM	→	→
B	conc 2 (1 μ M)	→	→	100 μ M	→	→	100 nM	→	→	1 nM	→	→
C	conc 3 (100 nM)	→	→	10 μ M	→	→	10 nM	→	→	100 pM	→	→
D	conc 4 (10 nM)	→	→	1 μ M	→	→	1 nM	→	→	10 pM	→	→
E	conc 5 (1 nM)	→	→	100 nM	→	→	100 pM	→	→	1 pM	→	→
F	conc 6 (100 pM)	→	→	10 nM	→	→	10 pM	→	→	0,1 pM	→	→
G	conc 7 (10 pM)	→	→	1 nM	→	→	1 pM	→	→	0,01 pM	→	→
H	VC	→	→	→	→	→	PC	→	→	→	→	→


VC: Vehicle control (0.1 % DMSO); PC: Positive control (1 nM E2)

Table 5

Example of plate concentration assignment of the reference standards in the assay plate in ER antagonist assay

Row	Tamoxifen			Flutamide			Test Chemical 1			Test Chemical 2		
	1	2	3	4	5	6	7	8	9	10	11	12
A	conc 1 (10 μ M)	→	→	10 μ M	→	→	10 μ M	→	→	10 μ M	→	→
B	conc 2 (1 μ M)	→	→	1 μ M	→	→	1 μ M	→	→	1 μ M	→	→
C	conc 3 (100 nM)	→	→	100 nM	→	→	100 nM	→	→	100 nM	→	→
D	conc 4 (10 nM)	→	→	10 nM	→	→	10 nM	→	→	10 nM	→	→
E	conc 5 (1 nM)	→	→	1 nM	→	→	1 nM	→	→	1 nM	→	→
F	conc 6 (100 pM)	→	→	100 pM	→	→	100 pM	→	→	100 pM	→	→
G	0.1% DMSO	→	→	→	→	→	1 μ M OHT	→	→	100 μ M Dig	→	→
H	VC	→	→	→	→	→	PC	→	→	→	→	→

VC: Vehicle control (0.1% DMSO), PC: Positive control (1 nM E2), OHT :4-Hydroxytamoxifen, Dig: Digitonin.

 = spiked with 25pM E2


28. To evaluate the antagonist activity of chemicals, assay wells located in rows from A to G should be spiked with 25pM E2. The reference standards (Tamoxifen and Flutamide) should be tested in every run. PC wells treated with 1 nM of E2 that can be used as quality control of hER α -HeLa-9903 cell line, VC wells treated with DMSO (or appropriate solvent), 0,1 % DMSO wells treated with DMSO addition to the spiked E2 corresponding to "Spike-in-control", wells treated with final concentration 1 μ M OHT and wells treated with 100 μ M Dig should be included in each test assay plate (Table 5). Subsequent assay plate should follow the same plate layout without reference standards wells (Table 6). If cells from different sources (e.g. different passage number, different lot, etc.) are used in the same experiment, the reference standards should be tested for each cell source.

Table 6

Example of plate concentration assignment of test and plate control chemicals in the assay plate in ER antagonist assay

Row	Test Chemical 1			Test Chemical 2			Test Chemical 3			Test Chemical 4		
	1	2	3	4	5	6	7	8	9	10	11	12
A	conc 1 (10 μ M)	→	→	10 μ M	→	→	10 μ M	→	→	10 μ M	→	→
B	conc 2 (1 μ M)	→	→	1 μ M	→	→	1 μ M	→	→	1 μ M	→	→
C	conc 3 (100 nM)	→	→	100 nM	→	→	100 nM	→	→	100 nM	→	→
D	conc 4 (10 nM)	→	→	10 nM	→	→	10 nM	→	→	10 nM	→	→
E	conc 5 (1 nM)	→	→	1 nM	→	→	1 nM	→	→	1 nM	→	→
F	conc 6 (100 pM)	→	→	100 pM	→	→	100 pM	→	→	100 pM	→	→
G	0.1% DMSO	→	→	→	→	→	1 μ M OHT	→	→	100 μ M Dig	→	→
H	VC	→	→	→	→	→	PC	→	→	→	→	→

VC: Vehicle control (0.1% DMSO), PC: Positive control (1 nM E2), OHT: 4-Hydroxytamoxifen, Dig: Digitonin.

 : Spiked with 25pM E2

29. The lack of edge effects should be confirmed, as appropriate, and if edge effects are suspected, the plate layout should be altered to avoid such effects. For example, a plate layout excluding the edge wells can be employed.
30. After adding the chemicals, the assay plates should be incubated in a 5 % CO₂ incubator at 37±1°C for 20-24 hours to induce the reporter gene products.
31. Special considerations will need to be applied to those compounds that are highly volatile. In such cases, nearby control wells may generate false positives and this should be considered in light of expected and historical control values. In the few cases where volatility may be of concern, the use of "plate sealers" may help to effectively isolate individual wells during testing, and is therefore recommended in such cases.
32. Repeat definitive tests for the same chemical should be conducted on different days, to ensure independence.

Luciferase assay

33. A commercial luciferase assay reagent [e.g. Steady-Glo® Luciferase Assay System (Promega, E2510, or equivalent)] or a standard luciferase assay system (e.g. Promega, E1500, or equivalent) can be used for the assay, as long as the acceptability criteria are met. The assay reagents should be selected based on the sensitivity of the luminometer to be used. When using the standard luciferase assay system, Cell Culture Lysis Reagent (e.g. Promega, E1531, or equivalent) should be used before adding the substrate. The luciferase reagent should be applied following the manufacturers' instructions.

ANALYSIS OF DATA

ER agonist assay

34. In case of ER agonist assay, to obtain the relative transcriptional activity to PC (1 nM of E2), the luminescence signals from the same plate can be analysed according to the following steps (other equivalent mathematical processes are also acceptable):

Step 1. Calculate the mean value for the VC.

Step 2. Subtract the mean value of the VC from each well value to normalise the data.

Step 3. Calculate the mean for the normalised PC.

Step 4. Divide the normalised value of each well in the plate by the mean value of the normalised PC (PC=100 %).

The final value of each well is the relative transcriptional activity for that well compared to the PC response.

Step 5. Calculate the mean value of the relative transcriptional activity for each concentration group of the test chemical. There are two dimensions to the response: the averaged transcriptional activity (response) and the concentration at which the response occurs (see following section).

EC₅₀, PC₅₀ and PC₁₀ induction considerations

35. The full concentration-response curve is required for the calculation of the EC₅₀, but this may not always be achievable or practical due to limitations of the test concentration range (for example due to cytotoxicity or solubility problems). However, as the EC₅₀ and maximum induction level (corresponding to the top value of the Hill-equation) are informative parameters, these parameters should be reported where possible. For the calculation of EC₅₀ and maximum induction level, appropriate statistical software should be used (e.g. Graphpad Prism statistical software). If the Hill's logistic equation is applicable to the concentration response data, the EC₅₀ should be calculated by the following equation (7):

$$Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + 10^{\exp((\log \text{EC}_{50} - X) \times \text{Hill slope})}) \text{ Where:}$$

X is the logarithm of concentration; and,

Y is the response and Y starts at the Bottom and goes to the Top in a sigmoid curve. Bottom is fixed at zero in the Hill's logistic equation.

36. For each test chemical, the following should be provided:

The RPCMax which is the maximum level of response induced by a test chemical, expressed as a percentage of the response induced by 1 nM E2 on the same plate, as well as the PC_{Max} (concentration associated with the RPCMax); and

For positive chemicals, the concentrations that induce the PC10 and, if appropriate, the PC50.

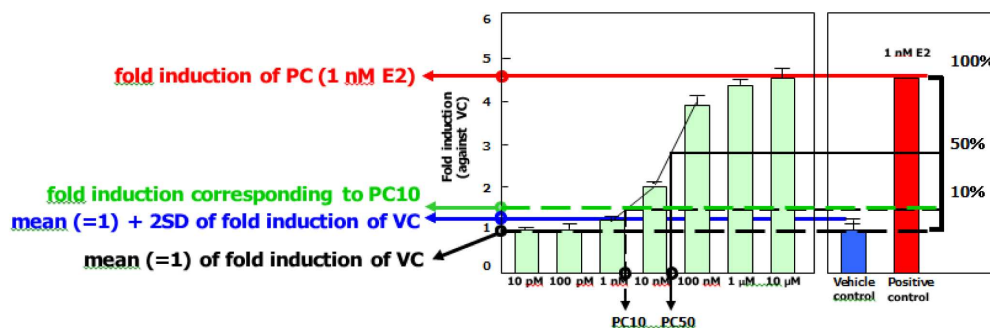
37. The PCx value can be calculated by interpolating between 2 points on the X-Y coordinate, one immediately above and one immediately below a PCx value. Where the data points lying immediately above and below the PCx value have the coordinates (a,b) and (c,d) respectively, then the PCx value may be calculated using the following equation:

$$\log[\text{PCx}] = \log[c] + (x-d)/(d-b)$$

38. Descriptions of PC values are provided in Figure 1 below.

Figure 1

Example of how to derive PC-values. The PC (1 nM of E2) is included on each assay plate



ER antagonist assay

39. In case of ER antagonist assay, to obtain the relative transcriptional activity (RTA) to spike in control (25 pM of E2), the luminescence signals from the same plate can be analysed according to the following steps (other equivalent mathematical processes are also acceptable):

Step 1. Calculate the mean value for the VC.

Step 2. Subtract the mean value of the VC from each well value to normalise the data. Step 3. Calculate the mean for the normalised spike in control.

Step 4. Divide the normalised value of each well in the plate by the mean value of the normalised spike in control (spike in control=100 %).

The final value of each well is the relative transcriptional activity for that well compared to the spike in control response.

Step 5. Calculate the mean value of the relative transcriptional activity for each treatment.

IC₃₀ and IC₅₀ induction considerations

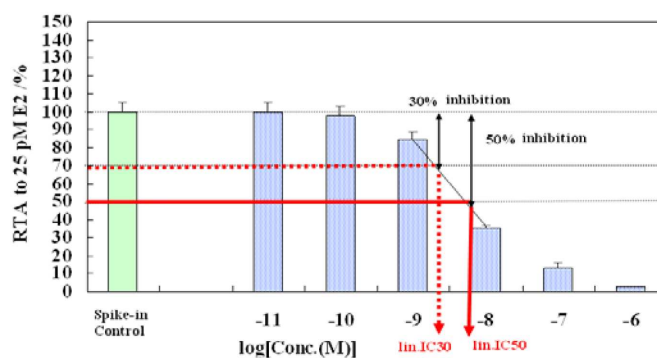
40. For positive chemicals, the concentrations that induce the IC30 and, if appropriate, the IC50 should be provided.

41. The IC_x value can be calculated by interpolating between 2 points on the X-Y coordinate, one immediately above and one immediately below a IC_x value. Where the data points lying immediately above and below the IC_x value have the coordinates (c,d) and (a,b) respectively, then the IC_x value may be calculated using the following equation:

$$\text{lin IC}_x = a - (b - (100 - x)) \cdot (a - c) / (b - d)$$

Figure 2

Example of how to derive IC-values. The spike in control (25 pM of E2) is included on each assay plate



RTA: relative transcriptional activity

42. The results should be based on two (or three) independent runs. If two runs give comparable and therefore reproducible results, it is not necessary to conduct a third run. To be acceptable, the results should:
- Meet the acceptability criteria (see Acceptability criteria para 14-20),
 - Be reproducible.

Data Interpretation Criteria

Table 7

Positive and negative decision criteria in ER agonist assay

Positive	If the RPCMax is obtained that is equal to or exceeds 10 % of the response of the positive control in at least two of two or two of three runs.
Negative	If the RPCMax fails to achieve at least 10 % of the response of the positive control in two of two or two of three runs.

Table 8

Positive and negative decision criteria in ER antagonist assay

Positive	If the IC30 is calculated in at least two of two or two of three runs.
Negative	If the IC30 fails to calculate in two of two or two of three runs.

43. Data interpretation criteria are shown in Tables 7 and 8. Positive results will be characterised by both the magnitude of the effect and the concentration at which the effect occurs. Expressing results as a concentration at which a 50 % (PC50) or 10 % (PC10) of PC values are reached for the agonist assay, and 50 % (IC50) or 30 % (IC30) of the spike-in control value is inhibited for the antagonist assay, accomplishes both of these goals. However, a test chemical is determined to be positive, if the maximum response induced by the test chemical (RPCMax) is equal to or exceeds 10 % of the response of the PC in at least two of two or two of three runs, while a test chemical is considered negative if the RPCMax fails to achieve at least 10 % of the response of the positive control in two of two or two of three runs.
44. The calculations of PC10, PC50 and PCMax in ER agonist assay and IC30 and IC50 in ER antagonist assay can be made by using a spreadsheet available with the Test Guideline on the OECD public website ⁽²⁾.
45. It should be sufficient to obtain PC10 or PC50 and IC30 or IC50 values at least twice. However, should the resulting base-line for data in the same concentration range show variability with an unacceptably high coefficient of variation (CV; %) the data may not be considered reliable and the source of the high variability should be identified. The CV of the raw data triplicates (i.e. luminescence intensity data) of the data points that are used for the calculation of PC10 should be less than 20 %.
46. Meeting the acceptability criteria indicates the assay system is operating properly, but it does not ensure that any particular run will produce accurate data. Duplicating the results of the first run is the best insurance that accurate data were produced.
47. In case of ER agonist assay, where more information is required in addition to the screening and prioritisation purposes of this TG for positive test chemicals, particularly for PC10-PC49 chemicals, as well as chemicals suspected to over-stimulate luciferase, it can be confirmed that the observed luciferase-activity is solely an ER α -specific response, using an ER α antagonist (see Appendix 2.1).

TEST REPORT

48. See paragraph 20 of "ER TA ASSAY COMPONENTS".

LITERATURE

- (1) OECD (2015). Report of the Inter-Laboratory Validation for Stably Transfected Transactivation Assay to detect Estrogenic and Anti-estrogenic Activity. Environment, Health and Safety Publications, Series on Testing and Assessment (No 225), Organisation for Economic Cooperation and Development, Paris.
- (2) Escande A., *et al.* (2006). Evaluation of Ligand Selectivity Using Reporter Cell Lines Stably Expressing Estrogen Receptor Alpha or Beta, *Biochem. Pharmacol.*, 71, 1459-1469.
- (3) Kuiper G.G., *et al.* (1998). Interaction of Estrogenic Chemicals and Phytoestrogens with Estrogen Receptor Beta, *Endocrinol.*, 139, 4252-4263.

⁽²⁾ <http://www.oecd.org/env/testguidelines>

-
- (4) Spaepen M., *et al.* (1992). Detection of Bacterial and Mycoplasma Contamination in Cell Cultures by Polymerase Chain Reaction, *FEMS Microbiol. Lett.*, 78(1), 89-94.
 - (5) Kobayashi H., *et al.* (1995). Rapid Detection of Mycoplasma Contamination in Cell Cultures by Enzymatic Detection of Polymerase Chain Reaction (PCR) Products, *J. Vet. Med. Sci.*, 57(4), 769- 71.
 - (6) Dussurget O. and Roulland-Dussoix D. (1994). Rapid, Sensitive PCR-Based Detection of Mycoplasmas in Simulated Samples of Animal Sera, *Appl. Environ. Microbiol.*, 60(3), 953-9.
 - (7) De Lean A., Munson P.J. and Rodbard D. (1978). Simultaneous Analysis of Families of Sigmoidal Curves: Application to Bioassay, Radioligand Assay, and Physiological Dose-Response Curves, *Am. J. Physiol.*, 235, E97-E102.

Appendix 2.1

FALSE POSITIVES: ASSESSMENT OF NON-RECEPTOR MEDIATED LUMINESCENCE SIGNALS

1. False positives in the ER agonist assay might be generated by non-ER-mediated activation of the luciferase gene, or direct activation of the gene product or unrelated fluorescence. Such effects are indicated by an incomplete or unusual dose-response curve. If such effects are suspected, the effect of an ER antagonist (e.g. 4- hydroxytamoxifen (OHT) at non-toxic concentration) on the response should be examined. The pure antagonist ICI 182780 may not be suitable for this purpose as a sufficient concentration of ICI 182780 may decrease the VC value, and this will affect the data analysis.
2. To ensure validity of this approach, the following needs to be tested in the same plate:
 - Agonistic activity of the unknown chemical with / without 10 μ M of OHT
 - VC (in triplicate)
 - OHT (in triplicate)
 - 1 nM of E2 (in triplicate) as agonist PC
 - 1 nM of E2 + OHT (in triplicate)

Data interpretation criteria

Note: All wells should be treated with the same concentration of the vehicle.

- If the agonistic activity of the unknown chemical is NOT affected by the treatment with ER antagonist, it is classified as "Negative".
- If the agonistic activity of the unknown chemical is completely inhibited, apply the decision criteria.
- If the agonistic activity at the lowest concentration is equal to, or is exceeding, PC10 response the unknown chemical is inhibited equal to or exceeding PC10 response. The difference in the responses between the non-treated and treated wells with the ER antagonist is calculated and this difference should be considered as the true response and should be used for the calculation of the appropriate parameters to enable a classification decision to be made.

Data analysis

Check the performance standard.

Check the CV between wells treated under the same conditions.

1. Calculate the mean of the VC
2. Subtract the mean of VC from each well value **not** treated with OHT
3. Calculate the mean of OHT
4. Subtract the mean of the VC from each well value treated with OHT
5. Calculate the mean of the PC
6. Calculate the relative transcriptional activity of all other wells relative to the PC.

Appendix 2.2

PREPARATION OF SERUM TREATED WITH DEXTRAN COATED CHARCOAL (DCC)

1. The treatment of serum with dextran-coated charcoal (DCC) is a general method for removal of estrogenic compounds from serum that is added to cell medium, in order to exclude the biased response associated with residual estrogens in serum. 500 ml of fetal bovine serum (FBS) can be treated by this procedure.

Components

2. The following materials and equipment will be required:

Materials

Activated charcoal

Dextran

Magnesium chloride hexahydrate ($MgCl_2 \cdot 6H_2O$)

Sucrose

1 M HEPES buffer solution (pH 7.4)

Ultrapure water produced from a filter system

Equipment

Autoclaved glass container (size should be adjusted as appropriate) General Laboratory Centrifuge (that can set temperature at 4 °C)

Procedure

3. The following procedure is adjusted for the use of 50 ml centrifuge tubes:

[Day-1] Prepare dextran-coated charcoal suspension with 1 l of ultrapure water containing 1,5 mM of $MgCl_2$, 0,25 M sucrose, 2,5 g of charcoal, 0,25 g dextran and 5 mM of HEPES and stir it at 4 °C, overnight.

[Day-2] Dispense the suspension in 50 ml centrifuge tubes and centrifuge at 10 000 rpm at 4 °C for 10 minutes. Remove the supernatant and store half of the charcoal sediment at 4 °C for the use on Day-3. Suspend the other half of the charcoal with FBS that has been gently thawed to avoid precipitation, and heat-inactivated at 56 °C for 30 minutes, then transfer into an autoclaved glass container such as an Erlenmeyer flask. Stir this suspension gently at 4 °C, overnight.

[Day-3] Dispense the suspension with FBS into centrifuge tubes for centrifugation at 10 000 rpm at 4 °C for 10 minutes. Collect FBS and transfer into the new charcoal sediment prepared and stored on Day-2. Suspend the charcoal sediment and stir this suspension gently in an autoclaved glass container at 4 °C, overnight.

[Day-4] Dispense the suspension for centrifugation at 10 000 rpm at 4 °C for 10 minutes and sterilise the supernatant by filtration through 0,2 µm sterile filter. This DCC treated FBS should be stored at -20 °C and can be used for up a year.

Appendix 3

VM7LUC ESTROGEN RECEPTOR TRANSACTIVATION ASSAY FOR IDENTIFYING ESTROGEN RECEPTOR AGONISTS AND ANTAGONISTS

INITIAL CONSIDERATIONS AND LIMITATIONS (SEE ALSO GENERAL INTRODUCTION)

1. This assay uses the VM7Luc4E2 cell line ⁽¹⁾. It has been validated by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), and the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (1). The VM7Luc cell lines predominantly express endogenous ER α and a minor amount of endogenous ER β (2) (3) (4).
2. This assay is applicable to a wide range of substances, provided they can be dissolved in dimethyl sulfoxide (DMSO; CASRN 67-68-5), do not react with DMSO or the cell culture medium, and are not cytotoxic at the concentrations being tested. If use of DMSO is not possible, another vehicle such as ethanol or water may be used (see paragraph 12). The demonstrated performance of the VM7Luc ER TA (ant)agonist assay suggests that data generated with this assay may inform upon ER mediated mechanisms of action and could be considered for prioritisation of substances for further testing.
3. This assay is specifically designed to detect hER and hER β -mediated TA by measuring chemiluminescence as the endpoint. Chemiluminescence use in bioassays is widespread because luminescence has a high signal-to-background ratio (10). However, the activity of firefly luciferase in cell-based assays can be confounded by substances that inhibit the luciferase enzyme, causing both apparent inhibition or increased luminescence due to protein stabilisation (10). In addition, in some luciferase-based ER reporter gene assays, non-receptor-mediated luminescence signals have been reported at phytoestrogen concentrations higher than 1 μ M due to the over-activation of the luciferase reporter gene (9) (11). While the dose-response curve indicates that true activation of the ER system occurs at lower concentrations, luciferase expression obtained at high concentrations of phytoestrogens or similar compounds suspected of producing phytoestrogen-like over-activation of the luciferase reporter gene needs to be examined carefully in stably transfected ER TA assay systems (see Appendix 2).
4. The "GENERAL INTRODUCTION" and "ER TA ASSAY COMPONENTS" should be read before using this assay for regulatory purposes. Definitions and abbreviations used in this test method are described in Appendix 1.

PRINCIPLE OF THE ASSAY (SEE ALSO GENERAL INTRODUCTION)

5. The assay is used to indicate ER ligand binding, followed by translocation of the receptor-ligand complex to the nucleus. In the nucleus, the receptor-ligand complex binds to specific DNA response elements and transactivates the reporter gene (*luc*), resulting in the production of luciferase and the subsequent emission of light, which can be quantified using a luminometer. Luciferase activity can be quickly and inexpensively evaluated with a number of commercially available kits. The VM7Luc ER TA utilises an ER responsive human breast adenocarcinoma cell line, VM7, which has been stably transfected with a firefly *luc* reporter construct under control of four estrogenresponse

⁽¹⁾ Before June 2016, this cell line was designated as BG1Luc cell line. BG-1 cells were originally described by Geisinger et al. (1998) (12) and were later characterized by researchers at the National Institute of Environmental Health Sciences (NIEHS) (13). Relatively recently, it was discovered that there exist two different variants of BG-1 cells being used by researchers, BG-1 Fr and BG-1 NIEHS. In-depth analysis, including DNA testing, of these two BG-1 variant cell lines carried out by Li and coworkers (2014) (14) showed that the BG-1 Fr was unique and that the BG-1 NIEHS, i.e. the original cell line used to develop the assay, was not the BG1 human ovarian carcinoma cell line, but was instead a variant of the MCF7 human breast cancer cell line. The cell line used in the assay, originally referred to as BG1Luc4E2 (15), will now be designated as VM7Luc4E2 ("V" = variant; "M7" = MCF7 cells). Likewise, the assay will now be designated as the VM7Luc ER TA. While this changes the origin of the cell line upon which the assay is based, it does not affect published validation studies nor the utility and application of this assay for screening of estrogenic/anti-estrogenic chemicals.

elements placed upstream of the mouse mammary tumour virus promoter (MMTV), to detect substances with *in vitro* ER agonist or antagonist activity. This MMTV promoter exhibits only minor cross-reactivity with other steroid and non-steroid hormones (8). Criteria for data interpretation are described in detail in paragraph 41. Briefly, a positive response is identified by a concentration-response curve containing at least three points with non-overlapping error bars (mean \pm SD), as well as a change in amplitude (normalised relative light unit [RLU]) of at least 20 % of the maximal value for the reference standard (17-estradiol [E2; CASRN 50-28-2] for the agonist assay, raloxifene HCl [Ral; CASRN 84449-90-1]/E2 for the antagonist assay).

PROCEDURE

Cell Line

6. The stably transfected VM7Luc4E2 cell line should be used for the assay. The cell line is currently only available with a technical licensing agreement from the University of California, Davis, California, USA ⁽²⁾, and from Xenobiotic Detection Systems Inc., Durham, North Carolina, USA ⁽³⁾.

Stability of the Cell Line

7. To maintain the stability and integrity of the cell line, the cells should be grown for more than one passage from the frozen stock in cell maintenance media (see paragraph 9). Cells should not be cultured for more than 30 passages. For the VM7Luc4E2 cell line, 30 passages will be approximately three months.

Cell Culture and Plating Conditions

8. Procedures specified in the Guidance on Good Cell Culture Practice (5) (6) should be followed to assure the quality of all materials and methods in order to maintain the integrity, validity, and reproducibility of any work conducted.
9. VM7Luc4E2 cells are maintained in RPMI 1640 medium supplemented with 0,9 % Pen-Strep and 8,0 % fetal bovine serum (FBS) in a dedicated tissue culture incubator at 37°C \pm 1°C, 90 % \pm 5 % humidity, and 5,0 % \pm 1 % CO₂/air.
10. Upon reaching ~80 % confluence, VM7Luc4E2 cells are subcultured and conditioned to an estrogen-free environment for 48 hours prior to plating the cells in 96-well plates for exposure to test chemicals and analysis of estrogen dependent induction of luciferase activity. The estrogen-free medium (EFM) contains Dulbecco's Modification of Eagle's Medium (DMEM) without phenol red, supplemented with 4.5 % charcoal/dextran-treated FBS, 1,9 % L-glutamine, and 0,9 % Pen-Strep. All plasticware should be free of estrogenic activity [see detailed protocol (7)].

Acceptability Criteria

11. Acceptance or rejection of a test is based on the evaluation of reference standard and control results from each experiment conducted on a 96-well plate. Each reference standard is tested in multiple concentrations and thereare

⁽²⁾ Michael S. Denison, Ph.D. Professor, Dept. of Environmental Toxicology, 4241 Meyer Hall, One Shields Ave, University of California, Davis, CA 95616, E: msdenison@ucdavis.edu, (530) 754-8649

⁽³⁾ Xenobiotic Detection Systems Inc. 1601 East Geer Street, Suite S, Durham NC, 27704 USA, email: info@dioxins.com, Telephone: 919-688-4804, Fax: 919-688-4404

multiple samples of each reference and control concentration. Results are compared to quality controls (QC) for these parameters that were derived from the agonist and antagonist historical databases generated by each laboratory during the demonstration of proficiency. The historical databases are updated with reference standard and control values on a continuous basis. Changes in equipment or laboratory conditions may necessitate generation of updated historical databases.

Agonist Test

Range Finder Test

- Induction: Plate induction should be measured by dividing the average highest E2 reference standard relative light unit (RLU) value by the average DMSO control RLU value. Five-fold induction is usually achieved, but for purpose of acceptance, induction should be greater than or equal to four-fold.
- DMSO control results: Solvent control RLU values should be within 2.5 times the standard deviation of the historical solvent control mean RLU value.
- An experiment that fails either acceptance criterion should be discarded and repeated.

Comprehensive Test

It includes acceptability criteria from the agonist range finder test and the following:

- Reference standard results: The E2 reference standard concentration-response curve should be sigmoidal in shape and have at least three values within the linear portion of the concentration-response curve.
- Positive control results: Methoxychlor control RLU values should be greater than the DMSO mean plus three times the standard deviation from the DMSO mean.
- An experiment that fails any single acceptance criterion should be discarded and repeated.

Antagonist Test

Range Finder Test

- Reduction: Plate reduction is measured by dividing the average highest Ral/E2 reference standard RLU value by the average DMSO control RLU value. Five-fold reduction is usually achieved, but for the purposes of acceptance, reduction should be greater than or equal to three-fold.
- E2 control results: E2 control RLU values should be within 2.5 times the standard deviation of the historical E2 control mean RLU value.
- DMSO control results: DMSO control RLU values should be within 2.5 times the standard deviation of the historical solvent control mean RLU value.

- An experiment that fails any single acceptance criterion will be discarded and repeated.

Comprehensive Test

It includes acceptance criteria from the antagonist range finder test and the following:

- Reference standard results: The Ral/E2 reference standard concentration-response curve should be sigmoidal in shape and have at least three values within the linear portion of the concentration-response curve.
- Positive control results: Tamoxifen/E2 control RLU values should be less than the E2 control mean minus three times the standard deviation from the E2 control mean.
- An experiment that fails any single acceptance criterion will be discarded and repeated.

Reference Standards, Positive, and Vehicle Controls

Vehicle Control (Agonist and Antagonist Assays)

12. The vehicle that is used to dissolve the test chemicals should be tested as a vehicle control. The vehicle used during the validation of the VM7Luc ER TA assay was 1 % (v/v) dimethylsulfoxide (DMSO, CASRN 67-68-5) (see paragraph 24). If a vehicle other than DMSO is used, all reference standards, controls, and test chemicals should be tested in the same vehicle, if appropriate.

Reference Standard (Agonist Range Finder)

13. The reference standard is E2 (CASRN 50-28-2). For range finder testing, the reference standard is comprised of a serial dilution of four concentrations of E2 (1.84×10^{-10} , 4.59×10^{-11} , 1.15×10^{-11} and 2.87×10^{-12} M), with each concentration tested in duplicate wells.

Reference Standard (Agonist Comprehensive)

14. E2 for comprehensive testing is comprised of a 1:2 serial dilution consisting of 11 concentrations (ranging from 3.67×10^{-10} to 3.59×10^{-13} M) of E2 in duplicate wells.

Reference Standard (Antagonist Range Finder)

15. The reference standard is a combination of Ral (CASRN 84449-90-1) and E2 (CASRN 50-28-2). Ral/E2 for range finder testing is comprised of a serial dilution of three concentrations of Ral (3.06×10^{-9} , 7.67×10^{-10} , and 1.92×10^{-10} M) plus a fixed concentration (9.18×10^{-11} M) of E2 in duplicate wells.

Reference Standard (Antagonist Comprehensive)

16. Ral/E2 for comprehensive testing is comprised of a 1:2 serial dilution of Ral (ranging from 2.45×10^{-8} to 9.57×10^{-11} M) plus a fixed concentration (9.18×10^{-11} M) of E2 consisting of nine concentrations of Ral/E2 in duplicate wells.

Weak Positive Control (Agonist)

17. The weak positive control is $9,06 \times 10^{-6}$ M *p,p'*-methoxychlor (methoxychlor; CASRN 72-43-5) in EFM.

Weak Positive Control (Antagonist)

18. The weak positive control consists of tamoxifen (CASRN 10540-29-1) $3,36 \times 10^{-6}$ M with $9,18 \times 10^{-11}$ M E2 in EFM.

E2 Control (Antagonist Assay Only)

19. The E2 control is $9,18 \times 10^{-11}$ M E2 in EFM and used as a base line negative control.

Fold-Induction (Agonist)

20. The induction of luciferase activity of the reference standard (E2) is measured by dividing the average highest E2 reference standard RLU value by the average DMSO control RLU value, and the result should be greater than four-fold.

Fold-Reduction (Antagonist)

21. The mean luciferase activity of the reference standard (Ral/E2) is measured by dividing the average highest Ral/E2 reference standard RLU value by the average DMSO control RLU value and should be greater than three-fold.

Demonstration of Laboratory Proficiency (see paragraph 14 and Tables 3 and 4 in "ER TA ASSAY COMPONENTS" of this test method)

Vehicle

22. Test chemicals should be dissolved in a solvent that solubilises the test chemical and is miscible with the cell medium. Water, ethanol (95 % to 100 % purity) and DMSO are suitable vehicles. If DMSO is used, the level should not exceed 1 % (v/v). For any vehicle, it should be demonstrated that the maximum volume used is not cytotoxic and does not interfere with the assay performance. Reference standards and controls are dissolved in 100 % solvent and then diluted down to appropriate concentrations in EFM.

Preparation of Test chemicals

23. The test chemicals are dissolved in 100 % DMSO (or appropriate solvent), and then diluted down to appropriate concentrations in EFM. All test chemicals should be allowed to equilibrate to room temperature before being dissolved and diluted. Test chemical solutions should be prepared fresh for each experiment. Solutions should not have noticeable precipitate or cloudiness. Reference standard and control stocks may be prepared in bulk; however, final reference standard, control dilutions and test chemicals should be freshly prepared for each experiment and used within 24 hours of preparation.

Solubility and Cytotoxicity: Considerations for Range Finding

24. Range finder testing consists of seven point - 1:10 serial dilutions run in duplicate. Initially, test chemicals are tested up to the maximum concentration of 1 mg/ml (~1 mM) for agonist testing and 20 µg/ml (~10 M) for antagonist testing. Range finder experiments are used to determine the following:

- Test chemical starting concentrations to be used during comprehensive testing
- Test chemical dilutions (1:2 or 1:5) to be used during comprehensive testing

25. An assessment of cell viability/cytotoxicity is included in the agonist and antagonist assay protocols (7) and is incorporated into range finder and comprehensive testing. The cytotoxicity method that was used to assess cell viability during the validation of the VM7Luc ER TA (1) was a scaled qualitative visual observation method; however, a quantitative method for the determination of cytotoxicity can be used (see protocol (7)). Data from test chemical concentrations that cause more than 20 % reduction in viability cannot be used.

Test chemical Exposure and Assay Plate Organisation

26. Cells are counted and plated into 96-well tissue culture plates (2×10^5 cells per well) in EFM and incubated for 24 hours to allow the cells to attach to the plate. The EFM is removed and replaced with test and reference chemicals in EFM and incubated for 19-24 hours. Special considerations will need to be applied to those substances that are highly volatile since nearby control wells may generate false positive results. In such cases, "plate sealers" may help to effectively isolate individual wells during testing, and are therefore recommended.

Range Finder Tests

27. Range finder testing uses all wells of the 96-well plate to test up to six test chemicals as seven point 1:10 serial dilutions in duplicate (see Figures 1 and 2).

— Agonist range finder testing uses four concentrations of E2 in duplicate as the reference standard and four replicate wells for the DMSO control.

— Antagonist range finder testing uses three concentrations of Ral/E2 with $9,18 \times 10^{-11}$ M E2 in duplicate as the reference standard, with three replicate wells for the E2 and DMSO controls.

Figure 1

Agonist Range Finder Test 96-well Plate Layout

	1	2	3	4	5	6	7	8	9	10	11	12
A	TC1-1	TC1-1	TC2-1	TC2-1	TC3-1	TC3-1	TC4-1	TC4-1	TC5-1	TC5-1	TC6-1	TC6-1
B	TC1-2	TC1-2	TC2-2	TC2-2	TC3-2	TC3-2	TC4-2	TC4-2	TC5-2	TC5-2	TC6-2	TC6-2
C	TC1-3	TC1-3	TC2-3	TC2-3	TC3-3	TC3-3	TC4-3	TC4-3	TC5-3	TC5-3	TC6-3	TC6-3
D	TC1-4	TC1-4	TC2-4	TC2-4	TC3-4	TC3-4	TC4-4	TC4-4	TC5-4	TC5-4	TC6-4	TC6-4
E	TC1-5	TC1-5	TC2-5	TC2-5	TC3-5	TC3-5	TC4-5	TC4-5	TC5-5	TC5-5	TC6-5	TC6-5
F	TC1-6	TC1-6	TC2-6	TC2-6	TC3-6	TC3-6	TC4-6	TC4-6	TC5-6	TC5-6	TC6-6	TC6-6
G	TC1-7	TC1-7	TC2-7	TC2-7	TC3-7	TC3-7	TC4-7	TC4-7	TC5-7	TC5-7	TC6-7	TC6-7
H	E2-1	E2-2	E2-3	E2-4	VC	VC	VC	VC	E2-1	E2-2	E2-3	E2-4

Abbreviations: E2-1 to E2-4 = concentrations of the E2 reference standard (from high to low); TC1-1 to TC1-7 = concentrations (from high to low) of test chemical 1 (TC1); TC2-1 to TC2-7 = concentrations (from high to low) of test chemical 2 (TC2); TC3-1 to TC3-7 = concentrations (from high to low) of test chemical 3 (TC3); TC4-1 to TC4-7 = concentrations (from high to low) of test chemical 4 (TC4); TC5-1 to TC5-7 = concentrations (from high to low) of test chemical 5 (TC5); TC6-1 to TC6-7 = concentrations (from high to low) of test chemical 6 (TC6); VC = vehicle control (DMSO [1 % v/v EFM.]).

Figure 2

Antagonist Range Finder Test 96-well Plate Layout

	1	2	3	4	5	6	7	8	9	10	11	12
A	TC1-1	TC1-1	TC2-1	TC2-1	TC3-1	TC3-1	TC4-1	TC4-1	TC5-1	TC5-1	TC6-1	TC6-1
B	TC1-2	TC1-2	TC2-2	TC2-2	TC3-2	TC3-2	TC4-2	TC4-2	TC5-2	TC5-2	TC6-2	TC6-2
C	TC1-3	TC1-3	TC2-3	TC2-3	TC3-3	TC3-3	TC4-3	TC4-3	TC5-3	TC5-3	TC6-3	TC6-3
D	TC1-4	TC1-4	TC2-4	TC2-4	TC3-4	TC3-4	TC4-4	TC4-4	TC5-4	TC5-4	TC6-4	TC6-4
E	TC1-5	TC1-5	TC2-5	TC2-5	TC3-5	TC3-5	TC4-5	TC4-5	TC5-5	TC5-5	TC6-5	TC6-5
F	TC1-6	TC1-6	TC2-6	TC2-6	TC3-6	TC3-6	TC4-6	TC4-6	TC5-6	TC5-6	TC6-6	TC6-6
G	TC1-7	TC1-7	TC2-7	TC2-7	TC3-7	TC3-7	TC4-7	TC4-7	TC5-7	TC5-7	TC6-7	TC6-7
H	Ral-1	Ral-2	Ral-3	VC	VC	VC	E2	E2	E2	Ral-1	Ral-2	Ral-3

Abbreviations: E2 = E2 control; Ral-1 to Ral-3 = concentrations of the Raloxifene/E2 reference standard (from high to low); TC1-1 to TC1-7 = concentrations (from high to low) of test chemical 1 (TC1); TC2-1 to TC2-7 = concentrations (from high to low) of test chemical 2 (TC2); TC3-1 to TC3-7 = concentrations (from high to low) of test chemical 3 (TC3); TC4-1 to TC4-7 = concentrations (from high to low) of test chemical 4 (TC4); TC5-1 to TC5-7 = concentrations (from high to low) of test chemical 5 (TC5); TC6-1 to TC6-7 = concentrations (from high to low) of test chemical 6 (TC6); VC = vehicle control (DMSO [1 % v/v EFM]).

Note: All test chemicals are tested in the presence of $9,18 \times 10^{-11}$ M E2.28.

The recommended final volume of media required for each well is 200 μ l. Only use test plates in which the cells in all wells give a viability of 80 % and above.

29. Determination of starting concentrations for comprehensive **agonist** testing is described in depth in the agonist protocol (7). Briefly, the following criteria are used:

- If there are no points on the test chemical concentration curve that are greater than the mean plus three times the standard deviation of the DMSO control, comprehensive testing will be conducted using an 11-point 1:2 serial dilution starting at the maximum soluble concentration.
- If there are points on the test chemical concentration curve that are greater than the mean plus three times the standard deviation of the DMSO control, the starting concentration to be used for the 11-point dilution scheme in comprehensive testing should be one log higher than the concentration giving the highest adjusted RLU value in the range finder. The 11-point dilution scheme will be based on either 1:2 or 1:5 dilutions according to the following criteria:

An 11-point 1:2 serial dilution should be used if the resulting concentration range will encompass the full range of responses based on the concentration response curve generated in the range finder test. Otherwise, use a 1:5 dilution.

- If a test chemical exhibits a biphasic concentration response curve in the range finder test, both phases should also be resolved in comprehensive testing.

30. Determination of starting concentrations for comprehensive **antagonist** testing is described in depth in the antagonist protocol (7). Briefly, the following criteria are used:

- If there are no points on the test chemical concentration curve that are less than the mean minus three times the standard deviation of the E2 control, comprehensive testing will be conducted using an 11-point 1:2 serial dilution starting at the maximum soluble concentration.

- If there are points on the test chemical concentration curve that are less than the mean minus three times the standard deviation of the E2 control, the starting concentration to be used for the 11-point dilution scheme in comprehensive testing should be one of the following:
 - The concentration giving the lowest adjusted RLU value in the range finder
 - The maximum soluble concentration (See antagonist protocol (7), Figure 14-2)
 - The lowest cytotoxic concentration (See antagonist protocol (7), Figure 14-3 for a related example).
- The 11-point dilution scheme will be based on either a 1:2 or 1:5 serial or dilution according to the following criteria:

An 11-point 1:2 serial dilution should be used if the resulting concentration range will encompass the full range of responses based on the concentration response curve generated in the range finder test. Otherwise a 1:5 dilution should be used.

Comprehensive Tests

31. Comprehensive testing consists of 11-point serial dilutions (either 1:2 or 1:5 serial dilutions based on the starting concentration for comprehensive testing criteria) with each concentration tested in triplicate wells of the 96-well plate (see Figures 3 and 4).
- *Agonist* comprehensive testing uses 11 concentrations of E2 in duplicate as the reference standard. Four replicate wells for the DMSO control and four replicate wells for the methoxychlor control (9.06×10^{-6} M) are included on each plate.
 - *Antagonist* comprehensive testing uses nine concentrations of Ral/E2 with $9,18 \times 10^{-11}$ M E2 in duplicate as the reference standard, with four replicate wells for the E2 $9,18 \times 10^{-11}$ M control, four replicate wells for DMSO controls, and four replicate wells for tamoxifen $3,36 \times 10^{-6}$ M.

Figure 3

Agonist Comprehensive Test 96-well Plate Layout

	1	2	3	4	5	6	7	8	9	10	11	12
A	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	TC1-9	TC1-10	TC1-11	VC
B	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	TC1-9	TC1-10	TC1-11	VC
C	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	TC1-9	TC1-10	TC1-11	VC
D	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	TC2-9	TC2-10	TC2-11	VC
E	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	TC2-9	TC2-10	TC2-11	Meth
F	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	TC2-9	TC2-10	TC2-11	Meth
G	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8	E2-9	E2-10	E2-11	Meth
H	E2-1	E2-2	E2-3	E2-4	E2-5	E2-6	E2-7	E2-8	E2-9	E2-10	E2-11	Meth

Abbreviations: TC1-1 to TC1-11 = concentrations (from high to low) of test chemical 1; TC2-1 to TC2-11 = concentrations (from high to low) of test chemical 2; E2-1 to E2-11 = concentrations of the E2 reference standard (from high to low); Meth = p,p' methoxychlor weak positive control; VC = DMSO (1 % v/v) EFM vehicle control

Figure 4

Antagonist Comprehensive Test 96-well Plate Layout

	1	2	3	4	5	6	7	8	9	10	11	12
A	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	TC1-9	TC1-10	TC1-11	VC
B	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	TC1-9	TC1-10	TC1-11	VC
C	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	TC1-9	TC1-10	TC1-11	VC
D	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	TC2-9	TC2-10	TC2-11	VC
E	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	TC2-9	TC2-10	TC2-11	Tam
F	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	TC2-9	TC2-10	TC2-11	Tam
G	Ral-1	Ral-2	Ral-3	Ral-4	Ral-5	Ral-6	Ral-7	Ral-8	Ral-9	E2	E2	Tam
H	Ral-1	Ral-2	Ral-3	Ral-4	Ral-5	Ral-6	Ral-7	Ral-8	Ral-9	E2	E2	Tam

Abbreviations: E2 = E2 control; Ral-1 to Ral-9 = concentrations of the Raloxifene/E2 reference standard (from high to low); Tam = Tamoxifen/E2 weak positive control; TC1-1 to TC1-11 = concentrations (from high to low) of test chemical 1 (TC1); TC2-1 to TC2-11 = concentrations (from high to low) of test chemical 2 (TC2); VC = vehicle control (DMSO [1 % v/v EFM.]).

Note: As noted, all reference and test wells contain a fixed concentration of E2 (9.18×10^{-11} M)³².

Repeat comprehensive tests for the same chemical should be conducted on different days, to ensure independence. At least two comprehensive tests should be conducted. If the results of the tests contradict each other (e.g. one test is positive, the other negative), or if one of the tests is inadequate, a third additional test should be conducted.

Measure of Luminescence

33. Luminescence is measured in the range of 300 to 650 nm, using an injecting luminometer and with software that controls the injection volume and measurement interval (7). Light emission from each well is expressed as RLU per well.

ANALYSIS OF DATA**EC₅₀ /IC₅₀ determination**

34. The EC₅₀ value (half maximal effective concentration of a test chemical [agonists]) and the IC₅₀ value (half maximal inhibitory concentration of a test chemical [antagonists]) are determined from the concentration-response data. For test chemicals that are positive at one or more concentrations, the concentration of test chemical that causes a half-maximal response (IC₅₀ or EC₅₀) is calculated using a Hill function analysis or an appropriate alternative. The Hill function is a four-parameter logistic mathematical model relating the test chemical concentration to the response (typically following a sigmoidal curve) using the equation below:

$$Y = Bottom + \frac{(Top - Bottom)}{1 + 10^{(lgEC_{50} - X)Hill\ Slope}}$$

Where:

Y= response (i.e. RLUs);

X= the logarithm of concentration;

Bottom= the minimum response;

Top= the maximum response;

lg EC₅₀ (or lg IC₅₀)= the logarithm of X as the response midway between Top and Bottom;

Hillslope= the steepness of the curve.

The model calculates the best fit for the Top, Bottom, Hillslope, and IC₅₀ and EC₅₀ parameters. For the calculation of EC₅₀ and IC₅₀ values, appropriate statistical software should be used (e.g. Graphpad Prism^R statistical software).

Determination of Outliers

35. Good statistical judgment could be facilitated by including (but not limited to) the Q-test (see agonist and antagonist protocols (7) for determining “unusable” wells that will be excluded from the data analysis.
36. For E2 reference standard replicates (sample size of two), any adjusted RLU value for a replicate at a given concentration of E2 is considered an outlier if its value is more than 20 % above or below the adjusted RLU value for that concentration in the historical database.

Collection and Adjustment of Luminometer Data for Range Finder Testing

37. Raw data from the luminometer should be transferred to a spreadsheet template designed for the assay. It should be determined whether there are outlier data points that need to be removed. (See Test Acceptance Criteria for parameters that are determined in the analyses.) The following calculations should be performed:

Agonist

Step 1 Calculate the mean value for the DMSO vehicle control (VC).

Step 2 Subtract the mean value of the DMSO VC from each well value to normalise the data.

Step 3 Calculate the mean fold induction for the reference standard (E2).

Step 4 Calculate the mean EC₅₀ value for the test chemicals.

Antagonist

Step 1 Calculate the mean value for the DMSO VC.

Step 2 Subtract the mean value of the DMSO VC from each well value to normalise the data.

Step 3 Calculate the mean fold reduction for the reference standard (R_{al}/E2).

Step 4 Calculate the mean value for the E2 reference standard.

Step 5 Calculate the mean IC₅₀ value for the test chemicals.

Collection and Adjustment of Luminometer Data for Comprehensive Testing

38. Raw data from the luminometer should be transferred to a spreadsheet template designed for the assay. It should be determined whether there are outlier data points that need to be removed. (See Test Acceptance Criteria for parameters that are determined in the analyses.) The following calculations are performed:

Agonist

Step 1 Calculate the mean value for the DMSO VC.

Step 2 Subtract the mean value of the DMSO VC from each well value to normalise the data.

Step 3 Calculate the mean fold induction for the reference standard (E2).

Step 4 Calculate the mean EC₅₀ value for E2 and the test chemicals.

Step 5 Calculate the mean adjusted RLU value for methoxychlor.

Antagonist

Step 1 Calculate the mean value for the DMSO VC.

Step 2 Subtract the mean value of the DMSO VC from each well value to normalise the data.

Step 3 Calculate the mean fold induction for the reference standard (Ral/E2).

Step 4 Calculate the mean IC₅₀ value for Ral/E2 and the test chemicals.

Step 5 Calculate the mean adjusted RLU value for tamoxifen.

Step 6 Calculate the mean value for the E2 reference standard.

Data Interpretation Criteria

39. The VM7Luc ER TA is intended as part of a weight of evidence approach to help prioritise substances for ED testing *in vivo*. Part of this prioritisation procedure will be the classification of the test chemical as positive or negative for either ER agonist or antagonist activity. The positive and negative decision criteria used in the VM7Luc ER TA validation study are described in Table 1.

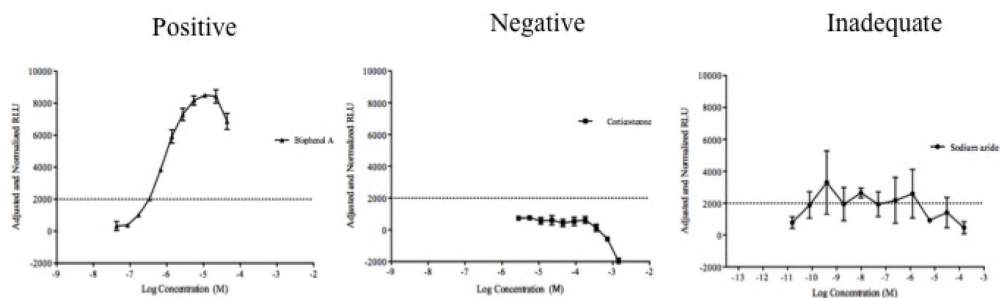
Table 1
Positive and Negative Decision Criteria

AGONIST ACTIVITY	
Positive	<ul style="list-style-type: none"> — All test chemicals classified as <i>positive for ER agonist activity</i> should have a concentration–response curve consisting of a baseline, followed by a positive slope, and concluding in a plateau or peak. In some cases, only two of these characteristics (baseline–slope or slope–peak) may be defined. — The line defining the positive slope should contain at least three points with non-overlapping error bars (mean \pm SD). Points forming the baseline are excluded, but the linear portion of the curve may include the peak or first point of the plateau. — A positive classification requires a response amplitude, the difference between baseline and peak, of at least 20 % of the maximal value for the reference standard, E2 (i.e. 2000 RLU or more when the maximal response value of the reference standards [E2] is adjusted to 10,000 RLU). — If possible, an EC₅₀ value should be calculated for each positive test chemical.
Negative	The average adjusted RLU for a given concentration is at or below the mean DMSO control RLU value plus three times the standard deviation of the DMSO RLU.
Inadequate	Data that cannot be interpreted as valid for showing either the presence or absence of activity because of major qualitative or quantitative limitations are considered inadequate and cannot be used to determine whether the test chemical is positive or negative. Chemicals should be retested.
ANTAGONIST ACTIVITY	
Positive	<ul style="list-style-type: none"> — Test chemical data produce a concentration-response curve consisting of a baseline, which is followed by a negative slope. — The line defining the negative slope should contain at least three points with non-overlapping error bars; points forming the baseline are excluded but the linear portion of the curve may include the first point of the plateau. — There should be at least a 20 % reduction in activity from the maximal value for the reference standard, Ral/E2 (i.e. 8000 RLU or less when the maximal response value of the reference standard [Ral/E2] is adjusted to 10,000 RLU). — The highest non-cytotoxic concentrations of the test chemical should be less than or equal to 1×10^{-5} M. — If possible, an IC₅₀ value should be calculated for each positive test chemical.
Negative	All data points are above the EC ₈₀ value (80 % of the E2 response, or 8000 RLU), at concentrations less than $1,0 \times 10^{-5}$ M.
Inadequate	Data that cannot be interpreted as valid for showing either the presence or absence of activity because of major qualitative or quantitative limitations are considered inadequate and cannot be used to determine whether the test chemical is positive or negative. Chemical should be retested.

40. Positive results will be characterised by both the magnitude of the effect and the concentration at which the effect occurs, where possible. Examples of positive, negative and inadequate data are shown in Figures 5 and 6.

Figure 5

Agonist Examples: Positive, Negative and Inadequate Data

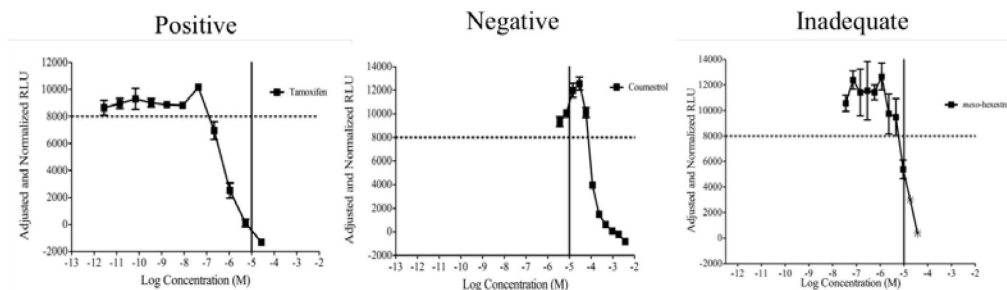


Dashed line indicates 20% of E2 response, 2000 adjusted and normalised RLUs.

Dashed line indicates 20 % of E2 response, 2000 adjusted and normalised RLUs.

Figure 6

Antagonist Examples: Positive, Negative, and Inadequate Data



Dashed line indicates 80 % of Ral/E2 response, 8000 adjusted and normalised RLUs.

Solid line indicates $1,00 \cdot 10^{-5}$ M. For a response to be considered positive, it should be below the 8000 RLU line, and at concentrations less than $1,00 \cdot 10^{-5}$ M.

Asterisked concentrations in the meso-hexestrol graph indicate viability scores of "2" or greater.

The test results for meso-hexestrol are considered inadequate data because the only response that is below 8,000 RLU occurs at $1,00 \cdot 10^{-5}$ M.

41. The calculations of EC_{50} and IC_{50} can be made using a four-parameter Hill Function (see agonist protocol and antagonist protocol for more details (7)). Meeting the acceptability criteria indicates the system is operating properly, but it does not ensure that any particular run will produce accurate data. Duplicating the results of the first run is the best assurance that accurate data were produced (see paragraph 19 of "ER TA ASSAY COMPONENTS").

TEST REPORT

42. See paragraph 20 of "ER TA ASSAY COMPONENTS".

LITERATURE

- (1) ICCVAM. (2011). ICCVAM Test Method Evaluation Report on the LUMI-CELL® ER (BG1Luc ER TA) Test Method: An *In Vitro* Method for Identifying ER Agonists and Antagonists, National Institute of Environmental Health Sciences: Research Triangle Park, NC.
- (2) Monje P., Boland R. (2001). Subcellular Distribution of Native Estrogen Receptor α and β Isoforms in Rabbit Uterus and Ovary, *J. Cell Biochem.*, 82(3): 467-479.
- (3) Pujol P., *et al.* (1998). Differential Expression of Estrogen Receptor-Alpha and -Beta Messenger RNAs as a Potential Marker of Ovarian Carcinogenesis, *Cancer Res.*, 58(23): 5367-5373.
- (4) Weihua Z., *et al.* (2000). Estrogen Receptor (ER) β , a Modulator of ER α in the Uterus, *Proceedings of the National Academy of Sciences of the United States of America* 97(11): 936-941.
- (5) Balls M., *et al.* (2006). The Importance of Good Cell Culture Practice (GCCP), *ALTEX*, 23(Suppl): p. 270-273.
- (6) Coecke S., *et al.* (2005). Guidance on Good Cell Culture Practice: a Report of the Second ECVAM Task Force on Good Cell Culture Practice, *Alternatives to Laboratory Animals*, 33: p. 261-287.
- (7) ICCVAM (2011). ICCVAM Test Method Evaluation Report, The LUMI-CELL® ER (BG1Luc ER TA) Test Method: An *In Vitro* Assay for Identifying Human Estrogen Receptor Agonist and Antagonist Activity of Chemicals, NIH Publication No 11-7850.
- (8) Rogers J.M., Denison M.S. (2000). Recombinant Cell Bioassays for Endocrine Disruptors: Development of a Stably Transfected Human Ovarian Cell Line for the Detection of Estrogenic and Anti-Estrogenic Chemicals, *In Vitro Mol. Toxicol.*, 13(1):67-82.
- (9) Escande A., *et al.* (2006). Evaluation of Ligand Selectivity Using Reporter Cell Lines Stably Expressing Estrogen Receptor Alpha or Beta, *Biochem. Pharmacol.*, 71(10):1459-69.
- (10) Thorne N., Inglese J., Auld D.S. (2010). Illuminating Insights into Firefly Luciferase and Other Bioluminescent Reporters Used in Chemical Biology, *Chemistry and Biology*, 17(6):646-57.
- (11) Kuiper G.G., *et al.* (1998). Interaction of Estrogenic Chemicals and Phytoestrogens with Estrogen Receptor Beta, *Endocrinology*, 139(10):4252-63.

-
- (12) Geisinger, *et al.* (1989). Characterization of a human ovarian carcinoma cell line with estrogen and progesterone receptors, *Cancer* 63, 280-288.
- (13) Baldwin, *et al.* (1998). BG-1 ovarian cell line: an alternative model for examining estrogen-dependent growth *in vitro*, *In Vitro Cell. Dev. Biol. – Animal*, 34, 649-654.
- (14) Li, Y., *et al.* (2014). Research resource: STR DNA profile and gene expression comparisons of human BG-1 cells and a BG-1/MCF-7 clonal variant, *Mol. Endo.* 28, 2072-2081.
- (15) Rogers, J.M. and Denison, M.S. (2000). Recombinant cell bioassays for endocrine disruptors: development of a stably transfected human ovarian cell line for the detection of estrogenic and anti-estrogenic chemicals, *In Vitro & Molec. Toxicol.* 13, 67-82.

*Appendix 4*STABLY TRANSFECTED HUMAN ESTROGEN RECEPTOR- α TRANSACTIVATION ASSAY FOR DETECTION OF ESTROGENIC AGONIST AND ANTAGONIST ACTIVITY OF CHEMICALS USING THE ER α CALUX CELL LINE

INITIAL CONSIDERATIONS AND LIMITATIONS (SEE ALSO GENERAL INTRODUCTION)

1. The ER α CALUX transactivation assay uses the human U2OS cell line to detect estrogenic agonist and antagonist activity mediated through human estrogen receptor alpha (hER α). The validation study of the stably transfected ER α CALUX bioassay by BioDetection Systems BV (Amsterdam, the Netherlands) demonstrated the relevance and reliability of the assay for its intended purpose (1). The ER α CALUX cell line expresses stably transfected human ER α only (2) (3).
2. This assay is specifically designed to detect hER α -mediated transactivation by measuring bioluminescence as the endpoint. The use of bioluminescence is commonly used in bioassays because of the high signal-to-noise ratio (4).
3. Phytoestrogen concentrations higher than 1 μ M have been reported to over-activate the luciferase reporter gene, resulting in non-receptor-mediated luminescence (5) (6) (7). Therefore, higher concentrations of phytoestrogens or other similar compounds that can over-activate the luciferase expression, have to be examined carefully in stably transfected ER transactivation assays (see Appendix 2).
4. The "GENERAL INTRODUCTION" and "ER TA ASSAY COMPONENTS" should be read before using this assay for regulatory purposes. Definitions and abbreviations used in this test method are described in Appendix 1.

PRINCIPLE OF THE ASSAY (SEE ALSO GENERAL INTRODUCTION)

5. The bioassay is used to assess ER ligand binding and subsequent translocation of the receptor-ligand complex to the nucleus. In the nucleus, the receptor-ligand complex binds specific DNA response elements and transactivates a firefly luciferase reporter gene, resulting in increased cellular expression of the luciferase enzyme. Following the addition of the luciferase substrate luciferine, the luciferine is transformed into a bioluminescent product. The light produced can easily be detected and quantified using a luminometer.
6. The test system utilises stably transfected ER α CALUX cells. ER α CALUX cells originated from the human osteoblastic osteosarcoma U2OS cell line. Human U2OS cells were stably transfected with 3xHRE-TATA-Luc and pSG5-neo-hER α using the calcium phosphate co-precipitation method. The U2OS cell line was selected as the best candidate to serve as the estrogen- (and other steroid hormone) responsive reporter cell line, based on the observation that the U2OS cell line showed little or no endogenous receptor activity. The absence of endogenous receptors was assessed using luciferase reporter plasmids only, showing no activity when receptor ligands were added. Furthermore, this cell line supported strong hormone-mediated responses when cognate receptors were transiently introduced (2) (3) (8).
7. Testing chemicals for estrogenic or anti-estrogenic activity using the ER α CALUX cell line include a prescreen run and comprehensive runs. During the prescreen run, the solubility, cytotoxicity and a refined concentration-range of test chemicals for comprehensive testing are determined. During the comprehensive runs, the refined concentration-ranges of test chemicals are tested in the ER α CALUX bioassays followed by the classification of the test chemicals for agonism or antagonism.

- Criteria for data interpretation are described in detail in paragraph 59. Briefly, a test chemical is considered positive for agonism in case at least two consecutive concentrations of the test chemical show a response that is equal or higher than 10 % of the maximum response of the reference standard 17 β -estradiol (PC₁₀). A test chemical is considered positive for antagonism in case at least two consecutive concentrations of the test chemical show a response that is equal or lower than 80 % of the maximum response of the reference standard tamoxifen (PC₈₀).

PROCEDURE

Cell lines

- The stably transfected U2OS ER α CALUX cell line should be used for the assay. The cell line can be obtained from BioDetection Systems BV, Amsterdam, the Netherlands with a technical licensing agreement.
- Only mycoplasma free cell cultures should be used. Cell batches used should either be certified negative for mycoplasma contamination, or a mycoplasma test should be performed before use. RT-PCR (Real Time Polymerase Chain Reaction) should be used for sensitive detection of mycoplasma infection (9).

Stability of the cell line

- To maintain the stability and integrity of the CALUX cells, the cells should be stored in liquid nitrogen (-80⁰C). Following thawing of cells to start a new culture, cells should be sub-cultured at least twice before being used to assess the estrogenic agonist and antagonist activity of chemicals. Cells should not be sub-cultured for more than 30 passages.
- To monitor the stability of the cell line over time, the responsiveness of the agonistic and antagonistic test system should be verified by evaluating the EC₅₀ or IC₅₀ of the reference standard. In addition, the relative induction of the positive control sample (PC) and the negative control sample (NC) should be monitored. The results should be in agreement with the acceptance criteria for the agonistic (Table 3C) or antagonistic ER α CALUX bioassay (Table 4C). The reference standards, positive and negative controls are given in Table 1 and Table 2 for the agonistic and antagonistic mode respectively.

Cell Culture and plating conditions

- The U2OS cells should be cultured in growth medium (DMEM/F12 (1:1) medium with phenol red as pH indicator, supplemented with fetal bovine serum (7.5 %), non-essential amino acids (1 %), 10 Units/ml of penicillin, streptomycin and geneticin (G-418) as selection marker). Cells should be placed in a CO₂ incubator (5 % CO₂) at 37⁰C and 100 % humidity. When cells reach an 85-95 % confluency, cells should either be subcultured or prepared for seeding in 96-well microtiter plates. In case of the latter, cells should be resuspended at 1x10⁵ cells/ml in estrogen free assay medium (DMEM/F12 (1:1) medium without phenol red, supplemented with Dextran-Coated Charcoal treated fetal bovine serum (5 % v/v), non-essential amino acids (1 % v/v), 10 Units/ml of penicillin and streptomycin) and plated into the wells of the 96-well microtiterplates (100 μ l of homogenised cell suspension). Cells should be pre-incubated in a CO₂ incubator (5 % CO₂, 37⁰C, 100 % humidity) for 24 hours prior to exposure. Plastic ware should be estrogen free.

Acceptability criteria

- Agonistic and antagonistic activities of the test chemical(s) are tested in test series. A test series consists of a maximum of 6 microtiter plates. Each test series contains at least 1 full series of dilutions of a reference standard, a positive control sample, a negative control sample and solvent controls. Figures 1 and 2 give the plate setup for agonistic and antagonistic tests series.

15. Each dilution of the reference standards, test chemicals, all solvent controls, and positive and negative controls should be analysed in triplicate. Each of the triplicate analyses should fulfil the requirements given in Table 3A and Table 4A.
16. A complete series of dilutions of the reference standard (17 β -estradiol for agonism; tamoxifen for antagonism) is measured on the first plate in each test series. To be able to compare the analysis results of the remaining 5 microtiter plates with the first microtiter plate containing the complete concentration-response curve of the reference standard, all plates should contain 3 control samples: solvent control, the highest concentration of the reference standard tested, and the approximate EC₅₀ (agonism) or IC₅₀ (antagonism) concentration of the reference standard. The ratio of the average control samples on the first plate and the remaining 5 plates should fulfil the requirements as given in Table 3C (agonism) or Table 4C (antagonism).
17. For each of the microtiter plates within a test series, the z-factor is calculated (10). The z-factor should be calculated using the responses at the highest and lowest concentration of the reference standard. A microtiter plate is considered valid in case it fulfils the requirements as stated in Table 3C (agonism) or Table 4C (antagonism).
18. The reference standard should demonstrate a sigmoidal dose-response curve. The EC₅₀ or IC₅₀ derived from the response of the series of dilutions of the reference standard, should fulfil the requirements as indicated in Table 3C (agonism) or Table 4C (antagonism).
19. Each test series should contain a positive control and negative control sample. The calculated relative induction of both the positive and negative control sample should fulfil the requirements as indicated in Table 3C (agonism) or Table 4C (antagonism).
20. During all measurements, the induction factor of the highest concentration of the reference standard should be measured by dividing the average highest 17 β -estradiol reference standard relative light unit (RLU) response by the average reference solvent control RLU response. This induction factor should fulfil the minimum requirements for the fold induction as indicated in Table 3C (agonism) or Table 4C (antagonism).
21. Only microtiter plates that fulfil all above mentioned acceptance criteria are considered valid and can be used to evaluate the response of test chemicals.
22. The acceptance criteria are applicable to both prescreen and comprehensive runs.

Table 1

Concentrations of reference standard, positive control (PC) and negative control (NC) for the agonistic CALUX bioassay

	Substance	CAS RN	Test range (M)
Reference standard	17 β -estradiol	50-28-2	1*10 ⁻¹³ – 1*10 ⁻¹⁰
Positive control (PC)	17 α -methyltestosterone	58-18-4	3*10 ⁻⁶
Negative control (NC)	corticosterone	50-22-6	1*10 ⁻⁸

Table 2

Concentrations of reference standard, positive control (PC) and negative control (NC) for the antagonistic CALUX bioassay

	Substance	CAS RN	Test range (M)
Reference standard	tamoxifen	10540-29-1	3×10^{-9} – 1×10^{-5}
Positive control (PC)	4-hydroxytamoxifen	68047-06-3	1×10^{-9}
Negative control (NC)	resveratrol	501-36-0	1×10^{-5}

Table 3

Acceptance criteria for the agonistic ER α CALUX bioassay

A - individual samples on a plate		Criterion
1	Maximum %SD of triplicate wells (for NC, PC, each dilution of the test chemical and the reference standard, except C0)	< 15 %
2	Maximum %SD of triplicate wells (for reference standard and test chemical solvent controls (C0, SC))	< 30 %
3	Maximum LDH leakage, as a measure of cytotoxicity.	< 120 %
B - within a single microtiter plate		
4	Ratio of the reference standard solvent control (C0; plate 1) and test chemical solvent control (SC; plates 2 to x)	0,5 to 2,0
5	Ratio of the appr. EC50 and highest reference standard concentrations on plate 1 and the appr. EC50 and highest reference standard concentrations on plates 2 to x (C4, C8)	0,70 to 1,30
6	Z-factor for each plate	>0.6
C - within a single series of analyses (all plates within one series)		
7	Sigmoidal curve of reference standard	Yes (17 β -estradiol)
8	EC50 range reference standard 17 β -estradiol	4×10^{-12} – 4×10^{-11} M
9	Minimum fold induction of the highest 17 β -estradiol concentration, with respect to the reference standard solvent control.	5
10	Relative induction (%) PC.	> 30 %
11	Relative induction (%) NC	<10 %

Appr.: approximative; PC: positive control; NC: negative control; SC: test chemical solvent control; C0: reference standard solvent control; SD: standard deviation; LDH: lactate dehydrogenase

Table 4

Acceptance criteria for the antagonistic ER α CALUX bioassay

A - individual samples on a plate		Criterion
1	Maximum %SD of triplicate wells (for NC, PC, each dilution of the test chemical and the reference standard, solvent control (C0))	< 15 %
2	Maximum %SD of triplicate wells (for vehicle control (VC) and highest reference standard concentration (C8))	< 30 %
3	Maximum LDH leakage, as a measure of cytotoxicity.	< 120 %
B - within a single microtiter plate		
4	Ratio of the reference standard solvent control (C0; plate 1) and test chemical solvent control (SC; plates 2 to x)	0,70 to 1,30
5	Ratio of the appr. IC50 reference standard concentrations on plate 1 and the appr. IC50 reference standard concentrations on plates 2 to x (C4)	0,70 to 1,30
6	Ratio of the highest reference standard concentrations on plate 1 and the highest reference standard concentrations on plates 2 to x (C8)	0,50 to 2,0
7	Z-factor for each plate	> 0,6
C - within a single series of analyses (all plates within one series)		
8	Sigmoidal curve of reference standard	Yes (Tamoxifen)
9	IC50 range reference standard (Tamoxifen)	$1 \cdot 10^{-8}$ - $1 \cdot 10^{-7}$ M
10	Minimum fold induction of the reference standard solvent control, with respect to the highest Tamoxifen concentration.	2,5
11	Relative induction (%) PC.	< 70 %
12	Relative induction (%) NC	> 85 %

Appr.: approximative; PC: positive control; NC: negative control; VC: vehicle control (solvent control without fixed concentration of agonist reference standard); SC: test chemical solvent control; C0: reference standard solvent control; SD: standard deviation; LDH: lactate dehydrogenase

Solvent/vehicle control, reference standards, positive controls, negative controls

23. For both the prescreen run and comprehensive runs, the same solvent/vehicle control, reference standards, positive controls and negative controls should be used. In addition, the concentration of reference standards, positive controls and negative controls should be the same.

Solvent control

24. The solvent used to dissolve the test chemicals should be tested as a solvent control. Dimethylsulfoxide (DMSO, 1 % (v/v); CASRN 67-68-5) was used as vehicle during the validation of the ER α CALUX bioassay. If a solvent other than DMSO is used, all reference standards, controls, and test chemicals should be tested in the same vehicle. Please note that the solvent control for antagonistic studies contains a fixed concentration of the agonist reference standard 17 β -estradiol (approximately EC₅₀ concentration). To test the solvent used for antagonistic studies, a vehicle control should be prepared and tested.

Vehicle control (antagonism)

25. For testing antagonism, the assay medium is supplemented with a fixed concentration of the agonist reference standard 17 β -estradiol (approximately EC₅₀ concentration). To test the solvent used to dissolve the test chemicals for antagonism, an assay medium without a fixed concentration of the agonist reference standard 17 β -estradiol should be prepared. This control sample is indicated as the vehicle control. Dimethylsulfoxide (DMSO, 1 % (v/v); CASRN 67-68-5) was used as vehicle during the validation of the ER α CALUX bioassay. If a solvent other than DMSO is used, all reference standards, controls, and test chemicals should be tested in the same vehicle.

Reference standards

26. The agonistic reference standard is 17 β -estradiol (Table 1). The reference standards comprise a series of dilutions of eight concentrations of 17 β -estradiol (1×10^{-13} , 3×10^{-13} , 1×10^{-12} , 3×10^{-12} , 6×10^{-12} , 1×10^{-11} , 3×10^{-11} , 1×10^{-10} M).
27. The antagonistic reference standard is tamoxifen (Table 2). The reference standards comprise a series of dilutions of eight concentrations of tamoxifen (3×10^{-9} , 1×10^{-8} , 3×10^{-8} , 1×10^{-7} , 3×10^{-7} , 1×10^{-6} , 3×10^{-6} , 1×10^{-5} M). Each of the concentrations of the antagonistic reference standard is co-incubated with a fixed concentration of the agonistic reference standard 17 β -estradiol (3×10^{-12} M).

Positive control

28. The positive control for agonistic studies is 17 α -methyltestosterone (Table 1).
29. The positive control for antagonistic studies is 4-hydroxytamoxifen (Table 2). The antagonistic positive control is co-incubated with a fixed concentration of the agonistic reference standard 17 β -estradiol (3×10^{-12} M).

Negative control

30. The negative control for agonistic studies is corticosterone (Table 1).
31. The negative control for antagonistic studies is resveratrol (Table 2). The antagonistic negative control is co-incubated with a fixed concentration of the agonistic reference standard 17 β -estradiol (3×10^{-12} M).

Demonstration of laboratory proficiency (see paragraph 14 and Tables 3 and 4 in «ER TA ASSAY COMPONENTS» of this test method)

Vehicle

32. The solvent used to dissolve test chemicals should solubilise the test chemical completely and should be miscible with the cell medium. DMSO, water and ethanol (95 % to 100 % purity) are suitable solvents. In case DMSO is used as solvent, the maximum concentration of DMSO during incubation should not exceed 1 % (v/v). Prior to use, the solvent should be tested for absence of cytotoxicity and interference with the assays performance.

Preparation of reference standards, positive controls, negative controls and test chemicals

33. Reference standards, positive controls, negative controls and test chemicals are dissolved in 100 % DMSO (or an appropriate solvent). Appropriate (serial) dilutions should then be prepared in the same solvent. Before being dissolved, all substances should be allowed to equilibrate to room temperature. Freshly prepared stock solutions of reference standards, positive controls, negative controls and test chemicals should not have noticeable precipitate or cloudiness. Reference standard and control stocks may be prepared in bulk. Stock solutions of test chemicals should be prepared fresh before each experiment. Final dilutions of reference standards, positive controls, negative controls and test chemicals should be prepared for each experiment fresh and used within 24 hours of preparation.

Solubility, cytotoxicity and range finding.

34. During the prescreen run, the solubility of the test chemicals in the solvent of choice is determined. A maximum stock concentration of 0,1 M is prepared. In case this concentration shows solubility problems, lower stock solutions should be prepared until test chemicals are fully solubilised. During the prescreen run, 1:10 serial dilutions of test chemical are tested. The maximum assay concentration for agonist or antagonist testing is 1 mM. Following prescreening, an appropriate refined concentration range for test chemicals is derived that should be tested during the comprehensive runs. The dilutions used for comprehensive testing should be 1x, 3x, 10x, 30x, 100x, 300x, 1000x and 3000x.
35. Cytotoxicity testing is included in the agonist and antagonist assay protocol (11). Cytotoxicity testing is incorporated in both the prescreen run and comprehensive runs. The method used to assess cytotoxicity during the validation of the ER α CALUX bioassay was the lactate dehydrogenase (LDH) leakage test in combination with qualitative visual inspection of cells (see Appendix 4.1) following exposure to test chemicals. However, other quantitative methods for the determination of cytotoxicity (e.g. tetrazolium-based colorimetric (MTT) assay or cytotoxicity CALUX bioassay) can be used. In general, test chemical concentrations that show more than 20 % reduction of cell viability are considered cytotoxic and therefore cannot be used for data evaluation. With respect to the LDH leakage assay, the concentration of the test chemical is regarded cytotoxic when the percentage LDH leakage is higher than 120 %.

Test chemical exposure and assay plate organisation

36. Following trypsination of a confluent flask of cultured cells, cells are re-suspended at 1×10^5 cells/ml in estrogen free assay medium. Hundred μ l of re-suspended cells are plated in the inner-wells of a 96-well microtiter plate. The outer wells are filled with 200 μ l of Phosphate Buffered Saline (PBS) (see Figures 1 and 2). The plated cells are pre-incubated for 24 hours in a CO₂ incubator (5 % CO₂, 37°C, 100 % humidity).
37. After pre-incubation, the plates are inspected for visual cytotoxicity (see Appendix 4.1), contamination and confluence. Only plates that show no visual cytotoxicity, contamination and have a minimum of 85 % confluence are used for testing. The medium from the inner wells is carefully removed and replaced by 200 μ l of estrogen free assay medium containing appropriate dilutions series of reference standards, test chemicals, positivecontrols,

negative controls and solvent controls (Table 5: agonist studies; Table 6: antagonist studies). All reference standards, test chemicals, positive controls, negative controls and solvent controls are tested in triplicate. In Figure 1, the plate layout for agonist testing is given. In Figure 2, the plate layout for antagonist testing is given. The plate layout for prescreen testing and comprehensive testing is identical. For antagonist testing, all inner-wells, except for the vehicle control wells (VC), also contain a fixed concentration of agonist reference standard 17β -estradiol (3×10^{-12} M). Note that reference standards C8 and C4 should be added to each TC plate.

38. Following exposure of the cells to all chemicals, the 96-well microtiter plates should be incubated for another 24 hours in a CO₂ incubator (5 % CO₂, 37°C, 100 % humidity).

Figure 1

Plate layout of the 96-well microtiter plates for prescreening and assessment of agonistic effect.

Plate 1

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		C0	C1	C2	C3	C4	C5	C6	C7	C8	PC	
C		C0	C1	C2	C3	C4	C5	C6	C7	C8	PC	
D		C0	C1	C2	C3	C4	C5	C6	C7	C8	PC	
E		SC	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	NC	
F		SC	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	NC	
G		SC	TC1-1	TC1-2	TC1-3	TC1-4	TC1-5	TC1-6	TC1-7	TC1-8	NC	
H												

Subsequent plates

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		SC	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	C8 (max)	
C		SC	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	C8 (max)	
D		SC	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	C8 (max)	
E		SC	TCx-1	TCx-2	TCx-3	TCx-4	TCx-5	TCx-6	TCx-7	TCx-8	C4 (EC50)	
F		SC	TCx-1	TCx-2	TCx-3	TCx-4	TCx-5	TCx-6	TCx-7	TCx-8	C4 (EC50)	
G		SC	TCx-1	TCx-2	TCx-3	TCx-4	TCx-5	TCx-6	TCx-7	TCx-8	C4 (EC50)	
H												

C0 = reference standard solvent.

C(1-8) = series of dilutions (1-8, low-to-high concentrations) of reference standard.

PC = positive control.

NC = negative control.

TCx-(1-8) = dilutions (1-8, low-to-high concentrations) of test chemical for the prescreen run and assessment of agonistic effect of test chemical x.

SC = solvent control of the test chemical (optimally the same solvent as in C0, but possibly from another batch).

Grey cells: = Outer wells, filled up with 200 µl of PBS.

Figure 2

Plate layout of the 96-well microtiter plates for antagonistic prescreening and assessment of antagonistic effect.**Plate 1**

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		C0	C1	C2	C3	C4	C5	C6	C7	C8	VC	
C		C0	C1	C2	C3	C4	C5	C6	C7	C8	VC	
D		C0	C1	C2	C3	C4	C5	C6	C7	C8	VC	
E		NC	TC1-	TC1-2	TC1-3	TC	TC1-5	TC1-	TC1-	TC1-	PC	
F		NC	TC1-	TC1-2	TC1-3	TC	TC1-5	TC1-	TC1-	TC1-	PC	
G		NC	TC1-	TC1-2	TC1-3	TC	TC1-5	TC1-	TC1-	TC1-	PC	
H												

Subsequent plates

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		SC	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	C8 (max)	
C		SC	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	C8 (max)	
D		SC	TC2-1	TC2-2	TC2-3	TC2-4	TC2-5	TC2-6	TC2-7	TC2-8	C8 (max)	
E		C4 (IC50)	TCx-1	TCx-2	TCx-3	TCx-4	TCx-5	TCx-6	TCx-7	TCx-8	C8 (max)	
F		C4 (IC50)	TCx-1	TCx-2	TCx-3	TCx-4	TCx-5	TCx-6	TCx-7	TCx-8	C8 (max)	
G		C4 (IC50)	TCx-1	TCx-2	TCx-3	TCx-4	TCx-5	TCx-6	TCx-7	TCx-8	C8 (max)	
H												

C0 = reference standard solvent.

C(1-8) = series of dilutions (1-8, low-to-high concentrations) of reference standard.

NC = negative control.

PC = positive control.

TCx-(1-8) = dilutions (1-8, low-to-high concentrations) of test chemical for the prescreen run and assessment of agonistic effect of test chemical x.

SC = solvent control of the test chemical (optimally the same solvent as in C0, but possibly from another batch).

VC = vehicle control (solvent control without fixed concentration of agonist reference standard 17 β -estradiol).

Grey cells: = Outer wells, filled up with 200 μ l of PBS.

Note: all inner-wells, except for the vehicle control wells (VC), also contain a fixed concentration of agonist reference standard 17 β -estradiol (3.0*10⁻¹² M)

Measurement of luminescence

39. The measurement of luminescence is described in detail in the agonist and antagonist assay protocol (10). The medium from the wells should be removed and the cells should be lysed following 24 hours of incubation in order to open up the cell membrane and allow measurement of luciferase activity.
40. For measuring the luminescence, this procedure requires a luminometer equipped with 2 injectors. The luciferase reaction is started by injection of the substrate luciferin. The reaction is stopped by addition of 0,2 M NaOH. The reaction is stopped to prevent carry over of luminescence from one well to the other.
41. Light emitted from each well is expressed as Relative Light Units (RLUs) per well.

Prescreen run

42. The prescreen analysis results are used to determine a refined concentration-range of test chemicals for comprehensive testing. Evaluation of prescreen analysis results and the determination of the refined concentration-range of test chemicals for comprehensive testing, is described in depth in the agonist and antagonist assay protocol (10). Here, a brief summary of the procedures for determining the concentration range of test chemicals for agonist and antagonist testing, is given. See Tables 5 and 6 for guidance of serial dilution design.

Selection of concentrations for assessment of agonistic effects

43. During the prescreen run, test chemicals should be tested using the series of dilutions as indicated in Tables 5 (agonism) and 6 (antagonism). All concentrations should be tested in triplicate wells according to the plate layout as indicated in Figure 1 (agonism) or 2 (antagonism).
44. Only analysis results that fulfil the acceptance criteria (Table 3) are considered valid and can be used to evaluate the response of test chemicals. In case one or more microtiter plates in an analysis series fail to fulfil the acceptance criteria, the respective microtiterplates should be re-analysed. In case the first plate containing the complete series of dilutions of the reference standard fails the acceptance criteria, the complete test series (6 plates) have to be re-analysed.
45. Initial concentration ranges of test chemicals should be adjusted and the prescreen run should be repeated in case:
 - cytotoxicity is observed. The prescreen procedure should be repeated with lower non-cytotoxic concentrations of the test chemical.
 - the prescreen of the test chemical does not show a full dose-response curve because the concentrations tested generate maximum induction. The prescreen run should be repeated using lower concentrations of the test chemical.
46. When a valid dose-related response is observed, the (lowest) concentration at which maximum induction is observed and does not show cytotoxicity, should be selected. The highest concentration of the test chemical to be tested in the comprehensive runs, should be 3-times this selected concentration.
47. A complete refined dilution series of the test chemical should be prepared with dilutions steps as indicated in Table 5, starting with the highest concentration as determined above.
48. A test chemical that does not elicit any agonistic effect, should be tested in the comprehensive runs starting with the highest, non-cytotoxic concentration identified during prescreening.

Selection of concentrations for assessment of antagonistic effects

49. Only analysis results that fulfil the acceptance criteria (Table 4) are considered valid and can be used to evaluate the response of test chemicals. In case one or more microtiter plates in an analysis series fail to fulfil the acceptance criteria, the respective microtiterplates should be re-analysed. In case the first plate containing the complete series of dilutions of the reference standard fails the acceptance criteria, the complete test series (6 plates) have to be re-analysed.
50. Initial concentration ranges of test chemicals should be adjusted and the prescreen run should be repeated in case:
- cytotoxicity is observed. The prescreen procedure should be repeated with lower non-cytotoxic concentrations of the test chemical.
 - the prescreen of the test chemical does not show a full dose-response curve because the concentrations tested generate maximum inhibition. The prescreen should be repeated using lower concentrations of the test chemical.
51. When a valid dose-related response is found, the (lowest) concentration at which maximum inhibition is observed and does not show cytotoxicity, should be selected. The highest concentration of the test chemical to be tested in the comprehensive runs, should be 3-times this selected concentration.
52. A complete refined dilution series of the test chemical should be prepared with the dilutions steps as indicated in Table 6, starting with the highest concentration as determined above.
53. Test chemicals that do not elicit any antagonistic effects, should be tested in the comprehensive runs starting with the highest, non-cytotoxic concentration tested during prescreening.

Comprehensive runs

54. Following the selection of the refined concentration ranges, test chemicals should be tested comprehensively using the series of dilutions as indicated in Tables 5 (agonism) and 6 (antagonism). All concentrations should be tested in triplicate wells according to the plate layout as indicated in Figure 1 (agonism) or 2 (antagonism).
55. Only analysis results that fulfil the acceptance criteria (Table 3 and 4) are considered valid and can be used to evaluate the response of test chemicals. In case one or more microtiter plates in an analysis series fail to fulfil the acceptance criteria, the respective microtiterplates should be re-analysed. In case the first plate containing the complete series of dilutions of the reference standard fails the acceptance criteria, the complete test series (6 plates) have to be re-analysed.

Table 5

Concentration and dilutions of reference standards, controls and test chemicals used for agonist testing

Reference 17 β -estradiol		TCx - prescreen run		TCx - comprehensive run		Controls	
conc. (M)		dilution		dilution		conc. (M)	
C0	0	TCx-1	10 000 000 x	TCx-1	3 000 x	PC	3*10 ⁻⁶
C1	1*10 ⁻¹³	TCx-2	1 000 000 x	TCx-2	1 000 x	NC	1*10 ⁻⁸
C2	3*10 ⁻¹³	TCx-3	100 000 x	TCx-3	300 x	C0	0
C3	1*10 ⁻¹²	TCx-4	10 000 x	TCx-4	100 x	SC	0
C4	3*10 ⁻¹²	TCx-5	1 000 x	TCx-5	30 x		
C5	6*10 ⁻¹²	TCx-6	100 x	TCx-6	10 x		
C6	1*10 ⁻¹¹	TCx-7	10 x	TCx-7	3 x		
C7	3*10 ⁻¹¹	TCx-8	1 x	TCx-8	1 x		
C8	1*10 ⁻¹⁰						

TCx - test chemical x
PC - positive control (17 α -methyltestosterone)
NC - negative control (corticosterone)
C0 - reference standard solvent control
SC - test chemical solvent control

Table 6

Concentration and dilutions of reference standards, controls and test chemicals used for antagonist testing

Reference tamoxifen		TCx - prescreen run		TCx - comprehensive run		Controls	
conc. (M)		dilution		dilution		conc. (M)	
C0	0	TCx-1	10 000 000 x	TCx-1	3 000 x	PC	$1 \cdot 10^{-9}$
C1	$3 \cdot 10^{-9}$	TCx-2	1 000 000 x	TCx-2	1 000 x	NC	$1 \cdot 10^{-5}$
C2	$1 \cdot 10^{-8}$	TCx-3	100 000 x	TCx-3	300 x	C0	0
C3	$3 \cdot 10^{-8}$	TCx-4	10 000 x	TCx-4	100 x	SC	0
C4	$1 \cdot 10^{-7}$	TCx-5	1 000 x	TCx-5	30 x		
C5	$3 \cdot 10^{-6}$	TCx-6	100 x	TCx-6	10 x	Supplemented agonist	
C6	$1 \cdot 10^{-6}$	TCx-7	10 x	TCx-7	3 x	conc. (M)	
C7	$3 \cdot 10^{-6}$	TCx-8	1 x	TCx-8	1 x	17 β -estradiol	$3 \cdot 10^{-12}$
C8	$1 \cdot 10^{-5}$						

TCx - test chemical x

PC - positive control (4-hydroxytamoxifen)

NC - negative control (resveratrol)

C0 - reference standard solvent control

SC - test chemical solvent control

VC - vehicle control (does not contain fixed concentration of the agonistic reference standard 17 β -estradiol ($3.0 \cdot 10^{-12}$ M))

Collection of data and data analysis

56. Following the prescreen and comprehensive runs, the EC₁₀, EC₅₀, PC₁₀, PC₅₀ and maximum induction (TCx_{max}) of a test chemical should be determined for agonistic testing. For antagonistic testing, the IC₂₀, IC₅₀, PC₈₀, PC₅₀ and minimum induction (TCx_{min}) should be calculated. In Figure 3 (agonism) and 4 (antagonism), a graphical representation of these parameters is given. The required parameters are calculated based on the relative induction of each test chemical (relative to the maximum induction of the reference standard (=100 %)). Non-linear regression (variable slope, 4 parameters) should be used for evaluation of data according to the following equation:

$$Y = Bottom + \frac{(Top - Bottom)}{(1 + 10^{(\lg EC_{50} - X) \times Hill\ Slope})}$$

Where:

X = Log of dose or concentration

Y = Response (relative induction (%))

Top = Maximum induction (%)

Bottom = Minimum induction (%)

LogEC₅₀ = Log of concentration at which 50 % of maximum response is observed

HillSlope = Slope factor or Hill slope

57. Raw data from the luminometer, expressed as Relative Light Units (RLUs), should be transferred to the data analysis spreadsheet designed for the prescreen and comprehensive runs. Raw data should meet the acceptance criteria as indicated in Table 3A and 3B (agonism) or 4A and 4B (antagonism). In case the raw data meet the acceptance criteria, the following calculation steps are performed to determine the required parameters:

Agonism

- Subtract the average RLU of the reference standard solvent control from each of the raw analysis data of the reference standards.

- Subtract the average RLU for the test chemical solvent control from each of the raw analysis data of the test chemicals.

- Calculate the relative induction of each concentration of the reference standard. Set the induction of the highest concentration of the reference standard at 100 %.

- Calculate the relative induction of each concentration of test chemical compared to the highest concentration of the reference standard as 100 %.

- Evaluate the analysis results following non-linear regression (variable slope, 4 parameters).

- Determine the EC_{50} and EC_{10} of the reference standard.

- Determine the EC_{50} and EC_{10} of the test chemicals.

- Determine the maximum relative induction of the test chemical (TC_{max}).

- Determine the PC_{10} and PC_{50} of the test chemicals.

For test chemicals, a full dose-response curve may not always be achieved due to e.g. cytotoxicity or solubility problems. Hence, the EC_{50} , EC_{10} and PC_{50} cannot be determined. In such case, only the PC_{10} and TC_{max} can be determined.

Antagonism

- Subtract the average RLU of the highest reference standard concentration from each of the raw analysis data of the reference standards.

- Subtract the average RLU of the highest reference standard concentration from each of the raw analysis data of the test chemicals.

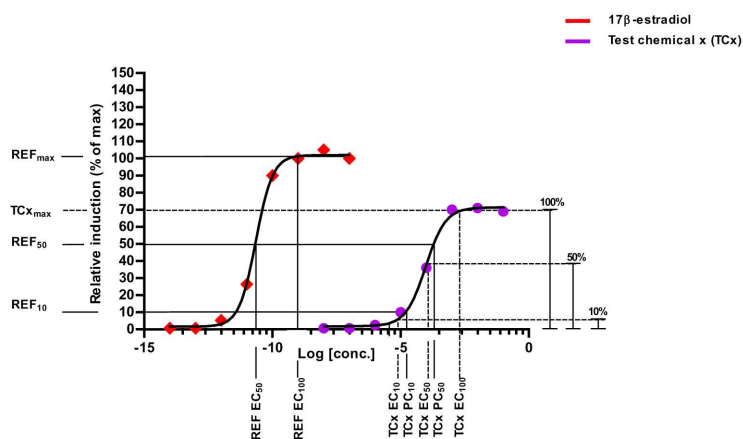
- Calculate the relative induction of each concentration of the reference standard. Set the induction of the lowest concentration of the reference standard at 100 %.

- Calculate the relative induction of each concentration of test chemical compared to the lowest concentration of the reference standard as 100 %.

- Evaluate the analysis results following non-linear regression (variable slope, 4 parameters).
- Determine the IC_{50} and IC_{20} of the reference standard.
- Determine the IC_{50} and IC_{20} of the test chemicals.
- Determine the minimum relative induction of the test chemical (TC_{min}).
- Determine the PC_{80} and PC_{50} of the test chemicals.

Figure 3

Overview of parameters determined in the agonist assay



EC_{10} = concentration of a substance at which 10 % of its maximum response is observed.

EC_{50} = concentration of a substance at which 50 % of its maximum response is observed.

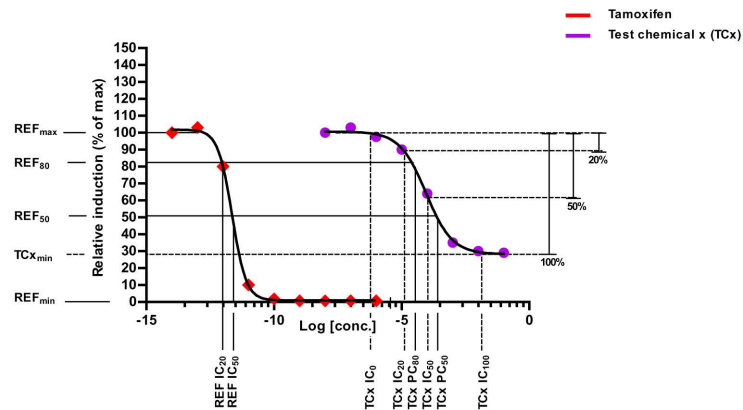
PC_{10} = concentration of a test chemical at which its response is equal to the EC_{10} of the reference standard.

PC_{50} = concentration of a test chemical at which its response is equal to the EC_{50} of the reference standard.

TCx_{max} = maximum relative induction of test chemical.

Figure 4

Overview of parameters determined in the antagonist assay



IC₂₀ = concentration of a substance at which 80 % of its maximum response is observed (20 % inhibition).

IC₅₀ = concentration of a substance at which 50 % of its maximum response is observed (50 % inhibition).

PC₈₀ = concentration of a test chemical at which its response is equal to the IC₂₀ of the reference standard.

PC₅₀ = concentration of a test chemical at which its response is equal to the IC₅₀ of the reference standard.

TCx_{min} = minimum relative induction of test chemical.

For test chemicals, a full dose-response curve may not always be achieved due to e.g. cytotoxicity or solubility problems. Hence, the IC₅₀, IC₂₀ and PC₅₀ cannot be determined. In such case, only the PC₂₀ and TC_{min} can be determined.

58. The results should be based on two (or three) independent runs. If two runs give comparable and therefore reproducible results, it is not necessary to conduct a third run. To be acceptable, the results should:

— Meet the acceptability criteria (see Acceptability criteria paragraphs 14-22),

— Be reproducible.

Data interpretation criteria

59. For the interpretation of data and the decision whether a test chemical is considered positive or negative, the following criteria are to be used:

Agonism

For each comprehensive run, a test chemical is considered **positive** in case:

- 1 The TC_{max} is equal or exceeds 10 % of the maximum response of the reference standard (REF_{10}).
- 2 At least 2 consecutive concentrations of the test chemical are equal to or exceed the REF_{10} .

For each comprehensive run, a test chemical is considered **negative** in case:

- 1 The TC_{max} does not exceed 10 % of the maximum response of the reference standard (REF_{10}).
- 2 Less than 2 concentrations of the test chemical are equal to or exceed the REF_{10} .

Antagonism

For each comprehensive run, a test chemical is considered **positive** in case:

- 1 The TC_{min} is equal or lower than 80 % of the maximum response of the reference standard ($REF_{80} = 20\%$ inhibition).
- 2 At least 2 consecutive concentrations of the test chemical are equal to or lower than the REF_{80} .

For each comprehensive run, a test chemical is considered **negative** in case:

- 1 The TC_{min} exceeds 80 % of the maximum response of the reference standard ($REF_{80} = 20\%$ inhibition).
- 2 Less than 2 concentrations of the test chemical are equal to or lower than the REF_{80} .

60. To characterise the potency of the positive response of a test chemical, the magnitude of the effect (agonism: TC_{max} ; antagonism: TC_{min}) and the concentration at which the effect occurs (agonism: EC_{10} , EC_{50} , PC_{10} , PC_{50} ; antagonism: IC_{20} , IC_{50} , PC_{80} , PC_{50}) should be reported.

TEST REPORT

61. See paragraph 20 of 'ER TA ASSAY COMPONENTS'

LITERATURE

- (1) OECD (2016). Draft Validation report of the (anti-) ER α CALUX bioassay - transactivation bioassay for the detection of compounds with (anti)estrogenic potential. Environmental Health and Safety Publications, Series on Testing and Assessment (No 240). Organisation for Economic Cooperation and Development, Paris
- (2) Sonneveld E, Jansen HJ, Riteco JA, Brouwer A, van der Burg B. (2005). Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid-responsive bioassays. *Toxicol Sci.* 83(1), 136-148.
- (3) Quaedackers ME, van den Brink CE, Wissink S, Schreurs RHMM, Gustafsson JA, van der Saag PT, and van der Burg B. (2001). 4-Hydroxytamoxifen trans-represses nuclear factor-kB Activity in human osteoblastic U2OS cells through estrogen receptor (ER) α and not through ER β . *Endocrinology* 142(3), 1156-1166.
- (4) Thorne N, Inglese J and Auld DS. (2010). Illuminating Insights into Firefly Luciferase and Other Bioluminescent Reporters Used in Chemical Biology, *Chemistry and Biology*17(6):646-57.
- (5) Escande A, Pillon A, Servant N, Cravedi JP, Larrea F, Muhn P, Nicolas JC, Cavallès V and Balaguer P. (2006). Evaluation of ligand selectivity using reporter cell lines stably expressing estrogen receptor alpha or beta. *Biochem. Pharmacol.*, 71, 1459-1469.
- (6) Kuiper GG, Lemmen JG, Carlsson B, Corton JC, Safe SH, van der Saag PT, van der Burg B and Gustafsson JA. (1998). Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. *Endocrinol.*, 139, 4252-4263.
- (7) Sotoca AM, Bovee TFH, Brand W, Velikova N, Boeren S, Murk AJ, Vervoort J, Rietjens IMCM. (2010). Super-induction of estrogen receptor mediated gene expression in luciferase based reporter gene assays is mediated by a post-transcriptional mechanism. *J. Steroid. Biochem. Mol. Biol.*, 122, 204–211.

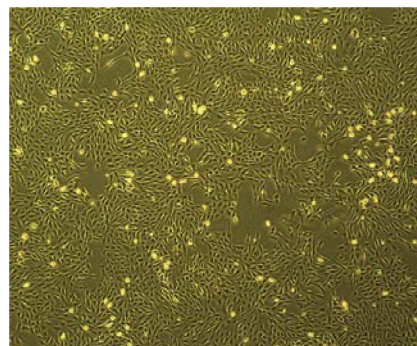
- (8) Sonneveld E, Riteco JAC, Jansen HJ, Pieterse B, Brouwer A, Schoonen WG, and van der Burg B. (2006). Comparison of *in vitro* and *in vivo* screening models for androgenic and estrogenic activities. *Toxicol. Sci.*, 89(1), 173–187.
- (9) Kobayashi H, Yamamoto K, Eguchi M, Kubo M, Nakagami S, Wakisaka S, Kaizuka M and Ishii H. (1995). Rapid detection of mycoplasma contamination in cell cultures by enzymatic detection of polymerase chain reaction (PCR) products. *J. Vet. Med. Sci.*, 57(4), 769-771.
- (10) Zhang J-H, Chung TDY, and Oldenburg KR. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Scr.*, 4, 67-73
- (11) Besselink H, Middelhof I, and Felzel, E. (2014). Transactivation assay for the detection of compounds with (anti)estrogenic potential using ER α CALUX cells. BioDetection Systems BV (BDS). Amsterdam, the Netherlands.

Appendix 4.1

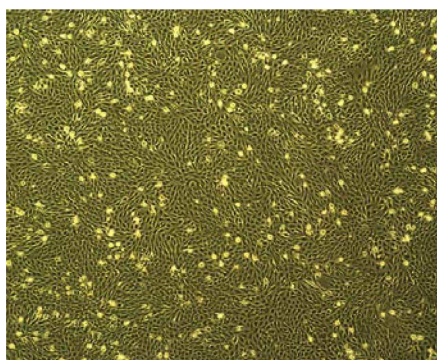
VISUAL INSPECTION OF CELL VIABILITY



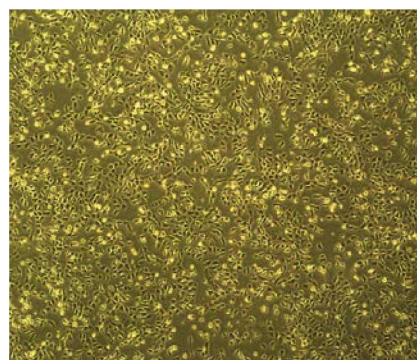
< 5 % confluency. Cells have just been seeded. 100 % cell viability. Classification: “no cytotoxicity”



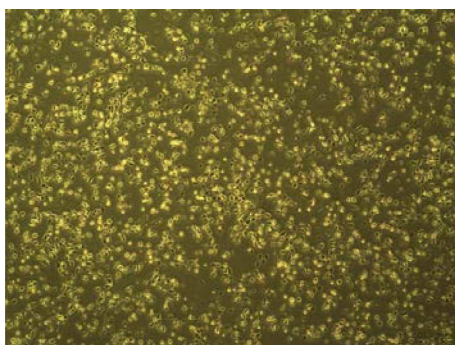
> 85 % confluency. At this stage, cells are exposed to test chemicals. > 95 % cell viability. Classification: “no cytotoxicity”



> 95 % confluency. Cells are densely packed and start to overgrow. > 95 % cell viability. Classification: “no cytotoxicity”



< 25 % cell viability. Cells become detached and contact between cells decreases. Cells are rounded. Classification: “cytotoxicity”



< 5 % cell viability. Cells are fully detached and contact between cells is broken. Cells are rounded. Classification: “cytotoxicity”

B.67 IN VITRO MAMMALIAN CELL GENE MUTATION TESTS USING THE THYMIDINE KINASE GENE

INTRODUCTION

1. This test method (TM) is equivalent to the OECD test guideline 490 (2016). Test methods are periodically reviewed and revised in the light of scientific progress, regulatory needs and animal welfare. The mouse lymphoma assay (MLA) and TK6 test using the thymidine kinase (TK) locus were originally contained in test method B.17. Subsequently, the MLA Expert Workgroup of the International Workshop for Genotoxicity Testing (IWGT) has developed internationally harmonised recommendations for assay acceptance criteria and data interpretation for the MLA (1)(2)(3)(4)(5), and these recommendations are incorporated into this new test method B.67. This test method is written for the MLA and, because it also utilises the TK locus, the TK6 test. While the MLA has been widely used for regulatory purposes, the TK6 has been used much less frequently. It should be noted that in spite of the similarity between the endpoints the two cell lines are not interchangeable and regulatory programs may validly express a preference for one over the other for a particular regulatory use. For instance, the validation of the MLA demonstrated its appropriateness for detecting not only gene mutation, but also, the ability of a test chemical to induce structural chromosomal damage. This test method is part of a series of test methods on genetic toxicology. A document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to genetic toxicity OECD test guidelines has been developed by OECD (6).
2. The purpose of the *in vitro* mammalian cell gene mutation tests is to detect gene mutations induced by chemicals. The cell lines used in these tests measure forward mutations in reporter genes, specifically the endogenous thymidine kinase gene (*TK* for human cells and *Tk* for rodent cells, collectively referred to as *TK* in this test method). This test method is intended for use with two cell lines: the L5178Y $TK^{+/-}$ -3.7.2C mouse lymphoma cell line (generally called L5178Y) and the TK6 human lymphoblastoid cell line (generally called TK6). Although the two cell lines vary because of their origin, cell growth, p53-status, etc., the *TK* gene mutation tests can be conducted in a similar way in both cell types as described in this test method.
3. The autosomal and heterozygous nature of the thymidine kinase gene enables the detection of viable colonies whose cells are deficient in the enzyme thymidine kinase following mutation from $TK^{+/-}$ to $TK^{-/-}$. This deficiency can result from genetic events affecting the *TK* gene including both gene mutations (point mutations, frame-shift mutations, small deletions, etc.) and chromosomal events (large deletions, chromosome rearrangements and mitotic recombination). The latter events are expressed as loss of heterozygosity, which is a common genetic change of tumor suppressor genes in human tumorigenesis. Theoretically, loss of the entire chromosome carrying the *TK* gene resulting from spindle impairment and/or mitotic non-disjunction can be detected in the MLA. Indeed, a combination of cytogenetic and molecular analysis clearly shows that some MLA *TK* mutants are the result of nondisjunction. However, the weight of evidence shows that the *TK* gene mutation tests cannot reliably detect aneuploids when applying standard cytotoxicity criteria (as described in this test method) and therefore, it is not appropriate to use these tests to detect aneuploids (7)(8)(9).
4. In the *TK* gene mutation tests, two distinct phenotypic classes of *TK* mutants are generated; the normal growing mutants that grow at the same rate as the *TK* heterozygous cells, and slow growing mutants which grow with prolonged doubling times. The normal growing and slow growing mutants are recognised as large colony and small colony mutants in the MLA and as early appearing colony and late appearing colony mutants in the TK6. The molecular and cytogenetic nature of both large and small colony MLA mutants has been explored in detail (8)(10)(11)(12)(13). The molecular and cytogenetic nature of the early appearing and late appearing TK6 mutants has also been extensively investigated (14)(15)(16)(17). Slow growing mutants for both cell types have suffered genetic damage that involves putative growth regulating gene(s) near the *TK* locus which results in prolonged doubling times and the formation of late appearing or small colonies (18). The induction of slow growing mutants has been associated with chemicals that induce gross structural changes at the chromosomal level. Cells whose damage does not involve the putative growth regulating gene(s) near the *TK* locus grow at rates similar to the parental cells and become normal growing mutants. The induction of primarily normal growing mutants is associated with chemicals primarily acting as point mutagens. Consequently it is essential to count both slow growing and normal growing mutants in order to recover all of the mutants and to provide some insight into the type(s) of damage (mutagens vs. clastogens) induced by the test chemical (10)(12)(18)(19).

5. The test method is organised so as to provide general information that applies to both MLA and TK6 and specialised guidance for the individual tests.
6. Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

7. Tests conducted *in vitro* generally require the use of an exogenous source of metabolic activation. The exogenous metabolic activation system does not entirely mimic *in vivo* conditions.
8. Care should be taken to avoid conditions that could lead to artefactual positive results (i.e. possible interaction with the test system) not caused by interaction between the test chemical and the genetic material of the cell; such conditions include changes in pH or osmolality, interaction with the medium components (20)(21), or excessive levels of cytotoxicity (22)(23)(24). Cytotoxicity exceeding the recommended top cytotoxicity levels as defined in paragraph 28 is considered excessive for the MLA and TK6. In addition, it should be noted that test chemicals that are thymidine analogues, or behave like thymidine analogues can increase the mutant frequency by selective growth of the spontaneous background mutants during cell treatment and require additional test methods for adequate evaluation (25).
9. For manufactured nanomaterials, specific adaptations of this test method may be needed but are not described in this test method.
10. Before using the test method for testing a mixture to generate data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing the mixture.
11. Mutant cells deficient in thymidine kinase enzyme activity because of a mutation $TK^{+/-}$ to $TK^{-/-}$ are resistant to the cytostatic effects of the pyrimidine analogue trifluorothymidine (TFT). The TK proficient cells are sensitive to TFT, which causes the inhibition of cellular metabolism and halts further cell division. Thus, mutant cells are able to proliferate in the presence of TFT and form visible colonies, whereas cells containing the TK enzyme are not.

PRINCIPLE OF THE TEST

12. Cells in suspension are exposed to the test chemical, both with and without an exogenous source of metabolic activation (see paragraph 19), for a suitable period of time (see paragraph 33), and then sub-cultured to determine cytotoxicity and to allow phenotypic expression prior to mutant selection. Cytotoxicity is determined by relative total growth (RTG—see paragraph 25) for the MLA and by relative survival (RS—see paragraph 26) for TK6. The treated cultures are maintained in growth medium for a sufficient period of time, characteristic of each cell type (see paragraph 37), to allow near-optimal phenotypic expression of induced mutations. Following phenotypic expression, mutant frequency is determined by seeding known numbers of cells in medium containing the selective agent to detect mutant colonies, and in medium without selective agent to determine the cloning efficiency (viability). After a suitable incubation time, colonies are counted. Mutant frequency is calculated based on the number of mutant colonies corrected by the cloning efficiency at the time of mutant selection.

DESCRIPTION OF THE METHOD

Preparations

Cells

13. For MLA: Because the MLA was developed and characterised using the $TK^{+/-}$ -3.7.2C subline of L5178Y cells, this specific subline has to be used for the MLA. The L5178Y cell line was derived from a methylcholanthrene-induced thymic lymphoma from a DBA-2 mouse (26). Clive and co-workers treated L5178Y cells (designated by Clive as $TK^{+/+}$ -3) with ethylmethane sulfonate and isolated a $TK^{-/-}$ (designated as $TK^{-/-}$ -3.7) clone using bromodeoxyuridine as the selective agent. From the $TK^{-/-}$ clone a spontaneous $TK^{+/-}$ clone (designated as $TK^{+/-}$ -3.7.2.) and a subclone

(designated as TK^{+/-}-3.7.2C) were isolated and characterised for use in the MLA (27). The karyotype for the cell line has been published (28)(29)(30)(31). The modal chromosome number is 40. There is one metacentric chromosome (t12;13) that should be counted as one chromosome. The mouse TK locus is located on the distal end of chromosome 11. The L5178Y TK^{+/-}-3.7.2C cell line has mutations in both p53 alleles and produces mutant-p53 protein (32) (33). The p53 status of the TK^{+/-}-3.7.2C cell line is likely responsible for the ability of the test to detect large-scale damage (17).

14. For TK6: The TK6 is a human lymphoblastoid cell line. The parent cell line is an Epstein-Barr virus-transformed cell line, WI-L2, which was originally derived from a 5-year-old male with hereditary spherocytosis. The first isolated clone, HH4, was mutagenised with ICR191 and a TK heterozygous cell line, TK6, was generated (34). TK6 cells are nearly diploid and the representative karyotype is 47, XY, 13+, t(14; 20), t(3; 21) (35). The human TK locus is located on the long arm of chromosome 17. The TK6 is a p53-competent cell line, because it has a wild-type p53 sequence in both alleles and expresses only wild-type p53 protein (36).
15. For both the MLA and the TK6, when first establishing or replenishing a master stock, it is advisable for the testing laboratory to assure the absence of *Mycoplasma* contamination, karyotype the cells or paint the chromosomes harboring the TK locus, and to check population doubling times. The normal cell cycle time for the cells used in the testing laboratory should be established and should be consistent with published cell characteristics (16)(19)(37). This master stock should be stored at -150° C or below and used to prepare all working cell stocks.
16. Either prior to establishing a large number of cryopreserved working stocks or just prior to use in an experiment, the culture may need to be cleansed of pre-existing mutant cells [unless the solvent control mutant frequency (MF) is already within the acceptable range—see Table 2 for the MLA)]. This is accomplished using methotrexate (aminopterin) to select against TK-deficient cells and adding thymidine, hypoxanthine and glycine (L5178Y) or 2'-deoxycytidine (TK6) to the culture to ensure optimal growth of the TK-competent cells (19)(38)(39), and (40) for TK6). General advice on good practice for the maintenance of cell cultures as well as specific advice for L5178Y and TK6 cells can be found in (19)(31)(37)(39)(41). For laboratories requiring master cell stocks to initiate either the MLA or TK6 or to obtain new master cell stocks, a cell repository of well characterised cells is available (37).

Media and culture conditions

17. For both tests, appropriate culture medium and incubation conditions (e.g. culture vessels, humidified atmosphere of 5 % CO₂, incubation temperature of 37°C) should be used for maintaining cultures. Cell cultures should always be maintained under conditions that ensure that they are growing in log phase. It is particularly important to choose media and culture conditions that ensure optimal growth of cells during the expression period and cloning for both mutant and non-mutant cells. For the MLA and the TK6, it is also important that the culture conditions ensure optimal growth of both the large colony/early appearing and the small colony/late appearing TK mutants. More culture details, including the need to properly heat inactivate horse serum if RPMI medium is used during mutant selection can be found in (19)(31)(38)(39)(40)(42).

Preparation of cultures

18. Cells are propagated from stock cultures, seeded in culture medium at a density such that the suspension cultures will continue to grow exponentially through the treatment and expression periods.

Metabolic activation

19. Exogenous metabolising systems should be used when employing L5178Y and TK6 cells because they have inadequate endogenous metabolic capacity. The most commonly used system that is recommended by default unless otherwise justified, is a co-factor-supplemented post-mitochondrial fraction (S9) prepared from the livers of rodents (generally rats) treated with enzyme-inducing agents such as Aroclor 1254 (43)(44)(45) or a combination of phenobarbital and β-naphthoflavone (46)(47)(48)(49)(50)(51). The latter combination does not conflict with the Stockholm Convention on Persistent Organic Pollutants (52) and has been shown to be as effective as Aroclor 1254 for inducing mixed-function oxidases (45)(46)(47)(48)(49). The S9 fraction typically is used at concentrations

ranging from 1-2 % but may be increased to 10 % (v/v) in the final test medium. The choice of type and concentration of exogenous metabolic activation system or metabolic inducer employed may be influenced by the class of test chemicals.

Test chemical preparations

20. Solid test chemicals should be prepared in appropriate solvents and diluted, if appropriate, prior to treatment of the cells (see paragraph 21). Liquid test chemicals may be added directly to the test system and/or diluted prior to treatment of the test system. Gaseous or volatile test chemicals should be tested by appropriate modifications to the standard protocols, such as treatment in sealed culture vessels (53)(54)(55). Preparations of the test chemical should be made just prior to treatment unless stability data demonstrate the acceptability of storage.

TEST CONDITIONS

Solvents

21. The solvent should be chosen to optimise the solubility of the test chemical without adversely impacting the conduct of the test, e.g. changing cell growth, affecting the integrity of the test chemical, reacting with culture vessels, impairing the metabolic activation system. It is recommended that, wherever possible, the use of an aqueous solvent (or culture medium) should be considered first. Well established solvents are water or dimethyl sulfoxide. Generally organic solvents should not exceed 1 % (v/v) and aqueous solvents (saline or water) should not exceed 10 % (v/v) in the final treatment medium. If other than well-established solvents are used (e.g. ethanol or acetone), their use should be supported by data indicating their compatibility with the test chemicals, the test system and their lack of genetic toxicity at the concentration used. In the absence of that supporting data, it is important to add untreated controls (see Appendix 1, Definitions) to demonstrate that no deleterious or mutagenic effects are induced by the chosen solvent.

MEASURING CYTOTOXICITY AND CHOOSING TREATMENT CONCENTRATIONS

22. When determining the highest test chemical concentration, concentrations that have the capability of producing artefactual positive responses, such as those producing excessive cytotoxicity (see paragraph 28), precipitation (see paragraph 29) in the culture medium, or marked changes in pH or osmolality (see paragraph 8), should be avoided. If the test chemical causes a marked change in the pH of the medium at the time of addition, the pH might be adjusted by buffering the final treatment medium so as to avoid artefactual positive results and to maintain appropriate culture conditions.
23. Concentration selection is based on cytotoxicity and other considerations (see paragraphs 27-30). While the evaluation of cytotoxicity in an initial test may be useful to better define the concentrations to be used in the main experiment, an initial test is not required. Even if an initial cytotoxicity evaluation is performed, the measurement of cytotoxicity for each culture is still required in the main experiment. If a range finding experiment is conducted, it should cover a wide range of concentrations and can either be terminated at day 1 after treatment or carried through the 2 day expression and to mutant selection (should it appear that the concentrations used are appropriate).
24. Cytotoxicity should be determined for each individual test culture and control culture: methods for MLA (2) and the TK6 (15) are defined by internationally agreed practice.
25. For both the agar and microwell versions of the MLA: Cytotoxicity should be evaluated using relative total growth (RTG) which was originally defined by Clive and Spector in 1975 (2). This measure includes the relative suspension growth (RSG: test culture vs. solvent control) during the cell treatment, the expression time and the relative cloning efficiency (RCE: test culture vs. solvent control) at the time that mutants are selected (2). It should be noted that the RSG includes any cell loss occurring in the test culture during treatment (See Appendix 2 for formulae).

26. For TK6: Cytotoxicity should be evaluated using relative survival (RS) i.e. cloning efficiency of cells plated immediately after treatment, adjusted for any cell loss during treatment, based on cell count as compared to the negative control (assigned a survival of 100 %) (See Appendix 2 for the formula).
27. At least four test concentrations (not including the solvent and positive controls) that meet the acceptability criteria (appropriate cytotoxicity, number of cells, etc) should be evaluated. While the use of duplicate cultures is advisable, either replicate or single treated cultures may be used at each concentration tested. The results obtained for replicate cultures at a given concentration should be reported separately but can be pooled for the data analysis (55). For test chemicals demonstrating little or no cytotoxicity, concentration intervals of approximately 2 to 3 fold will usually be appropriate. Where cytotoxicity occurs, concentrations should be selected to cover the cytotoxicity range from that producing cytotoxicity as described in paragraph 28 and including concentrations at which there is moderate and little or no cytotoxicity. Many test chemicals exhibit steep concentration response curves and in order to cover the whole range of cytotoxicity or to study the concentration response in detail, it may be necessary to use more closely spaced concentrations and more than four concentrations, in particular in situations where a repeat experiment is required (see paragraph 70). The use of more than 4 concentrations may be particularly important when using single cultures.
28. If the maximum concentration is based on cytotoxicity, the highest concentration should aim to achieve between 20 and 10 % RTG for the MLA, and between 20 and 10 % RS for the TK6 (paragraph 67).
29. For poorly soluble test chemicals that are not cytotoxic at concentrations below the lowest insoluble concentration, the highest concentration analysed should produce turbidity or a precipitate visible by eye or with the aid of an inverted microscope at the end of the treatment with the test chemical. Even if cytotoxicity occurs above the lowest insoluble concentration, it is advisable to test at only one concentration producing turbidity or with a visible precipitate because artifactual effects may result from the precipitate. Because the MLA and TK6 use suspension cultures, particular care should be taken to assure that the precipitate does not interfere with the conduct of the test. The determination of solubility in the culture medium prior to the experiment may also be useful.
30. If no precipitate or limiting cytotoxicity is observed, the highest test concentration should correspond to 10 mM, 2 mg/ml or 2 µl/ml, whichever is the lowest (57)(58). When the test chemical is not of defined composition e.g. substance of unknown or variable composition, complex reaction products or biological materials [i.e. Chemical Substances of unknown or Variable Composition (UVCBs)], environmental extracts etc., the top concentration, may need to be higher (e.g. 5 mg/ml), in the absence of sufficient cytotoxicity, to increase the concentration of each of the components. It should be noted however that these requirements may differ for human pharmaceuticals (59).

Controls

31. Concurrent negative controls (see paragraph 21), consisting of the solvent alone in the treatment medium and handled in the same way as the treatment cultures, should be included for every experimental condition.
32. Concurrent positive controls are needed to demonstrate the ability of the laboratory to identify mutagens under the conditions of the test protocol used, the effectiveness of the exogenous metabolic activation system (when applicable), and to demonstrate adequate detection of both small/late appearing and large/early appearing TK mutants. Examples of positive controls are given in the table 1 below. Alternative positive control substances can be used, if justified. Because *in vitro* mammalian cell tests for genetic toxicity are sufficiently standardised for short-term treatments (3-4 hours) done concurrently with and without metabolic activation using the same treatment duration, the use of positive controls may be confined to a mutagen requiring metabolic activation. In this case, this single positive control response will demonstrate both the activity of the metabolic activation system and the responsiveness of the test system. If used, long term treatment (i.e. 24 hours without S9) should however have its own positive control, as the treatment duration will differ from the test using metabolic activation. Each positive control should be used at one or more concentrations expected to give reproducible and detectable increases over background in order to demonstrate the sensitivity of the test system, and the response should not be compromised by cytotoxicity exceeding the limits specified in this TM (see paragraph 28).

Table 1

Reference substances recommended for assessing laboratory proficiency and for selection of positive controls

Category	Substance	CASRN
1. Mutagens active without metabolic activation		
	Methyl methanesulphonate	66-27-3
	Mitomycin C	50-07-7
	4-Nitroquinoline-N-Oxide	56-57-5
2. Mutagens requiring metabolic activation		
	Benzo(a)pyrene	50-32-8
	Cyclophosphamide (monohydrate)	50-18-0(6055-19-2)
	7,12-Dimethylbenzanthracene	57-97-6
	3-Methylcholanthrene	56-49-5

PROCEDURE

Treatment with test chemical

33. Proliferating cells are treated with the test chemical in the presence and absence of a metabolic activation system. Exposure should be for a suitable period of time (usually 3 to 4 hours is adequate). It should be noted however that these requirements may differ for human pharmaceuticals (59). For MLA, in cases where the short-term treatment yields negative results, and there is information suggesting the need for longer treatment [e.g. nucleoside analogs, poorly soluble chemicals, (5)(59)], consideration should be given to conducting the test with longer treatment, i.e. 24 hours without S9.
34. The minimum number of cells used for each test (control and treated) culture at each stage in the test should be based on the spontaneous mutant frequency. A general guide is to treat and passage sufficient cells in each experimental culture so as to maintain at least 10 but ideally 100 spontaneous mutants in all phases of the test (treatment, phenotypic expression and mutant selection) (56).
35. For MLA the recommended acceptable spontaneous mutant frequency is between $35-140 \times 10^{-6}$ (agar version) and $50-170 \times 10^{-6}$ (microwell version) (see Table 2). To have at least 10 and ideally 100 spontaneous mutants surviving treatment for each test culture, it is necessary to treat at least 6×10^6 cells. Treating this number of cells, and maintaining sufficient cells during expression and cloning for mutant selection, provides for a sufficient number of spontaneous mutants (10 or more) during all phases of the experiment, even for the cultures treated at concentrations that result in 90 % cytotoxicity (as measured by an RTG of 10 %) (19)(38)(39).
36. For the TK6, the spontaneous mutant frequency is generally between 2 and 10×10^{-6} . To have at least 10 spontaneous mutants surviving treatment for each culture, it is necessary to treat at least 20×10^6 cells. Treating this number of cells provides for a sufficient number of spontaneous mutants (10 or more) even for the cultures treated at concentrations that cause 90 % cytotoxicity during treatment (10 % RS). In addition a sufficient number of cells must be cultured during the expression period and plated for mutant selection (60).

Phenotypic expression time and measurement of cytotoxicity and mutant frequency

37. At the end of the treatment period, cells are cultured for a defined time to allow near optimal phenotypic expression of newly induced mutants; specific to each cell line. For the MLA, the phenotypic expression period is 2 days. For the TK6, the phenotypic expression period is 3-4 days. If a 24 hr treatment is used, the expression period begins after the end of treatment.
38. During the phenotypic expression period, cells are enumerated on a daily basis. For the MLA the daily cell counts are used to calculate the daily suspension growth (SG). Following the 2 day expression period, cells are suspended in medium with and without selective agent for the determination of the numbers of mutants (selection plates) and for cloning efficiency (viability plates), respectively. For MLA there are two equally acceptable methods for mutant selection cloning; one using soft agar and the other using liquid medium in 96-well plates (19) (38) (39). Cloning in the TK6 is conducted using liquid media and 96-well plates (16).
39. Trifluorothymidine (TFT) is the only recommended selective agent for TK mutants (61).
40. For the MLA, agar plates and microwell plates are counted after 10-12 days incubation. For the TK6, colonies in microwell plates are scored after 10-14 days for the early appearing mutants. In order to recover the slow growing (late appearing) TK6 mutants, it is necessary to re-feed the cells with growth medium and TFT after counting the early appearing mutants and then to incubate the plates for an additional 7-10 days (62). See paragraphs 42 & 44 for a discussion concerning the enumeration of the slow and normal growth TK mutants.
41. The appropriate calculations for the two tests including the two methods (agar and microwell) for the MLA are in Appendix 2. For the agar method of the MLA, colonies are counted and the number of mutant colonies adjusted by the cloning efficiency to calculate a MF. For the microwell version of the MLA and the TK6, cloning efficiency both for the selection and cloning efficiency plates is determined according to the Poisson distribution (63). The MF is calculated from these two cloning efficiencies.

Mutant Colony characterisation

42. For the MLA, if the test chemical is positive (see paragraphs 62-63), colony characterisation by colony sizing or growth should be performed on at least one of the test cultures (generally the highest acceptable positive concentration) and on the negative and positive controls. If the test chemical is negative (see paragraph 64), mutant colony characterisation should be performed on the negative and positive controls. For the microwell method of the MLA, small colony mutants are defined as those covering less than 25 % of the well's diameter and large colony mutants as those that cover more than 25 % of the well's diameter. For the agar method, an automatic colony counter is used to enumerate the mutant colonies and for colony sizing. Approaches to colony sizing are detailed in the literature (19)(38)(40). Colony characterisation on the negative and positive control is needed to demonstrate that the studies are adequately conducted.
43. The test chemical cannot be determined to be negative if the both the large and small colony mutants are not adequately detected in the positive control. Colony characterisation can be used to provide general information concerning the ability of the test chemical to cause point mutations and/or chromosomal events (paragraph 4).
44. TK6: Normal growing and slow growing mutants are differentiated by a difference in incubation time (see paragraph 40). For the TK6 generally both the early and late appearing mutants are scored for all of the cultures including the negative and positive controls. Colony characterisation of the negative and positive control is needed to demonstrate that the studies are adequately conducted. The test chemical cannot be determined to be negative if both the early appearing and late appearing mutants are not adequately detected in the positive control. Colony characterisation can be used to provide general information concerning the ability of the test chemical to cause point mutations and/or chromosomal events (paragraph 4).

Proficiency of the laboratory

45. In order to demonstrate sufficient experience with the test prior to using it for routine testing, the laboratory should have performed a series of experiments with reference positive substances acting via different mechanisms (at least one active with and one active without metabolic activation selected from the substances listed in Table 1) and various negative controls (including untreated cultures and various solvents/vehicles). These positive and negative control responses should be consistent with the literature. This requirement is not applicable to laboratories that have experience, i.e. that have an historical data base available as defined in paragraphs 47-50. For the MLA the values obtained for both positive and negative controls should be consistent with the IWGT recommendations (see Table 2).
46. A selection of positive control substances (see Table 1) should be investigated with short and long treatments (if using long treatments) in the absence of metabolic activation, and also with short treatment in the presence of metabolic activation, in order to demonstrate proficiency to detect mutagenic chemicals, to determine the effectiveness of the metabolic activation system and to demonstrate the appropriateness of the cell growth conditions during treatment, phenotypic expression and mutant selection and of the scoring procedures. A range of concentrations of the selected substances should be chosen so as to give reproducible and concentration-related increases above the background in order to demonstrate the sensitivity and dynamic range of the test system.

Historical control data

47. The laboratory should establish:
 - A historical positive control range and distribution,
 - A historical negative (untreated, solvent) control range and distribution.
48. When first acquiring data for an historical negative control distribution, concurrent negative controls should be consistent with published negative control data. As more experimental data are added to the control distribution, concurrent negative controls should ideally be within the 95 % control limits of that distribution (64)(65).
49. The laboratory's historical negative control database should initially be built with a minimum of 10 experiments but would preferably consist of at least 20 experiments conducted under comparable experimental conditions. Laboratories should use quality control methods, such as control charts (e.g. C-charts or X-bar charts (65), to identify how variable their positive and negative control data are, and to show that the methodology is 'under control' in their laboratory (66). Further details and recommendations on how to build and use the historical data can be found in the literature (64).
50. Negative control data should consist of mutant frequencies from single or preferably replicate cultures as described in paragraph 27. Concurrent negative controls should ideally be within the 95 % control limits of the distribution of the laboratory's historical negative control database. Where negative control data fall outside the 95 % control limit they may be acceptable for inclusion in the historical control distribution as long as these data are not extreme outliers, there is evidence that the test system is 'under control' (see paragraph 49) and there is evidence of no technical or human failure.
51. Any changes to the experimental protocol should be considered in terms of the consistency of the data with the laboratory's existing historical control databases. Any major inconsistencies should result in the establishment of a new historical control database.

DATA AND REPORTING

Presentation of the results

52. The presentation of data for both the MLA and TK6 should include, for both treated and control cultures, data required for the calculation of cytotoxicity (RTG or RS, respectively) and mutant frequencies, as described below.
53. For MLA, individual culture data should be provided for RSG, RTG, the cloning efficiency at the time of mutant selection and the number of mutant colonies (for agar version) or number of empty wells (for microwell version). MF should be expressed as number of mutant cells per million surviving cells. If the response is positive, small and large colony MFs (and/or percentage of the total MF) should be given for at least one concentration of the test chemical (generally the highest positive concentration) and the negative and positive controls. In the case of a negative response, the small and large colony MF should be given for the negative control and the positive control.
54. For TK6, individual culture data should be provided for RS, the cloning efficiency at the time of mutant selection and the number of empty wells for early appearing and late appearing mutants. MF should be expressed as number of mutant cells per number of surviving cells, and should include the total MF as well as the MF (and/or percentage of the total MF) of the early and late appearing mutants.

Acceptability Criteria

55. For both the MLA and the TK6 the following criteria should be met before determining the overall results for a specific test chemical:
- Two experimental conditions (short treatment with and without metabolic activation see paragraph 33) were conducted unless one resulted in positive results.
 - Adequate number of cells and concentrations should be analysable (see paragraphs 27, 34-36).
 - The criteria for the selection of top concentration are consistent with those described in paragraphs 28-30.

Acceptability criteria for negative and positive controls

56. The IWGT Expert MLA Workgroup analysis of an extensive amount of MLA data resulted in international consensus for specific acceptability criteria for the MLA (1)(2)(3)(4)(5). Therefore, this test method provides specific recommendations for determining the acceptability of negative and positive controls and for evaluating individual substance results in the MLA. The TK6 has a much smaller database and has not undergone evaluation by a workgroup.
57. For MLA, every experiment should be evaluated as to whether the untreated/solvent control meets the IWGT MLA Workgroup acceptance criteria ((4) and Table 2, below) for the: (1) MF (note that the IWGT acceptable MFs are different for the agar and microwell versions of the MLA), (2) cloning efficiency (CE) at the time of mutant selection and (3) suspension growth (SG) for the solvent control (see Appendix 2 for formulae).

Table 2

Acceptability criteria for the MLA

Parameter	Soft Agar Method	Microwell Method
Mutant Frequency	35 – 140 × 10 ⁻⁶	50 – 170 × 10 ⁻⁶
Cloning Efficiency	65 – 120 %	65 – 120 %
Suspension Growth	8 – 32 fold (3-4 hour treatment) 32 – 180 fold (24 hour treatment, if conducted)	8 – 32 fold (3-4 hour treatment) 32 – 180 fold (24 hour treatment, if conducted)

58. For MLA, every test should also be evaluated as to whether the positive control(s) meets at least one of the following two acceptance criteria developed by the IWGT workgroup:
- The positive control should demonstrate an absolute increase in total MF, that is, an increase above the spontaneous background MF [an induced MF (IMF)] of at least 300×10^{-6} . At least 40 % of the IMF should be reflected in the small colony MF.
 - The positive control has an increase in the small colony MF of at least 150×10^{-6} above that seen in the concurrent untreated/solvent control (a small colony IMF of 150×10^{-6}).
59. For the TK6, a test will be acceptable if the concurrent negative control is considered acceptable for addition to the laboratory historical negative control database as described in paragraphs 48-49. In addition, the concurrent positive controls (see paragraph 32) should induce responses that are compatible with those generated in the historical positive control data base and produce a statistically significant increase compared with the concurrent negative control.
60. For both tests, the upper limit of cytotoxicity observed in the positive control culture should be the same as of the experimental cultures. That is, the RTG/RS should not be less than 10 %. It is sufficient to use a single concentration (or one of the concentrations of the positive control cultures if more than one concentration is used) to demonstrate that the acceptance criteria for the positive control have been satisfied. Further, the MF of the positive control must be within the acceptable range established for the laboratory.

Evaluation and interpretation of results

61. For the MLA, significant work on biological relevance and criteria for a positive response has been conducted by The Mouse Lymphoma Expert Workgroup of the IWGT (4). Therefore, this test method provides specific recommendations for the interpretation of test chemical results from the MLA (see paragraphs 62-64). The TK6 has a much smaller database and has not undergone evaluation by a workgroup. Therefore, the recommendations for the interpretation of data for the TK6 are given in more general terms (see paragraphs 65-66). Additional recommendations apply to both tests (see paragraphs 67-71).

MLA

62. An approach for defining positive and negative responses is recommended to assure that the increased MF is biologically relevant. In place of statistical analysis generally used for other tests, it relies on the use of a predefined induced mutant frequency (i.e. increase in MF above concurrent control), designated the Global Evaluation Factor (GEF), which is based on the analysis of the distribution of the negative control MF data from participating laboratories (4). For the agar version of the MLA the GEF is 90×10^{-6} and for the microwell version of the MLA the GEF is 126×10^{-6} .
63. Providing that all acceptability criteria are fulfilled, a test chemical is considered to be clearly positive if, in any of the experimental conditions examined (see paragraph 33), the increase in MF above the concurrent background exceeds the GEF and the increase is concentration related (e.g. using a trend test). The test chemical is then considered able to induce mutation in this test system.
64. Providing that all acceptability criteria are fulfilled, a test chemical is considered to be clearly negative if, in all experimental conditions examined (see paragraph 33) there is no concentration related response or, if there is an increase in MF, it does not exceed the GEF. The test chemical is then considered unable to induce mutations in this test system.

TK6

65. Providing that all acceptability criteria are fulfilled, a test chemical is considered to be clearly positive if, in any of the experimental conditions examined (see paragraph 33):

- at least one of the test concentrations exhibits a statistically significant increase compared with the concurrent negative control
- the increase is concentration-related when evaluated with an appropriate trend test (see paragraph 33)
- any of the results are outside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limit; see paragraph 48).

When all of these criteria are met, the test chemical is then considered able to induce mutation in this test system. Recommendations for the most appropriate statistical methods can be found in the literature (66)(67).

66. Providing that all acceptability criteria are fulfilled, a test chemical is considered clearly negative if, in all experimental conditions examined (see paragraph 33):

- none of the test concentrations exhibits a statistically significant increase compared with the concurrent negative control,
- there is no concentration-related increase when evaluated with an appropriate trend test
- all results are inside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limit; see paragraph 48).

The test chemical is then considered unable to induce mutations in this test system.

For both the MLA and TK6:

67. If the maximum concentration is based on cytotoxicity, the highest concentration should aim to achieve between 20 and 10 % RTG/RS. The consensus is that care should be taken when interpreting positive results only found between 20 and 10 % RTG/RS and a result would not be considered positive if the increase in MF occurred only at or below 10 % RTG/RS (if evaluated) (2)(59).

68. There are some circumstances under which additional information may assist in determining that a test chemical is not mutagenic when there is no culture showing an RTG value between 10-20 % RTG/RS. These situations are outlined as follows: (1) There is no evidence of mutagenicity (e.g. no dose response, no mutant frequencies above those seen in the concurrent negative control or historical background ranges, etc.) in a series of data points within 100 % to 20 % RTG/RS and there is at least one data point between 20 and 25 % RTG/RS. (2) There is no evidence of mutagenicity (e.g. no dose response, no mutant frequencies above those seen in the concurrent negative control or historical background ranges, etc.) in a series of data points between 100 % to 25 % RTG/RS and there is also a negative data point slightly below 10 % RTG/RS. In both of these situations the test chemical can be concluded to be negative.

69. There is no requirement for verification of a clearly positive or negative response.

70. In cases when the response is neither clearly negative nor clearly positive as described above and/or in order to assist in establishing the biological relevance of a result the data should be evaluated by expert judgement and/or further investigations. Performing a repeat experiment possibly using modified experimental conditions [e.g. concentration spacing to increase the probability of attaining data points within the 10-20 % RTG/RS range, using other metabolic activation conditions (i.e. S9 concentration or S9 origin) and duration of treatment] could be useful.

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

TEST REPORT

72. The test report should include the following information:

Test chemical:

- source, lot number, limit date for use, if available;
- stability of the test chemical itself, if known;
- solubility and stability of the test chemical in solvent, if known;
- measurement of pH, osmolality, and precipitate in the culture medium to which the test chemical was added, as appropriate.

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Solvent:

- justification for choice of solvent;
- percentage of solvent in the final culture medium.

Cells:

For Laboratory master cultures:

- type and source of cells, and history in the testing laboratory;
- karyotype features and/or modal number of chromosomes;
- methods for maintenance of cell cultures;
- absence of mycoplasma;
- cell doubling times.

Test conditions:

- rationale for selection of concentrations and number of cell cultures; including e.g. cytotoxicity data and solubility limitations;

- composition of media, CO₂ concentration, humidity level;
- concentration of test chemical expressed as final concentration in the culture medium (e.g. µg or mg/ml or mM of culture medium);
- concentration (and/or volume) of solvent and test chemical added in the culture medium;
- incubation temperature;
- incubation time;
- duration of treatment;
- cell density during treatment;
- type and composition of metabolic activation system (source of S9, method of preparation of the S9 mix, the concentration or volume of S9 mix and S9 in the final culture medium, quality controls of S9);
- positive and negative control substances, final concentrations for each conditions of treatment;
- length of expression period (including number of cells seeded, and subcultures and feeding schedules, if appropriate);
- identity of the selective agent and its concentration;
- for the MLA, the version used (agar or microwell) should be indicated
- criteria for acceptability of the tests;
- methods used to enumerate numbers of viable and mutant cells;
- methods used for the measurements of cytotoxicity;
- any supplementary information relevant to cytotoxicity and method used;
- duration of incubation times after plating;
- definition of colonies of which size and type are considered (including criteria for "small" and "large" colonies, as appropriate);
- criteria for considering studies as positive, negative or equivocal;
- methods used to determine pH, osmolality, if performed and precipitation if relevant.

Results:

- number of cells treated and number of cells sub-cultured for each culture;
- toxicity parameters (RTG for MLA and RS for TK6);
- signs of precipitation and time of the determination;
- number of cells plated in selective and non-selective medium;

- number of colonies in non-selective medium and number of resistant colonies in selective medium and related mutant frequencies;
- colony sizing for the negative and positive controls and if the test chemical is positive, at least one concentration, and related mutant frequencies;
- concentration-response relationship, where possible;
- concurrent negative (solvent) and positive control data (concentrations and solvents);
- historical negative (solvent) and positive control data (concentrations and solvents) with ranges, means and standard deviations; number of tests upon which the historical controls are based;
- statistical analyses (for individual cultures and pooled replicates if appropriate), and p-values if any; and for the MLA, the GEF evaluation.

Discussion of the results

Conclusion

LITERATURE

- (1) Moore, M.M., Honma, M., Clements, J. (Rapporteur), Awogi, T., Bolcsfoldi, G., Cole, J., Gollapudi, B., Harrington-Brock, K., Mitchell, A., Muster, W., Myhr, B., O'Donovan, M., Ouldelhkim, M-C., San, R., Shimada, H. and Stankowski, L.F. Jr. (2000). Mouse Lymphoma Thymidine Kinase Locus (TK) Gene Mutation Assay: International Workshop on Genotoxicity Test Procedures (IWGTP) Workgroup Report, *Environ. Mol. Mutagen.*, 35 (3): 185-190.
- (2) Moore, M.M., Honma, M., Clements, J., Harrington-Brock, K., Awogi, T., Bolcsfoldi, G., Cifone, M., Collard, D., Fellows, M., Flanders, K., Gollapudi, B., Jenkinson, P., Kirby, P., Kirchner, S., Kraycer, J., McEnaney, S., Muster, W., Myhr, B., O'Donovan, M., Oliver, Ouldelhkim, M-C., Pant, K., Preston, R., Riach, C., San, R., Shimada, H. and Stankowski, L.F. Jr. (2002). Mouse Lymphoma Thymidine Kinase Locus Gene Mutation Assay: Follow-Up International Workshop on Genotoxicity Test Procedures, New Orleans, Louisiana, (April 2000), *Environ. Mol. Mutagen.*, 40 (4): 292-299.
- (3) Moore, M.M., Honma, M., Clements, J., Bolcsfoldi, G., Cifone, M., Delongchamp, R., Fellows, M., Gollapudi, B., Jenkinson, P., Kirby, P., Kirchner, S., Muster, W., Myhr, B., O'Donovan, M., Ouldelhkim, M-C., Pant, K., Preston, R., Riach, C., San, R., Stankowski, L.F. Jr., Thakur, A., Wakuri, S. and Yoshimura, I. (2003). Mouse Lymphoma Thymidine Kinase Locus Gene Mutation Assay: International Workshop (Plymouth, UK) on Genotoxicity Test Procedures Workgroup Report, *Mutation Res.*, 540: 127-140.
- (4) Moore, M.M., Honma, M., Clements, J., Bolcsfoldi, G., Burlinson, B., Cifone, M., Clarke, J., Delongchamp, R., Durward, R., Fellows, M., Gollapudi, B., Hou, S., Jenkinson, P., Lloyd, M., Majeska, J., Myhr, B., O'Donovan, M., Omori, T., Riach, C., San, R., Stankowski, L.F. Jr., Thakur, A.K., Van Goethem, F., Wakuri, S. and Yoshimura, I. (2006). Mouse Lymphoma Thymidine Kinase Gene Mutation Assay: Follow-Up Meeting of the International Workshop on Genotoxicity Tests – Aberdeen, Scotland, 2003 – Assay Acceptance Criteria, Positive Controls, and Data Evaluation, *Environ. Mol. Mutagen.*, 47 (1): 1-5.
- (5) Moore, M.M., Honma, M., Clements, J., Bolcsfoldi, G., Burlinson, B., Cifone, M., Clarke, J., Clay, P., Doppalapudi, R., Fellows, M., Gollapudi, B., Hou, S., Jenkinson, P., Muster, W., Pant, K., Kidd, D.A., Lorge, E., Lloyd, M., Myhr, B., O'Donovan, M., Riach, C., Stankowski, L.F. Jr., Thakur, A.K. and Van Goethem, F. (2007). Mouse Lymphoma Thymidine Kinase Mutation Assay: Meeting of the International Workshop on Genotoxicity Testing, San Francisco, 2005, Recommendations for 24-h Treatment, *Mutation. Res.*, 627 (1): 36-40.
- (6) OECD (2016). Overview of the set of OECD Genetic Toxicology Test Guidelines and updates performed in 2014-2015. ENV Publications. Series on Testing and Assessment, No 234, OECD, Paris.

- (7) Fellows M.D., Luker, T., Cooper, A. and O'Donovan, M.R. (2012). Unusual Structure-Genotoxicity Relationship in Mouse Lymphoma Cells Observed with a Series of Kinase Inhibitors. *Mutation, Res.*, 746 (1): 21-28.
- (8) Honma, M., Momose, M., Sakamoto, H., Sofuni, T. and Hayashi, M. (2001). Spindol Poisons Induce Allelic Loss in Mouse Lymphoma Cells Through Mitotic Non-Disjunction. *Mutation Res.*, 493 (1-2): 101-114.
- (9) Wang, J., Sawyer, J.R., Chen, L., Chen, T., Honma, M., Mei, N. and Moore, M.M. (2009). The Mouse Lymphoma Assay Detects Recombination, Deletion, and Aneuploidy, *Toxicol. Sci.*, 109 (1): 96-105.
- (10) Applegate, M.L., Moore, M.M., Broder, C.B., Burrell, A., and Hozier, J.C. (1990). Molecular Dissection of Mutations at the Heterozygous Thymidine Kinase Locus in Mouse Lymphoma Cells. *Proc. National. Academy. Sci. USA*, 87 (1): 51-55.
- (11) Hozier, J., Sawyer, J., Moore, M., Howard, B. and Clive, D. (1981). Cytogenetic Analysis of the L5178Y/TK^{+/-} Leads to TK^{-/-} Mouse Lymphoma Mutagenesis Assay System, *Mutation Res.*, 84 (1): 169-181.
- (12) Hozier, J., Sawyer, J., Clive, D. and Moore, M.M. (1985). Chromosome 11 Aberrations in Small Colony L5178Y TK^{-/-} Mutants Early in their Clonal History, *Mutation Res.*, 147 (5): 237-242.
- (13) Moore, M.M., Clive, D., Hozier, J.C., Howard, B.E., Batson, A.G., Turner, N.T. and Sawyer, J. (1985). Analysis of Trifluorothymidine-Resistant (TFTr) Mutants of L5178Y/TK^{+/-} Mouse Lymphoma Cells. *Mutation Res.*, 151 (1): 161-174.
- (14) Liber H.L., Call K.M. and Little J.B. (1987). Molecular and Biochemical Analyses of Spontaneous and X-Ray-Induced Mutants in Human Lymphoblastoid Cells. *Mutation Res.*, 178 (1): 143-153.
- (15) Li C.Y., Yandell D.W. and Little J.B. (1992). Molecular Mechanisms of Spontaneous and Induced Loss of Heterozygosity in Human Cells *In Vitro*. *Somat. Cell Mol. Genet.*, 18 (1): 77-87.
- (16) Honma M., Hayashi M. and Sofuni T. (1997). Cytotoxic and Mutagenic Responses to X-Rays and Chemical Mutagens in Normal and P53-Mutated Human Lymphoblastoid Cells. *Mutation. Res.*, 374 (1): 89-98.
- (17) Honma, M., Momose, M., Tanabe, H., Sakamoto, H., Yu, Y., Little, J.B., Sofuni, T. and Hayashi, M. (2000). Requirement of Wild-Type P53 Protein for Maintenance of Chromosomal Integrity. *Mol. Carcinogen.*, 28 (4): 203-14.
- (18) Amundson S.A. and Liber H.L. (1992). A Comparison of Induced Mutation at Homologous Alleles of the TK Locus in Human Cells. II. Molecular Analysis of Mutants. *Mutation Res.*, 267 (1): 89-95.
- (19) Schisler M.R., Moore M.M. and Gollapudi B.B. (2013). *In Vitro* Mouse Lymphoma (L5178Y TK^{+/-} -3.7.2C) Forward Mutation Assay. In *Protocols in Genotoxicity Assessment* A. Dhawan and M. Bajpayee (Eds.), Springer Protocols, Humana Press: 27-50.
- (20) Long, L.H., Kirkland, D., Whitwell, J. and Halliwell, B. (2007). Different Cytotoxic and Clastogenic Effects of Epigallocatechin Gallate in Various Cell-Culture Media Due to Variable Rates of its Oxidation in the Culture Medium, *Mutation Res.*, 634 (1-2): 177-183.
- (21) Nesslany, F., Simar-Meintieres, S., Watzinger, M., Talahari, I. and Marzin, D. (2008). Characterization of the Genotoxicity of Nitrilotriacetic Acid. *Environ. Mol. Mutagen.*, 49 (6): 439-452.
- (22) Brusick D. (1986). Genotoxic Effects in Cultured Mammalian Cells Produced by Low pH Treatment Conditions and Increased Ion Concentrations. *Environ. Mutagen.*, 8 (6): 879-886.

- (23) Morita, T., Nagaki, T., Fukuda, I. and Okumura, K. (1992). Clastogenicity of Low pH to Various Cultured Mammalian Cells. *Mutation Res.*, 268 (2): 297-305.
- (24) Scott, D., Galloway, S.M., Marshall, R.R., Ishidate, M.Jr, Brusick, D., Ashby, J. and Myhr, B.C. (1991). Genotoxicity under Extreme Culture Conditions. A report from ICPEMC Task Group 9. *Mutation Res.*, 257: 147-204.
- (25) Wang J., Heflich R.H. and Moore M.M. (2007). A Method to Distinguish Between the *De Novo* Induction of Thymidine Kinase Mutants and the Selection of Pre-Existing Thymidine Kinase Mutants in the Mouse Lymphoma Assay. *Mutation Res.*, 626 (1-2): 185-190.
- (26) Fischer, G.A. (1958). Studies on the Culture of Leukemic Cells *In Vitro*. *Ann. N.Y. Academy Sci.*, 76: 673-680.
- (27) Clive, D., Johnson, K.O., Spector, J.F.S., Batson, A.G. and Brown, M.M.M. (1979). Validation and Characterization of the L5178Y/TK^{+/-} Mouse Lymphoma Mutagen Assay System. *Mutation Res.*, 59(1): 61-108.
- (28) Sawyer, J., Moore, M.M., Clive, D. and Hozier, J. (1985). Cytogenetic Characterization of the L5178Y TK^{+/-} 3.7.2C Mouse Lymphoma Cell Line, *Mutation Res.*, 147 (5): 243-253.
- (29) Sawyer J.R., Moore M.M. and Hozier J.C. (1989). High-Resolution Cytogenetic Characterization of the L5178Y TK^{+/-} Mouse Lymphoma Cell Line, *Mutation Res.*, 214 (2): 181-193.
- (30) Sawyer, J.R., Binz, R.L., Wang, J. and Moore, M.M. (2006). Multicolor Spectral Karyotyping of the L5178Y TK^{+/-}-3.7.2C Mouse Lymphoma Cell Line, *Environ. Mol. Mutagen.*, 47 (2): 127-131.
- (31) Fellows, M.D., McDermott, A., Clare, K.R., Doherty, A. and Aardema, M.J. (2014). The Spectral Karyotype of L5178Y TK^{+/-} Mouse Lymphoma Cells Clone 3.7.2C and Factors Affecting Mutant Frequency at the Thymidine Kinase (TK) Locus in the Microtitre Mouse Lymphoma Assay, *Environ. Mol. Mutagen.*, 55 (1): 35-42.
- (32) Storer, R.D., Jraynak, A.R., McKelvey, T.W., Elia, M.C., Goodrow, T.L. and DeLuca, J.G. (1997). The Mouse Lymphoma L5178Y TK^{+/-} Cell Line is Heterozygous for a Codon 170 Mutation in the P53 Tumor Suppressor Gene. *Mutation. Res.*, 373 (2): 157-165.
- (33) Clark L.S., Harrington-Brock, K., Wang, J., Sargent, L., Lowry, D., Reynolds, S.H. and Moore, M.M. (2004). Loss of P53 Heterozygosity is not Responsible for the Small Colony Thymidine Kinase Mutant Phenotype in L5178Y Mouse Lymphoma Cells. *Mutagen.*, 19 (4): 263-268.
- (34) Skopek T.R., Liber, H.L., Penman, B.W. and Thilly, W.G. (1978). Isolation of a Human Lymphoblastoid Line Heterozygous at the Thymidine Kinase Locus: Possibility for a Rapid Human Cell Mutation Assay. *Biochem. Biophys. Res. Commun.*, 84 (2): 411-416.
- (35) Honma M. (2005). Generation of Loss of Heterozygosity and its Dependency on P53 Status in Human Lymphoblastoid Cells. *Environ. Mol. Mutagen.*, 45 (2-3): 162-176.
- (36) Xia, F., Wang, X., Wang, Y.H., Tsang, N.M., Yandell, D.W., Kelsey, K.T. and Liber, H.L. (1995). Altered P53 Status Correlates with Differences in Sensitivity to Radiation-Induced Mutation and Apoptosis in Two Closely Related Human Lymphoblast Lines. *Cancer. Res.*, 55 (1): 12-15.
- (37) Lorge, E., M. Moore, J. Clements, M. O Donovan, M. Honma, A. Kohara, J. van Benthem, S. Galloway, M.J. Armstrong, V. Thybaud, B. Gollapudi, M. Aardema, J. Kim, A. Sutter, D.J. Kirkland (2015). Standardized Cell Sources and Recommendations for Good Cell Culture Practices in Genotoxicity Testing. (Manuscript in preparation).

- (38) Lloyd M. and Kidd D. (2012). The Mouse Lymphoma Assay. Springer Protocols: Methods in Molecular Biology 817, Genetic Toxicology Principles and Methods, ed. Parry and Parry, Humana Press. ISBN, 978-1-61779-420-9, 35-54.
- (39) Mei N., Guo X. and Moore M.M. (2014). Methods for Using the Mouse Lymphoma Assay to Screen for Chemical Mutagenicity and Photo-Mutagenicity. In: Optimization in Drug Discover: *In Vitro* Methods: Yan Z and Caldwell(Eds), 2nd Edition, GW; Humana Press, Totowa, NJ.
- (40) Liber H.L. and Thilly W.G. (1982). Mutation Assay at the Thymidine Kinase Locus in Diploidhuman Lymphoblasts. *Mutation Res.*, 94 (2): 467-485.
- (41) Coecke, S., Balls, M., Bowe, G., Davis, J., Gstraunthaler, G., Hartung, T., Hay, R., Merten, OW., Price, A., Schechtman, L., Stacey, G. and Stokes, W. (2005). Guidance on Good Cell Culture Practice. A Report of the Second ECVAM Task Force on Good Cell Culture Practice. *ATLA*, 33 (3): 261-287.
- (42) Moore M.M. and Howard B.E. (1982). Quantitation of Small Colony Trifluorothymidine-Resistant Mutants of L5178Y/TK+/- Mouse Lymphoma Cells in RPMI-1640 Medium, *Mutation Res.*, 104 (4-5): 287-294.
- (43) Ames B.N., McCann J. and Yamasaki E. (1975). Methods for Detecting Carcinogens and Mutagens with the Salmonella/Mammalian Microsome Mutagenicity Test. *Mutation Res.*, 31 (6): 347-364.
- (44) Maron D.M. and Ames B.N. (1983). Revised Methods for the Salmonella Mutagenicity Test. *Mutation Res.*, 113 (3-4): 173-215.
- (45) Natarajan, A.T., Tates, A.D, Van Buul, P.P.W., Meijers, M. and De Vogel, N. (1976). Cytogenetic Effects of Mutagens/Carcinogens After Activation in a Microsomal System *In Vitro*, I. Induction of Chromosomal Aberrations and Sister Chromatid Exchanges by Diethylnitrosamine (DEN) and Dimethylnitrosamine (DMN) in CHO Cells in the Presence of Rat-Liver Microsomes. *Mutation Res.*, 37 (1): 83-90.
- (46) Matsuoka A., Hayashi M. and Ishidate M. Jr. (1979). Chromosomal Aberration Tests on 29 Chemicals Combined with S9 Mix *In Vitro*. *Mutation Res.*, 66 (3): 277-290.
- (47) Ong T.M., *et al.* (1980). Differential Effects of Cytochrome P450-Inducers on Promutagen Activation Capabilities and Enzymatic Activities of S-9 from Rat Liver, *J. Environ. Pathol. Toxicol.*, 4 (1): 55-65
- (48) Elliott, B.M., Combes, R.D., Elcombe, C.R., Gatehouse, D.G., Gibson, G.G., Mackay, J.M. and Wolf, R.C. (1992). Report of UK Environmental Mutagen Society Working Party. Alternatives to Aroclor 1254-Induced S9 in *In Vitro* Genotoxicity Assays. *Mutagen.*, 7 (3): 175-177.
- (49) Matsushima, T., Sawamura, M., Hara, K. and Sugimura, T. (1976). A Safe Substitute for Polychlorinated Biphenyls as an Inducer of Metabolic Activation Systems. In: *In Vitro* Metabolic Activation in Mutagenesis Testing. de Serres F.J., *et al.* (Eds, Elsevier, North-Holland, pp. 85-88.
- (50) Galloway S.M., *et al.* (1994). Report from Working Group on *In Vitro* Tests for Chromosomal Aberrations. *Mutation Res.*, 312 (3): 241-261.
- (51) Johnson T.E., Umbenhauer D.R. and Galloway S.M. (1996). Human Liver S-9 Metabolic Activation: Proficiency in Cytogenetic Assays and Comparison with Phenobarbital/Beta-Naphthoflavone or Aroclor 1254 Induced Rat S-9, *Environ. Mol. Mutagen.*, 28 (1): 51-59.
- (52) UNEP (2001). Stockholm Convention on Persistent Organic Pollutants, United Nations Environment Programme (UNEP).

- (53) Krahn D.F., Barsky F.C. and McCooley K.T. (1982). CHO/HGPRT Mutation Assay: Evaluation of Gases and Volatile Liquids. In: Genotoxic Effects of Airborne Agents Tice R.R., Costa D.L. and Schaich K.M. (Eds.). New York, Plenum, pp. 91-103.
- (54) Zamora, P.O., Benson, J.M., Li, A.P. and Brooks, A.L. (1983). Evaluation of an Exposure System Using Cells Grown on Collagen Gels for Detecting Highly Volatile Mutagens in the CHO/HGPRT Mutation Assay. *Environ. Mutagen.*, 5 (6): 795-801.
- (55) Asakura M., Sasaki T., Sugiyama T., Arito H., Fukushima, S. and Matsushima, T. (2008). An Improved System for Exposure of Cultured Mammalian Cells to Gaseous Compounds in the Chromosomal Aberration Assay. *Mutation Res.*, 652 (2): 122-130.
- (56) Arlett C.F., *et al.* (1989). Mammalian Cell Gene Mutation Assays Based upon Colony Formation. In: Statistical Evaluation of Mutagenicity Test Data, Kirkland, D.J. (Ed.), Cambridge University Press, pp. 66-101.
- (57) Morita T., Honma M. and Morikawa K. (2012). Effect of Reducing the Top Concentration Used in the *In Vitro* Chromosomal Aberration Test in CHL Cells on the Evaluation of Industrial Chemical Genotoxicity. *Mutation Res.*, 741 (1-2): 32-56.
- (58) Brookmire L., Chen J.J. and Levy D.D. (2013). Evaluation of the Highest Concentrations Used in the *In Vitro* Chromosome Aberrations Assay. *Environ. Mol. Mutagen.*, 54 (1): 36-43.
- (59) USFDA (2012). International Conference on Harmonisation (ICH) Guidance S2 (R1) on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended For Human Use. Available at: [<https://www.federalregister.gov/a/2012-13774>].
- (60) Honma M. and Hayashi M. (2011). Comparison of *In Vitro* Micronucleus and Gene Mutation Assay Results for P53-Competent Versus P53-Deficient Human Lymphoblastoid Cells. *Environ. Mol. Mutagen.*, 52 (5): 373-384.
- (61) Moore-Brown, M.M., Clive, D., Howard, B.E., Batson, A.G. and Johnson, K.O. (1981). The Utilization of Trifluorothymidine (TFT) to Select for Thymidine Kinase-Deficient (TK^{-/-}) Mutants from L5178Y/TK^{+/-} Mouse Lymphoma Cells, *Mutation Res.*, 85 (5): 363-378.
- (62) Liber H.L., Yandell D.W. and Little J.B. (1989). A Comparison of Mutation Induction at the TK and HRPT Loci in Human Lymphoblastoid Cells; Quantitative Differences are Due to an Additional Class of Mutations at the Autosomal TK locus. *Mutation Res.*, 216 (1): 9-17.
- (63) Furth E.E., Thilly, W.G., Penman, B.W., Liber, H.L. and Rand, W.M. (1981). Quantitative Assay for Mutation in Diploid Human Lymphoblasts Using Microtiter Plates. *Anal. Biochem.*, 110 (1): 1-8.
- (64) Hayashi, M., Dearfield, K., Kasper, P., Lovell, D., Martus, H. J. and Thybaud, V. (2011). Compilation and Use of Genetic Toxicity Historical Control Data, *Mutation Res.*, 723 (2): 87-90.
- (65) Ryan T.P. (2000). *Statistical Methods for Quality Improvement*. John Wiley and Sons, New York 2nd Edition.
- (66) OECD (2014). *Statistical analysis supporting the revision of the genotoxicity Test Guidelines*. Environmental, Health and Safety Publications, Series on Testing and Assessment (No 199.), Organisation for Economic Cooperation and Development, Paris.
- (67) Fleiss J.L., Levin B. and Paik M.C. (2003). *Statistical Methods for Rates and Proportions*, Third Edition, New York: John Wiley & Sons.

Appendix 1

DEFINITIONS

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

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

Base-pair-substitution mutagens: Chemicals that cause substitution of base pairs in the DNA.

Chemical: A substance or a mixture.

Cloning efficiency: The percentage of cells plated at a low density that are able to grow into a colony that can be counted.

Clastogen: Any chemical or process which causes structural chromosomal aberrations in populations of cells or organisms.

Cytotoxicity: For the assays covered in this test method, cytotoxicity is identified as a reduction in relative total growth (RTG) or relative survival (RS) for the MLA and TK6, respectively.

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

Frameshift mutagens: Chemicals which cause the addition or deletion of single or multiple base pairs in the DNA molecule.

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

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

Mutagenic: Produces a heritable change of DNA base-pair sequences(s) in genes or of the structure of chromosomes (chromosome aberrations).

Mutant frequency (MF): The number of mutant cells observed divided by the number of viable cells.

Phenotypic expression time: The time after treatment during which the genetic alteration is fixed within the genome and any pre-existing gene products are depleted to the point that the phenotypic trait is altered.

Relative survival (RS): RS is used as the measure of treatment-related cytotoxicity in the TK6. It is the relative cloning efficiency (CE) of cells plated immediately after the cell treatment adjusted by any loss of cells during treatment as compared with the cloning efficiency of the negative control.

Relative suspension growth (RSG): For the MLA, the relative total two day suspension growth of the test culture compared to the total two-day suspension growth of the negative/solvent control (Clive and Spector, 1975). The RSG should include the relative growth of the test culture compared to the negative/solvent control during the treatment period.

Relative total growth (RTG): RTG is used as the measure of treatment-related cytotoxicity in the MLA. It is a measure of relative (to the vehicle control) growth of test cultures during the treatment, two-day expression and mutant selection cloning phases of the test. The RSG of each test culture is multiplied by the relative cloning efficiency of the test culture at the time of mutant selection and expressed relative to the cloning efficiency of the negative/solvent control (Clive and Spector, 1975).

S9 liver fractions: Supernatant of liver homogenate after 9 000g centrifugation, i.e. raw liver extract

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

Suspension growth (SG): The fold-increase in the number of cells over the course of the treatment and expression phases of the MLA. The SG is calculated by multiplying the fold-increase on day 1 by the fold-increase on day 2 for the short (3 or 4 hr) treatment. If a 24 hr treatment is used the SG is the fold-increase during the 24 hr treatment multiplied by the fold increases on expression days 1 and 2.

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

Test chemical: Any substance or mixture tested using this test method.

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

Appendix 2

FORMULAS

Cytotoxicity

For both versions (agar and microwell) of the MLA

Cytotoxicity is defined as the Relative Total Growth (RTG) which includes the Relative Suspension Growth (RSG) during the 2 day expression period and the Relative Cloning Efficiency (RCE) obtained at the time of mutant selection. RTG, RSG and RCE are all expressed as a percentage.

Calculation of RSG: Suspension Growth one (SG_1) is the growth rate between day 0 and day 1 (cell concentration at day 1 / cell concentration at day 0) and Suspension Growth two (SG_2) is the growth rate between day 1 and day 2 (cell concentration at day 2 / cell concentration at day 1). The RSG is the total SG ($SG_1 \times SG_2$) for the treated culture compared to the untreated/solvent control. That is: $RSG = [SG_{1(test)} \times SG_{2(test)}] / [SG_{1(control)} \times SG_{2(control)}]$ The SG_1 should be calculated from the initial cell concentration used at the beginning of cell treatment. This accounts for any differential cytotoxicity that occurs in the test culture(s) during the cell treatment.

RCE is the relative cloning efficiency of the test culture compared to the relative cloning efficiency of the untreated/solvent control obtained at the time of mutant selection.

Relative Total Growth (RTG): $RTG = RSG \times RCE$

TK6

Relative Survival (RS):

Cytotoxicity is evaluated by relative survival, i.e. cloning efficiency (CE) of cells plated immediately after treatment adjusted by any loss of cells during treatment as compared with cloning efficiency in the negative controls (assigned a survival of 100 %). The adjustment for cell loss during treatment can be calculated as:

$$\text{Adjusted CE} = \text{CE} \times \frac{\text{Number of cells at the end of treatment}}{\text{Number of cells at the beginning of treatment}}$$

The RS for a culture treated by a test chemical is calculated as:

$$RS = \frac{\text{Adjusted CE in the treated culture}}{\text{Adjusted CE in the solvent control}} \times 100$$

Mutant frequency for both the MLA and TK6

Mutant frequency (MF) is the cloning efficiency of mutant colonies in selective medium (CE_M) adjusted by the cloning efficiency in non-selective medium at the time of mutant selection (CE_V). That is, $MF = CE_M / CE_V$. The calculation of these two cloning efficiencies is described below for the agar and microwell cloning methods.

MLA Agar Version: In the soft agar version of the MLA, the number of colonies on the mutant selection plate (C_M) and number of colonies on the unselected or cloning efficiency (viable count) plate (C_V) are obtained by directly counting the clones. When 600 cells are plated for cloning efficiency (CE) for the mutant selection (CE_M) plates and the unselected or cloning efficiency (viable count) plates (CE_V) and 3×10^6 cells are used for mutant selection,

$$CE_M = C_M / (3 \times 10^6) = (C_M / 3) \times 10^{-6}$$

$$CE_V = C_V / 600$$

MLA and TK6 Microwell Version: In the microwell version of the MLA, C_M and C_V are determined as the product of the total number of microwells (TW) and the probable number of colonies per well (P) on microwell plates.

$$C_M = P_M \times TW_M$$

$$C_V = P_V \times TW_V$$

From the zero term of the Poisson distribution (Furth et al., 1981), the P is given by

$$P = -\ln (EW / TW)$$

Where, EW is empty wells and TW is total wells. Therefore,

$$CE_M = C_M / T_M = (P_M \times TW_M) / T_M$$

$$CE_V = C_V / T_V = (P_V \times TW_V) / T_V$$

For the microwell version of the MLA, small and large colony mutant frequencies will be calculated in an identical manner, using the relevant number of empty wells for small and large colonies.

For TK6, small and large colony mutant frequencies are based on the early appearing and late appearing mutants.

B.68 SHORT TIME EXPOSURE IN VITRO TEST METHOD FOR IDENTIFYING i) CHEMICALS INDUCING SERIOUS EYE DAMAGE AND ii) CHEMICALS NOT REQUIRING CLASSIFICATION FOR EYE IRRITATION OR SERIOUS EYE DAMAGE

INTRODUCTION

1. This test method (TM) is equivalent to OECD test guideline (TG) 491 (2017). The Short Time Exposure (STE) test method is an *in vitro* method that can be used under certain circumstances and with specific limitations for hazard classification and labelling of chemicals (substances and mixtures) that induce serious eye damage as well as those that do not require classification for either serious eye damage or eye irritation, as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (1) and the European Union (EU) Regulation 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) (1).
2. For many years, the eye hazard potential of chemicals has been evaluated primarily using an *in vivo* rabbit eye test (TM B.5 (8), equivalent to OECD TG 405). It is generally accepted that, in the foreseeable future, no single *in vitro* alternative test will be able to fully replace the *in vivo* rabbit eye test to predict across the full range of serious eye damage/eye irritation responses for different chemical classes. However, strategic combinations of alternative test methods used in a (tiered) testing strategy may well be able to fully replace the rabbit eye test (2). The top-down approach is designed for the testing of chemicals that can be expected, based on existing information, to have a high irritancy potential or induce serious eye damage. Conversely, the bottom-up approach is designed for the testing of chemicals that can be expected, based on existing information, not to cause sufficient eye irritation to require a classification. While the STE test method is not considered to be a complete replacement for the *in vivo* rabbit eye test, it is suitable for use as part of a tiered testing strategy for regulatory classification and labelling, such as the top-down/bottom-up approach, to identify without further testing (i) chemicals inducing serious eye damage (UN GHS/CLP Category 1) and (ii) chemicals (excluding highly volatile substances and all solid chemicals other than surfactants) that do not require classification for eye irritation or serious eye damage (UN GHS/CLP No Category) (1)(2). However, a chemical that is neither predicted to cause serious eye damage (UN GHS/CLP Category 1) nor UN GHS/CLP No Category (does not induce either serious eye damage or eye irritation) by the STE test method would require additional testing to establish a definitive classification. Furthermore, the appropriate regulatory authorities should be consulted before using the STE in a bottom-up approach under classification schemes other than the UN GHS/CLP. The choice of the most appropriate test method and the use of this test method should be seen in the context of the OECD Guidance Document on an Integrated Approaches on Testing and Assessment for Serious Eye Damage and Eye irritation (14).
3. The purpose of this test method is to describe the procedures used to evaluate the eye hazard potential of a test chemical based on its ability to induce cytotoxicity in the Short Time Exposure Test method. The cytotoxic effect of chemicals on corneal epithelial cells is an important mode of action (MOA) leading to corneal epithelium damage and eye irritation. Cell viability in the STE test method is assessed by the quantitative measurement, after extraction from cells, of blue formazan salt produced by the living cells by enzymatic conversion of the vital dye MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide), also known as Thiazolyl Blue Tetrazolium Bromide (3). The obtained cell viability is compared to the solvent control (relative viability) and used to estimate the potential eye hazard of the test chemical. A test chemical is classified as UN GHS/CLP Category 1 when both the 5 % and 0,05 % concentrations result in a cell viability smaller than or equal to (\leq) 70 %. Conversely, a chemical is predicted as UN GHS/CLP No Category when both 5 % and 0,05 % concentrations result in a cell viability higher than ($>$) 70 %.
4. The term “test chemical” is used in this test method to refer to what is tested and is not related to the applicability of the STE test method to the testing of substances and/or mixtures. Definitions are provided in the Appendix.

INITIAL CONSIDERATIONS AND LIMITATIONS

5. This test method is based on a protocol developed by Kao Corporation (4), which was the subject of two different validation studies: one by the Validation Committee of the Japanese Society for Alternative to Animal Experiments

(1) Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006, OJ L 353/1, 31.12.2008

(JSAAE) (5) and another by the Japanese Center for the Validation of Alternative Methods (JaCVAM) (6). A peer review was conducted by NICEATM/ICCVAM based on the validation study reports and background review documents on the test method (7).

6. When used to identify chemicals (substances and mixtures) inducing serious eye damage (UN GHS/CLP Category 1 (1), data obtained with the STE test method on 125 chemicals (including both substances and mixtures), showed an overall accuracy of 83 % (104/125), a false positive rate of 1 % (1/86), and a false negative rate of 51 % (20/39) as compared to the *in vivo* rabbit eye test (7). The false negative rate obtained is not critical in the present context, since all test chemicals that induce a cell viability of ≤ 70 % at a 5 % concentration and > 70 % at 0,05 % concentration would be subsequently tested with other adequately validated *in vitro* test methods or, as a last option, in the *in vivo* rabbit eye test, depending on regulatory requirements and in accordance with the sequential testing strategy and weight-of-evidence approaches currently recommended (1) (8). Mainly mono-constituent substances were tested, although a limited amount of data also exist on the testing of mixtures. The test method is nevertheless technically applicable to the testing of multi-constituent substances and mixtures. However, before use of this test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed when there is a regulatory requirement for testing of the mixture. The STE test method showed no other specific shortcomings when used to identify test chemicals as UN GHS/CLP Category 1. Investigators could consider using this test method on test chemicals, whereby cell viability ≤ 70 % at both 5 % and 0,05 % concentration should be accepted as indicative of a response inducing serious eye damage that should be classified as UN GHS/CLP Category 1 without further testing.
7. When used to identify chemicals (substances and mixtures) not requiring classification for eye irritation and serious eye damage (i.e. UN GHS/CLP No Category), data obtained with the STE test method on 130 chemicals (including both substances and mixtures), showed an overall accuracy of 85 % (110/130), a false negative rate of 12 % (9/73), and a false positive rate of 19 % (11/57) as compared to the *in vivo* rabbit eye test (7). If highly volatile substances and solid substances other than surfactants are excluded from the dataset, the overall accuracy improves to 90 % (92/102), the false negative rate to 2 % (1/54), and the false positive to 19 % (9/48) (7). As a consequence, the potential shortcomings of the STE test method when used to identify test chemicals not requiring classification for eye irritation and serious eye damage (UN GHS/CLP No Category) are a high false negative rate for i) highly volatile substances with a vapor pressure over 6 kPa and ii) Solid chemicals (substances and mixtures) other than surfactants and mixtures composed only of surfactants. Such chemicals are excluded from the applicability domain of the STE test method (7).
8. In addition to the chemicals mentioned in paragraphs 6 and 7, the STE test method generated dataset also contains in-house data on 40 mixtures, which when compared to the *in vivo* Draize eye test, showed an accuracy of 88 % (35/40), a false positive rate of 50 % (5/10), and a false negative rate of 0 % (0/30) for predicting mixtures that do not require classification under the UN GHS/CLP classification systems (9). The STE test method can therefore be applied to identify mixtures as UN GHS/CLP No Category in a bottom-up approach with the exception of solid mixtures other than those composed only of surfactants as an extension of its limitation to solid substances. Furthermore, mixtures containing substances with vapour pressure higher than 6kPa should be evaluated with care to avoid potential under-predictions, and should be justified on a case-by-case basis.
9. The STE test method cannot be used for the identification of test chemicals as UN GHS/CLP Category 2, or UN GHS Category 2A (eye irritation) or 2B (mild eye irritation), due to the considerable number of UN GHS/CLP Category 1 chemicals under-predicted as Category 2, 2A, or 2B and UN GHS/CLP No Category chemicals over-predicted as Category 2, 2A, or 2B (7). For this purpose, further testing with another suitable method may be required.

10. The STE test method is suitable for test chemicals that are dissolved or uniformly suspended for at least 5 minutes in physiological saline, 5 % dimethyl sulfoxide (DMSO) in saline, or mineral oil. The STE test method is not suitable for test chemicals that are insoluble or cannot be uniformly suspended for at least 5 minutes in physiological saline, 5 % DMSO in saline, or mineral oil. The use of mineral oil in the STE test method is possible because of the short-time exposure. Therefore, the STE test method is suitable for predicting the eye hazard potential of water-insoluble test chemicals (e.g., long-chain fatty alcohols or ketones) provided that they are miscible in at least one of the three above proposed solvents (4).
11. The term "test chemical" is used in this test method to refer to what is being tested ⁽¹⁾ and is not related to the applicability of the STE test method to the testing of substances and/or mixtures.

PRINCIPLE OF THE TEST

12. The STE test method is a cytotoxicity-based *in vitro* assay that is performed on a confluent monolayer of Statens Seruminstitut Rabbit Cornea (SIRC) cells, cultured on a 96-well polycarbonate microplate (4). After five-minute exposure to a test chemical, the cytotoxicity is quantitatively measured as the relative viability of SIRC cells using the MTT assay (4). Decreased cell viability is used to predict potential adverse effects leading to ocular damage.
13. It has been reported that 80 % of a solution dropped into the eye of a rabbit is excreted through the conjunctival sac within three to four minutes, while greater than 80 % of a solution dropped into the human eye is excreted within one to two minutes (10). The STE test method attempts to approximate these exposure times and makes use of cytotoxicity as an endpoint to assess the extent of damage to SIRC cells following a five-minute exposure to the test chemical.

DEMONSTRATION OF PROFICIENCY

14. Prior to routine use of the STE test method described in this test method, laboratories should demonstrate technical proficiency by correctly classifying the eleven substances recommended in Table 1. These substances were selected to represent the full range of responses for serious eye damage or eye irritation based on results of *in vivo* rabbit eye tests (TG 405) and the UN GHS/CLP classification system (1). Other selection criteria included that the substances should be commercially available, that high-quality *in vivo* reference data should be available, and that high-quality *in vitro* data from the STE test method should be available (3). In situations where a listed substance is unavailable or where justifiable, another substance for which adequate *in vivo* and *in vitro* reference data are available could be used provided that the same criteria as described here are used.

Table 1

List of Proficiency Substances

Substance	CASRN	Chemical class ⁽¹⁾	Physical state	<i>In Vivo</i> UN GHS/CLP Cat. ⁽²⁾	Solvent in STE test	STE UN GHS/CLP Cat.
Benzalkonium chloride(10 %, aqueous)	8 001-54-5	Onium compound	Liquid	Category 1	Saline	Category 1

⁽¹⁾ In June 2013, the Joint Meeting agreed that where possible, a more consistent use of the term "test chemical" describing what is being tested should now be applied in new and updated test methods.

Substance	CASRN	Chemical class ⁽¹⁾	Physical state	<i>In Vivo</i> UN GHS/CLP Cat. ⁽²⁾	Solvent in STE test	STE UN GHS/CLP Cat.
Triton X-100 (100 %)	9 002-93-1	Ether	Liquid	Category 1	Saline	Category 1
Acid Red 92	18 472-87-2	Heterocyclic compound; Bromine compound; Chlorine compound	Solid	Category 1	Saline	Category 1
Sodium hydroxide	1 310-73-2	Alkali; Inorganic chemical	Solid	Category 1 ⁽³⁾	Saline	Category 1
Butyrolactone	96-48-0	Lactone; Heterocyclic compound	Liquid	Category 2A (Category 2 in CLP)	Saline	No prediction can be made
1-Octanol	111-87-5	Alcohol	Liquid	Category 2A/B ⁽⁴⁾ (Category 2 in CLP)	Mineral Oil	No prediction can be made
Cyclopentanol	96-41-3	Alcohol; Hydrocarbon, cyclic	Liquid	Category 2A/B ⁽⁵⁾ (Category 2 in CLP)	Saline	No prediction can be made
2-Ethoxyethyl acetate	111-15-9	Alcohol; Ether	Liquid	No Category	Saline	No Category
Dodecane	112-40-3	Hydrocarbon, acyclic	Liquid	No Category	Mineral Oil	No Category
Methyl isobutyl ketone	108-10-1	Ketone	Liquid	No Category	Mineral Oil	No Category

Substance	CASRN	Chemical class ⁽¹⁾	Physical state	<i>In Vivo</i> UN GHS/CLP Cat. ⁽²⁾	Solvent in STE test	STE UN GHS/CLP Cat.
1,1-Dimethyl-guanidine sulfate	598-65-2	Amidine; Sulfur compound	Solid	No Category	Saline	No Category

⁽¹⁾ Chemical classes were assigned using information obtained from previous NICEATM publications and if not available, using the National Library of Medicine's Medical Subject Headings (MeSH[®]) (via ChemIDplus[®] [National Library of Medicine], available at <http://chem.sis.nlm.nih.gov/chemidplus/>) and structure determinations made by NICEATM.

⁽²⁾ Based on results from the *in vivo* rabbit eye test (OECD TG 405) and using the UN GHS/CLP (1).

⁽³⁾ Classification as Cat.1 is based on skin corrosive potential of 100 % sodium hydroxide (listed as a proficiency chemical with skin corrosive potential in OECD TG 435) and the criterion for UN GHS/CLP category 1 (1).

⁽⁴⁾ Classification as 2A or 2B depends on the interpretation of the UN GHS criterion for distinguishing between these two categories, i.e., 2 out of 6 vs 4 out of 6 animals with effects at day 7 necessary to generate a Category 2A classification. The *in vivo* dataset included 2 studies with 3 animals each. In one study two out of three animals showed effects at day 7 warranting a Cat. 2A classification (11), whereas in the second study all endpoints in all three animals recovered to a score of zero by day 7 warranting a Cat. 2B classification (12).

⁽⁵⁾ Classification as 2A or 2B depends on the interpretation of the UN GHS criterion for distinguishing between these two categories, i.e., 1 out of 3 vs 2 out of 3 animals with effects at day 7 necessary to generate a Category 2A classification. The *in vivo* study included 3 animals. All endpoints apart from corneal opacity and conjunctivae redness in one animal recovered to a score of zero by day 7 or earlier. The one animal that did not fully recover by day 7 had a corneal opacity score of 1 and a conjunctivae redness of 1 (at day 7) that fully recovered at day 14 (11).

Abbreviations: CAS RN = Chemical Abstracts Service Registry Number

PROCEDURE

Preparation of the Cellular Monolayer

15. The rabbit cornea cell line, SIRC should be used for performing the STE test method. It is recommended that SIRC cells are obtained from a well-qualified cell bank, such as American Type Culture Collection CCL60.

16. SIRC cells are cultured at 37° C under 5 % CO₂ and humidified atmosphere in a culture flask containing a culture medium comprising Eagle's minimum essential medium (MEM) supplemented with 10 % fetal bovine serum (FBS), 2 mM L-glutamine, 50–100 units/ml penicillin and 50–100 µg/ml streptomycin. Cells that have become confluent in the culture flask should be separated using trypsin-ethylenediaminetetraacetic acid solution, with or without the use of a cell scraper. Cells are propagated (e.g. 2 to 3 passages) in a culture flask before being employed for routine testing, and should undergo no more than 25 passages from thawing.

17. Cells ready to be used for the STE test are then prepared at the appropriate density and seeded into 96-well plates. The recommended cell seeding density is $6,0 \times 10^3$ cells per well when cells are used four days after seeding, or $3,0 \times 10^3$ cells per well when cells are used five days after seeding, at a culture volume of 200 µl. Cells used for the STE test that are seeded in a culture medium at the appropriate density will reach a confluence of more than 80 % at the time of testing, i.e., four or five days after seeding.

Application of the Test Chemicals and Control Substances

18. The first choice of solvent for dissolving or suspending test chemicals is physiological saline. If the test chemical demonstrates low solubility or cannot be dissolved or suspended uniformly for at least five minutes in saline, 5 % DMSO (CAS#67-68-5) in saline is used as a second choice solvent. For test chemicals that cannot be dissolved or suspended uniformly for at least five minutes in either saline or 5 % DMSO in saline, mineral oil (CAS#8 042-47-5) is used as a third choice solvent.

19. Test chemicals are dissolved or suspended uniformly in the selected solvent at 5 % (w/w) concentration and further diluted by serial 10-fold dilution to 0,5 % and 0,05 % concentration. Each test chemical is to be tested at both 5 % and 0,05 % concentrations. Cells cultured in the 96-well plate are exposed to 200 µl/well of either a 5 % or a 0,05 % concentration of the test chemical solution (or suspension), for five minutes at room temperature. Test chemicals (mono-constituent substances or multi-constituent substances or mixtures) are considered as neat substances and diluted or suspended according to the method, regardless of their purity.

20. The culture medium described in paragraph 16 is used as a medium control in each plate of each repetition. Furthermore, cells are to be exposed also to solvent control samples in each plate of each repetition. The solvents listed in paragraph 18 have been confirmed to have no adverse effects on the viability of SIRC cells.

21. In the STE test method, 0,01 % Sodium lauryl sulfate (SLS) in saline is to be used as a positive control in each plate of each repetition. In order to calculate cell viability of the positive control, each plate of each repetition has to also include a saline solvent control.

22. A blank is necessary to determine compensation for optical density and should be performed on wells containing only phosphate buffered saline, but no calcium and magnesium (PBS-) or cells.

23. Each sample (test chemical at 5 % and 0,05 %, medium control, solvent control, and positive control) should be tested in triplicate in each repetition by exposing the cells to 200 µl of the appropriate test or control chemical for five minutes at room temperature.

24. Benchmark substances are useful for evaluating the ocular irritancy potential of unknown chemicals of a specific chemical or product class, or for evaluating the relative irritancy potential of an ocular irritant within a specific range of irritant responses.

Cell Viability Measurement

25. After exposure, cells are washed twice with 200 µl of PBS and 200 µl of MTT solution (0.5 mg MTT/ml of culture medium) is added. After a two-hour reaction time in an incubator (37° C, 5 % CO₂), the MTT solution is decanted, MTT formazan is extracted by adding 200 µl of 0,04 N hydrochloric acid-isopropanol for 60 minutes in the dark at room temperature, and the absorbance of the MTT formazan solution is measured at 570 nm with a plate reader. Interference of test chemicals with the MTT assay (by colorants or direct MTT reducers) only occurs if significant amount of test chemical is retained in the test system following rinsing after exposure which is the case for 3D Reconstructed human cornea or Reconstructed human epidermis tissues but is not relevant for the 2D cell cultures used for the STE test method.

Interpretation of Results and Prediction Model

26. The optical density (OD) values obtained for each test chemical are then used to calculate cell viability relative to the solvent control, which is set at 100 %. The relative cell viability is expressed as a percentage and obtained by dividing the OD of test chemical by the OD of the solvent control after subtracting the OD of blank from both values.

$$\text{Cell viability(\%)} = \frac{(\text{OD}_{570} \text{ of test chemical}) - (\text{OD}_{570} \text{ of blank})}{(\text{OD}_{570} \text{ of solvent control}) - (\text{OD}_{570} \text{ of blank})} \times 100$$

Similarly, the relative cell viability of each solvent control is expressed as a percentage and obtained by dividing the OD of each solvent control by the OD of the medium control after subtracting the OD of blank from both values.

27. Three independent repetitions, each containing three replicate wells (*i.e.*, $n=9$), should be performed. The arithmetic mean of the three wells for each test chemical and solvent control in each independent repetition is used to calculate the arithmetic mean of relative cell viability. The final arithmetic mean of the cell viability is calculated from the three independent repetitions.
28. The cell viability cut-off values for identifying test chemicals inducing serious eye damage (UN GHS/CLP Category 1) and test chemicals not requiring classification for eye irritation or serious eye damage (UN GHS/CLP No Category) are given hereafter.

Table 2

Prediction model of the STE test method

Cell viability		UN GHS/CLP Classification	Applicability
At 5 %	At 0,05 %		
> 70 %	> 70 %	No Category	Substances and mixtures, with the exception of: i) highly volatile substances with a vapor pressure over 6 kPa ⁽¹⁾ and ii) Solid chemicals (substances and mixtures) other than surfactants and mixtures composed only of surfactants
≤ 70 %	> 70 %	No prediction can be made	Not applicable
≤ 70 %	≤ 70 %	Category 1	Substances and mixtures ⁽²⁾

⁽¹⁾ Mixtures containing substances with vapour pressure higher than 6kPa should be evaluated with care to avoid potential under-predictions, and should be justified on a case-by-case basis.

⁽²⁾ Based on results obtained mainly with mono-constituent substances, although a limited amount of data also exist on the testing of mixtures. The test method is nevertheless technically applicable to the testing of multi-constituent substances and mixtures. Before use of this test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture.

Acceptance Criteria

29. Test results are judged to be acceptable when the following criteria are all satisfied:
- a) Optical density of the medium control (exposed to culture medium) should be 0,3 or higher after subtraction of blank optical density.

- b) Viability of the solvent control should be 80 % or higher relative to the medium control. If multiple solvent controls are used in each repetition, each of those controls should show cell viability greater than 80 % to qualify the test chemicals tested with those solvents.
- c) The cell viability obtained with the positive control (0.01 % SLS) should be within two standard deviations of the historical mean. The upper and lower acceptance boundaries for the positive control should be frequently updated i.e., every three months, or each time an acceptable test is conducted in laboratories where tests are conducted infrequently (i.e., less than once a month). Where a laboratory does not complete a sufficient number of experiments to establish a statistically robust positive control distribution, it is acceptable that the upper and lower acceptance boundaries established by the method developer are used, i.e., between 21.1 % and 62.3 % according to its laboratory historical data, while an internal distribution is built during the first routine tests.
- d) Standard deviation of the final cell viability derived from three independent repetitions should be less than 15 % for both 5 % and 0,05 % concentrations of the test chemical.

If one or several of these criteria is not met, the results should be discarded and another three independent repetitions should be conducted.

DATA AND REPORTING

Data

30. Data for each individual well (e.g., cell viability values) of each repetition as well as overall mean, SD, and classification are to be reported.

Test Report

31. The test report should include the following information:

Test Chemical and Control Substances

- Mono-constituent substance: chemical identification, such as IUPAC or CAS name(s), CAS registry number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Multi-constituent substance, UVCB and mixture: Characterisation as far as possible by e.g., chemical identity (see above), purity, quantitative occurrence and relevant physicochemical properties (see above) of the constituents, to the extent available;
- Physical state, volatility, pH, LogP, molecular weight, chemical class, and additional relevant physicochemical properties relevant to the conduct of the study, to the extent available;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc;
- Treatment prior to testing, if applicable (e.g., warming, grinding);
- Storage conditions and stability to the extent available.

Test Method Conditions and Procedures

- Name and address of the sponsor, test facility and study director;
- Description of the test method used;

- Cell line used, its source, passage number and confluence of cells used for testing;
- Details of test procedure used;
- Number of repetitions and replicates used;
- Test chemical concentrations used (if different than the ones recommended);
- Justification for choice of solvent for each test chemical;
- Duration of exposure to the test chemical (if different than the one recommended);
- Description of any modifications of the test procedure;
- Description of evaluation and decision criteria used;
- Reference to historical positive control mean and Standard Deviation (SD);
- Demonstration of proficiency of the laboratory in performing the test method (e.g. by testing of proficiency substances) or demonstration of reproducible performance of the test method over time.

Results

- For each test chemical and control substance, and each tested concentration, tabulation should be given for the individual OD values per replicate well, the arithmetic mean OD values for each independent repetition, the % cell viability for each independent repetition, and the final arithmetic mean % cell viability and SD over the three repetitions;
- Results for the medium, solvent and positive control demonstrating suitable study acceptance criteria;
- Description of other effects observed;
- The overall derived classification with reference to the prediction model/decision criteria used.

Discussion of the Results

Conclusions

LITERATURE

- (1) United Nations UN (2013). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Fifth revised edition. New York & Geneva: United Nations Publications. ISBN: 978-92-1-1170 06-1. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html
- (2) Scott L, *et al.* (2010). A proposed Eye Irritation Testing Strategy to Reduce and Replace *in vivo* Studies Using Bottom-Up and Top-Down Approaches. *Toxicol. In Vitro* 24, 1-9.

- (3) Mosmann T. (1983). Rapid Colorimetric Assay for Cellular Growth and Survival: Application to 7 Proliferation and Cytotoxicity Assays. *J. Immunol. Methods* 65, 55-63.
- (4) Takahashi Y, et al. (2008). Development of the Short Time Exposure (STE) Test: an *In Vitro* Eye Irritation Test Using SIRC Cells. *Toxicol. In Vitro* 22,760-770.
- (5) Sakaguchi H, et al. (2011). Validation Study of the Short Time Exposure (STE) Test to Assess the Eye Irritation Potential of Chemicals. *Toxicol. In Vitro* 25,796-809.
- (6) Kojima H, et al. (2013). Second-Phase Validation of Short Time Exposure Tests for Assessment of Eye Irritation Potency of Chemicals. *Toxicol. In Vitro* 27, pp.1 855-1 869.
- (7) ICCVAM (2013). Short Time Exposure (STE) Test Method Summary Review Document, NIH. Available at: [http://www.ntp.niehs.nih.gov/iccvam/docs/ocutox_docs/STE-SRD-NICEATM-508.pdf].
- (8) Chapter B.5 of this Annex, Acute Eye Irritation/Corrosion.
- (9) Saito K, et al. (2015). Predictive Performance of the Short Time Exposure Test for Identifying Eye Irritation Potential of Chemical Mixtures.
- (10) Mikkelsen TJ, Chrai SS and Robinson JR. (1973). Altered Bioavailability of Drugs in the Eye Due to Drug-Protein Interaction. *J. Pharm. Sci.* 1 648-1 653.
- (11) ECETOC (1998). Eye Irritation Reference Chemicals Data Bank. Technical Report (No 48. (2)), Brussels, Belgium.
- (12) Gautheron P, et al. (1992). Bovine Corneal Opacity and Permeability Test: an *In Vitro* Assay of Ocular Irritancy. *Fundam Appl Toxicol.* 18, 442-449.
- (13) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environmental, Health and Safety Publications, Series on Testing and Assessment (No 34). Organisation for Economic Cooperation and Development, Paris.
- (14) OECD (2017). Guidance Document on an Integrated Approaches on Testing and Assessment for Serious Eye Damage and Eye irritation. Environmental, Health and Safety Publications, Series on Testing and Assessment (No 263). Organisation for Economic Cooperation and Development, Paris.

Appendix

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with concordance to mean the proportion of correct outcomes of a test method (13).

Benchmark substance: A substance used as a standard for comparison to a test chemical. A benchmark substance should have the following properties; (i) a consistent and reliable source(s); (ii) structural and functional similarity to the class of substances being tested; (iii) known physical/chemical characteristics; (iv) supporting data on known effects, and (v) known potency in the range of the desired response.

Bottom-Up Approach: A step-wise approach used for a test chemical suspected of not requiring classification for eye irritation or serious eye damage, which starts with the determination of chemicals not requiring classification (negative outcome) from other chemicals (positive outcome)

Chemical: A substance or mixture.

Eye irritation: Production of change in the eye following the application of a test chemical to the anterior surface of the eye, which are fully reversible within 21 days of application. Interchangeable with “reversible effects on the eye” and with UN GHS/CLP Category 2

False negative rate: The proportion of all positive chemicals falsely identified by a test method as negative. It is one indicator of test method performance.

False positive rate: The proportion of all negative chemicals that are falsely identified by a test method as positive. It is one indicator of test method performance.

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

Medium control: An untreated replicate containing all components of a test system. This sample is processed with test chemical-treated samples and other control samples to determine whether the solvent interacts with the test system.

Mixture: A mixture or a solution composed of two or more substances.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

MTT: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide.

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration $\geq 10\%$ (w/w) and $< 80\%$ (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

OD: Optical Density.

Positive control: A replicate containing all components of a test system and treated with a substance known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (10).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability (13).

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (10).

Serious eye damage: Production of tissue damage in the eye, or serious physical decay of vision, following application of a test chemical to the anterior surface of the eye, which is not fully reversible within 21 days of application. Interchangeable with "irreversible effects on the eye" and with UN GHS/CLP Category 1.

Solvent/vehicle control: An untreated sample containing all components of a test system, including the solvent or vehicle that is processed with the test chemical-treated and other control samples to establish the baseline response for the samples treated with the test chemical dissolved in the same solvent or vehicle. When tested with a concurrent medium control, this sample also demonstrates whether the solvent or vehicle interacts with the test system.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (13).

Substance: A chemical element and its compounds in the natural state or obtained by any production process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Surfactant: Also called surface-active agent, this is a chemical such as a detergent, that can reduce the surface tension of a liquid and thus allow it to foam or penetrate solids; it is also known as a wetting agent.

Test chemical: Any substance or mixture tested using this Test Method.

Tiered testing strategy: A stepwise testing strategy where all existing information on a test chemical is reviewed, in a specified order, using a weight of evidence process at each tier to determine if sufficient information is available for a hazard classification decision, prior to progression to the next tier. If the irritancy potential of a test chemical can be assigned based on the existing information, no additional testing is required. If the irritancy potential of a test chemical cannot be assigned based on the existing information, a step-wise sequential animal testing procedure is performed until an unequivocal classification can be made.

Top-Down Approach: step-wise approach used for a test chemical suspected of causing serious eye damage, which starts with the determination of chemicals inducing serious eye damage (positive outcome) from other chemicals (negative outcome).

United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

UN GHS/CLP Category 1: See "Serious eye damage".

UN GHS/CLP Category 2: See "Eye irritation".

UN GHS/CLP No Category: Chemicals that are not classified as UN GHS/CLP Category 1 or 2 (or UN GHS Category 2A or 2B).

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

B.69 RECONSTRUCTED HUMAN CORNEA-LIKE EPITHELIUM (RhCE) TEST METHOD FOR IDENTIFYING CHEMICALS NOT REQUIRING CLASSIFICATION AND LABELLING FOR EYE IRRITATION OR SERIOUS EYE DAMAGE

INTRODUCTION

1. This test method (TM) is equivalent to OECD test guideline (TG) 492 (2017). *Serious eye damage* refers to the production of tissue damage in the eye, or serious physical decay of vision, following application of a test chemical to the anterior surface of the eye, which is not fully reversible within 21 days of application, as defined by the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) (1) and the European Union (EU) Regulation 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) (1). Also according to UN GHS and CLP, *eye irritation* refers to the production of changes in the eye following the application of a test chemical to the anterior surface of the eye, which are fully reversible within 21 days of application. Test chemicals inducing serious eye damage are classified as UN GHS and CLP Category 1, while those inducing eye irritation are classified as UN GHS and CLP Category 2. Test chemicals not classified for eye irritation or serious eye damage are defined as those that do not meet the requirements for classification as UN GHS and CLP Category 1 or 2 (2A or 2B) i.e., they are referred to as UN GHS and CLP No Category.
2. The assessment of serious eye damage/eye irritation has typically involved the use of laboratory animals (TM B.5 (2)). The choice of the most appropriate test method and the use of this test method should be seen in the context of the OECD Guidance Document on an Integrated Approaches on Testing and Assessment for Serious Eye Damage and Eye irritation (39).
3. This test method describes an *in vitro* procedure allowing the identification of chemicals (substances and mixtures) not requiring classification and labelling for eye irritation or serious eye damage in accordance with UN GHS and CLP. It makes use of reconstructed human cornea-like epithelium (RhCE) which closely mimics the histological, morphological, biochemical and physiological properties of the human corneal epithelium. Four other *in vitro* test methods have been validated, considered scientifically valid and adopted as TM B.47 (3), B.48 (4), B.61 (5) and B.68 (6) to address the human health endpoint serious eye damage/eye irritation.
4. Two validated tests using commercially available RhCE models are included in this test method. Validation studies for assessing eye irritation/serious eye damage have been conducted (7)(8)(9)(10)(11)(12)(13) using the EpiOcular™ Eye Irritation Test (EIT) and the SkinEthic™ Human Corneal Epithelium (HCE) Eye Irritation Test (EIT). Each of these tests makes use of commercially available RhCE tissue constructs as test system, which are referred to in the following text as the Validated Reference Methods – VRM 1 and VRM2, respectively. From these validation studies and their independent peer review (9)(12) it was concluded that the EpiOcular™ EIT and SkinEthic™ HCE EIT are able to correctly identify chemicals (both substances and mixtures) not requiring classification and labelling for eye irritation or serious eye damage according to UN GHS, and the tests were recommended as scientifically valid for that purpose (13).
5. It is currently generally accepted that, in the foreseeable future, no single *in vitro* test method will be able to fully replace the *in vivo* Draize eye test (2)(14) to predict across the full range of serious eye damage/eye irritation responses for different chemical classes. However, strategic combinations of several alternative test methods within (tiered) testing strategies such as the Bottom-Up/Top-Down approach may be able to fully replace the Draize eye test (15). The Bottom-Up approach (15) is designed to be used when, based on existing information, a chemical is expected not to cause sufficient eye irritation to require a classification, while the Top-Down approach (15) is designed to be used when, based on existing information, a chemical is expected to cause serious eye damage. The EpiOcular™ EIT and SkinEthic™ HCE EIT are recommended to identify chemicals that do not require classification for eye irritation or serious eye damage according to UN GHS/CLP (No Category) without further testing, within a testing strategy such as the Bottom-Up/Top-Down approach suggested by Scott *et al.* e.g. as an initial stepin

(1) Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006, OJ L 353/1, 31.12.2008

a Bottom-Up approach or as one of the last steps in a Top-Down approach (15). However, the EpiOcular™ EIT and SkinEthic™ HCE EIT are not intended to differentiate between UN GHS/CLP Category 1 (serious eye damage) and UN GHS/CLP Category 2 (eye irritation). This differentiation will need to be addressed by another tier of a test strategy (15). A test chemical that is identified as requiring classification for eye irritation/serious eye damage with EpiOcular™ EIT or SkinEthic™ HCE EIT will thus require additional testing (*in vitro* and/or *in vivo*) to reach a definitive conclusion (UN GHS/CLP No Category, Category 2 or Category 1), using e.g. TM B.47, B.48, B.61 or B.68.

6. The purpose of this test method is to describe the procedure used to evaluate the eye hazard potential of a test chemical based on its ability to induce cytotoxicity in a RhCE tissue construct, as measured by the MTT assay (16) (see paragraph 21). The viability of the RhCE tissue following exposure to a test chemical is determined in comparison to tissues treated with the negative control substance (% viability), and is then used to predict the eye hazard potential of the test chemical.
7. Performance standards (17) are available to facilitate the validation of new or modified *in vitro* RhCE-based tests similar to EpiOcular™ EIT and SkinEthic™ HCE EIT, in accordance with the principles of the OECD Guidance Document No 34 (18), and allow for timely amendment of OECD TG 492 for their inclusion. Mutual Acceptance of Data (MAD) according to the OECD agreement will only be guaranteed for tests validated according to the performance standards, if these tests have been reviewed and included in the corresponding test guideline by the OECD.

DEFINITIONS

8. Definitions are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

9. This test method is based on commercial three-dimensional RhCE tissue constructs that are produced using either primary human epidermal keratinocytes (i.e., EpiOcular™ OCL-200) or human immortalised corneal epithelial cells (i.e., SkinEthic™ HCE/S). The EpiOcular™ OCL-200 and SkinEthic™ HCE/S RhCE tissue constructs are similar to the *in vivo* corneal epithelium three-dimensional structure and are produced using cells from the species of interest (19)(20). Moreover, the tests directly measure cytotoxicity resulting from penetration of the chemical through the cornea and production of cell and tissue damage; the cytotoxic response then determines the overall *in vivo* serious eye damage/eye irritation outcome. Cell damage can occur by several modes of action (see paragraph 20), but cytotoxicity plays an important, if not the primary, mechanistic role in determining the overall serious eye damage/eye irritation response of a chemical, manifested *in vivo* mainly by corneal opacity, iritis, conjunctival redness and/or conjunctival chemosis, regardless of the physicochemical processes underlying tissue damage.
10. A wide range of chemicals, covering a large variety of chemical types, chemical classes, molecular weights, LogPs, chemical structures, etc., have been tested in the validation study underlying this test method. The EpiOcular™ EIT validation database contained 113 chemicals in total, covering 95 different organic functional groups according to an OECD QSAR toolbox analysis (8). The majority of these chemicals represented mono-constituent substances, but several multi-constituent substances (including 3 homopolymers, 5 copolymers and 10 quasi polymers) were also included in the study. In terms of physical state and UN GHS/CLP Categories, the 113 tested chemicals were distributed as follows: 13 Category 1 liquids, 15 Category 1 solids, 6 Category 2A liquids, 10 Category 2A solids, 7 Category 2B liquids, 7 Category 2B solids, 27 No Category liquids and 28 No Category solids (8). The SkinEthic™ HCE EIT validation database contained 200 chemicals in total, covering 165 different organic functional groups (8)(10)(11). The majority of these chemicals represented mono-constituent substances, but several multi-constituent substances (including 10 polymers) were also included in the study. In terms of physical state and UN GHS/CLP Categories, the 200 tested chemicals were distributed as follows: 27 Category 1 liquids, 24 Category 1 solids, 19 Category 2A liquids, 10 Category 2A solids, 9 Category 2B liquids, 8 Category 2B solids, 50 No Category liquids and 53 No Category solids (10)(11).

11. This test method is applicable to substances and mixtures, and to solids, liquids, semi-solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other pre-treatment of the sample is required. Gases and aerosols have not been assessed in a validation study. While it is conceivable that these can be tested using RhCE technology, the current test method does not allow testing of gases and aerosols.
12. Test chemicals absorbing light in the same range as MTT formazan (naturally or after treatment) and test chemicals able to directly reduce the vital dye MTT (to MTT formazan) may interfere with the tissue viability measurements and need the use of adapted controls for corrections. The type of adapted controls that may be required will vary depending on the type of interference produced by the test chemical and the procedure used to quantify MTT formazan (see paragraphs 36-42).
13. Results generated in pre-validation (21)(22) and full validation (8)(10)(11) studies have demonstrated that both EpiOcular™ EIT and SkinEthic™ HCE EIT are transferable to laboratories considered to be naïve in the conduct of the assays and also to be reproducible within- and between laboratories. Based on these studies, the level of reproducibility in terms of concordance of predictions that can be expected from EpiOcular™ EIT from data on 113 chemicals is in the order of 95 % within laboratories and 93 % between laboratories. The level of reproducibility in terms of concordance of predictions that can be expected from SkinEthic™ HCE EIT from data on 120 chemicals is in the order of 92 % within laboratories and 95 % between laboratories.
14. The EpiOcular™ EIT can be used to identify chemicals that do not require classification for eye irritation or serious eye damage according to the UN GHS and CLP classification system. Considering the data obtained in the validation study (8), the EpiOcular™ EIT has an overall accuracy of 80 % (based on 112 chemicals), sensitivity of 96 % (based on 57 chemicals), false negative rate of 4 % (based on 57 chemicals), specificity of 63 % (based on 55 chemicals) and false positive rate of 37 % (based on 55 chemicals), when compared to reference *in vivo* rabbit eye test data (TM B.5) (2)(14) classified according to the UN GHS and CLP classification system. A study where 97 liquid agrochemical formulations were tested with EpiOcular™ EIT demonstrated a similar performance of the test method for this type of mixtures as obtained in the validation study (23). The 97 formulations were distributed as follows: 21 Category 1, 19 Category 2A, 14 Category 2B and 43 No Category, classified according to the UN GHS classification system based on reference *in vivo* rabbit eye test data (TM B.5) (2)(14). An overall accuracy of 82 % (based on 97 formulations), sensitivity of 91 % (based on 54 formulations), false negative rate of 9 % (based on 54 formulations), specificity of 72 % (based on 43 formulations) and false positive rate of 28 % (based on 43 formulations) were obtained (23).
15. The SkinEthic™ HCE EIT can be used to identify chemicals that do not require classification for eye irritation or serious eye damage according to the UN GHS and CLP classification system. Considering the data obtained in the validation study (10)(11), the SkinEthic™ HCE EIT has an overall accuracy of 84 % (based on 200 chemicals), sensitivity of 95 % (based on 97 chemicals), false negative rate of 5 % (based on 97 chemicals), specificity of 72 % (based on 103 chemicals) and false positive rate of 28 % (based on 103 chemicals), when compared to reference *in vivo* rabbit eye test data (TM B.5) (2)(14) classified according to the UN GHS and CLP classification system.
16. The false negative rates obtained with both RhCE tests, with either substances or mixtures, fall within the 12 % overall probability that chemicals are identified as either UN GHS and CLP Category 2 or UN GHS and CLP No Category by the *in vivo* Draize eye test, in repeated tests; this is due to the method's inherent within-test variability (24). The false positive rates obtained with both RhCE test methods with either substances or mixtures are not critical

in the context of this test method since all test chemicals that produce a tissue viability equal or lower than the established cut-offs (see paragraph 44) will require further testing with other *in vitro* test methods, or as a last option in rabbits, depending on regulatory requirements, using a sequential testing strategy in a weight-of-evidence approach. These test methods can be used for all types of chemicals, whereby a negative result should be accepted for not classifying a chemical for eye irritation and serious eye damage (UN GHS and CLP No Category). The appropriate regulatory authorities should be consulted before using the EpiOcular™ EIT and SkinEthic™ HCE EIT under classification schemes other than UN GHS/CLP.

17. A limitation of this test method is that it does not allow discrimination between eye irritation/reversible effects on the eye (Category 2) and serious eye damage/irreversible effects on the eye (Category 1) as defined by UN GHS and CLP, nor between eye irritants (optional Category 2A) and mild eye irritants (optional Category 2B), as defined by UN GHS (1). For these purposes, further testing with other *in vitro* test methods is required.

18. The term 'test chemical' is used in this test method to refer to what is being tested ⁽²⁾ and is not related to the applicability of the RhCE test method to the testing of substances and/or mixtures.

PRINCIPLE OF THE TEST

19. The test chemical is applied topically to a minimum of two three-dimensional RhCE tissue constructs and tissue viability is measured following exposure and a post-treatment incubation period. The RhCE tissues are reconstructed from primary human epidermal keratinocytes or human immortalised corneal epithelial cells, which have been cultured for several days to form a stratified, highly differentiated squamous epithelium morphologically similar to that found in the human cornea. The EpiOcular™ RhCE tissue construct consists of at least 3 viable layers of cells and a non-keratinised surface, showing a cornea-like structure analogous to that found *in vivo*. The SkinEthic™ HCE RhCE tissue construct consists of at least 4 viable layers of cells including columnar basal cells, transitional wing cells and superficial squamous cells similar to that of the normal human corneal epithelium (20)(26).

20. Chemical-induced serious eye damage/eye irritation, manifested *in vivo* mainly by corneal opacity, iritis, conjunctival redness and/or conjunctival chemosis, is the result of a cascade of events beginning with penetration of the chemical through the cornea and/or conjunctiva and production of damage to the cells. Cell damage can occur by several modes of action, including: cell membrane lysis (e.g. by surfactants, organic solvents); coagulation of macromolecules (particularly proteins) (e.g. by surfactants, organic solvents, alkalis and acids); saponification of lipids (e.g. by alkalis); and alkylation or other covalent interactions with macromolecules (e.g. by bleaches, peroxides and alkylators) (15)(27)(28). However, it has been shown that cytotoxicity plays an important, if not the primary, mechanistic role in determining the overall serious eye damage/eye irritation response of a chemical regardless of the physico-chemical processes underlying tissue damage (29)(30). Moreover, the serious eye damage/eye irritation potential of a chemical is principally determined by the extent of initial injury (31), which correlates with the extent of cell death (29) and with the extent of the subsequent responses and eventual outcomes (32). Thus, slight irritants generally only affect the superficial corneal epithelium, the mild and moderate irritants damage principally the epithelium and superficial stroma and the severe irritants damage the epithelium, deep stroma and at times the corneal endothelium (30)(33). The measurement of viability of the RhCE tissue construct after topical exposure to a test chemical to identify chemicals not requiring classification for serious eye damage/eye irritancy (UN GHS and CLP No Category) is based on the assumption that all chemicals inducing serious eye damage or eye irritation will induce cytotoxicity in the corneal epithelium and/or conjunctiva.

⁽²⁾ In June 2013, the OECD Joint Meeting agreed that where possible, a more consistent use of the term "test chemical" describing what is being tested should now be applied in new and updated OECD test guidelines.

21. RhCE tissue viability is classically measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide; CAS number 298-93-1] by the viable cells of the tissue into a blue MTT formazan salt that is quantitatively measured after extraction from tissues (16). Chemicals not requiring classification and labelling according to UN GHS/CLP (No Category) are identified as those that do not decrease tissue viability below a defined threshold (i.e., tissue viability > 60 %, in EpiOcular™ EIT and SkinEthic™ HCE EITL⁽³⁾, or > 50 %, in SkinEthic™ HCE EITS⁽⁴⁾) (see paragraph 44).

DEMONSTRATION OF PROFICIENCY

22. Prior to routine use of RhCE tests for regulatory purposes, laboratories should demonstrate technical proficiency by correctly predicting the fifteen proficiency chemicals listed in Table 1. These chemicals were selected from the chemicals used in the validation studies of the VRMs (8)(10)(11). The selection includes, to the extent possible, chemicals that: (i) cover different physical states; (ii) cover the full range of *in vivo* serious eye damage/eye irritation responses based on high quality results obtained in the reference *in vivo* rabbit eye test (TM B.5) (2)(14) and the UN GHS classification system (i.e., Categories 1, 2A, 2B, or No Category) (1) and CLP classification system (i.e., Categories 1, 2 or No Category); (iii) cover the various *in vivo* drivers of classification (24)(25); (iv) are representative of the chemical classes used in the validation study (8)(10)(11); (v) cover a good and wide representation of organic functional groups (8)(10)(11); (vi) have chemical structures that are well-defined (8)(10)(11); (vii) are coloured and/or direct MTT reducers; (viii) produced reproducible results in RhCE test methods during their validations; (ix) were correctly predicted by RhCE test methods during their validation studies; (x) cover the full range of *in vitro* responses based on high quality RhCE test methods data (0 to 100 % viability); (xi) are commercially available; and (xii) are not associated with prohibitive acquisition and/or disposal costs. In situations where a listed chemical is unavailable or cannot be used for other justified reasons, another chemical fulfilling the criteria described above, e.g. from the chemicals used in the validation of the VRM, could be used. Such deviations should however be justified.

Table 1

List of proficiency chemicals

Chemical Name	CASRN	Organic Functional Group ⁽¹⁾	Physical State	VRM1 viability (%) ⁽²⁾	VRM2 viability (%) ⁽³⁾	VRM Prediction	MTT Reducer	Colour interf.
In Vivo Category 1⁽⁴⁾								
Methylthioglycolate	2 365-48-2	Carboxylic acid ester; Thioalcohol	L	10,9±6,4	5,5±7,4	No prediction can be made	Y(strong)	N
Hydroxyethyl acrylate	818-61-1	Acrylate;Alcohol	L	7,5±4,7 ⁽⁵⁾	1,6±1,0	No prediction can be made	N	N
2,5-Dimethyl-2,5-hexanediol	110-03-2	Alcohol	S	2,3±0,2	0,2±0,1	No prediction can be made	N	N
Sodium oxalate	62-76-0	Oxocarboxylic acid	S	29,0±1,2	5,3±4,1	No prediction can be made	N	N
In Vivo Category 2A⁽⁴⁾								
2,4,11,13-Tetraazatetradecane-diimidamide, N,N''-bis(4-chlorophenyl)-3,12-diimino-, di-D-gluconate(20 %, aqueous) ⁽⁶⁾	18 472-51-0	Aromatic heterocyclic halide; Aryl halide; Dihydroxyl group; Guanidine	L	4,0±1,1	1,3±0,6	No prediction can be made	N	Y(weak)

⁽³⁾ EITL: EIT for liquids in the case of SkinEthic™ HCE

⁽⁴⁾ EITS: EIT for solids in the case of SkinEthic™ HCE

Chemical Name	CASRN	Organic Functional Group ⁽¹⁾	Physical State	VRM1 viability (%) ⁽²⁾	VRM2 viability (%) ⁽³⁾	VRM Prediction	MTT Reducer	Colour interf.
Sodium benzoate	532-32-1	Aryl; Carboxylic acid	S	3,5±2,6	0,6±0,1	No prediction can be made	N	N

In Vivo Category 2B ⁽⁴⁾

Diethyl toluamide	134-62-3	Benzamide	L	15,6±6,3	2,8±0,9	No prediction can be made	N	N
2,2-Dimethyl-3-methylenebicyclo [2.2.1] heptane	79-92-5	Alkane, branched with tertiary carbon; Alkene; Bicycloheptane; Bridged-ring carbocycles; Cycloalkane	S	4,7±1,5	15,8±1,1	No prediction can be made	N	N

In Vivo No Category ⁽⁴⁾

1-Ethyl-3-methylimidazolium ethylsulphate	3425 73-75-5	Alkoxy; Ammonium salt; Aryl; Imidazole; Sulphate	L	79,9±6,4	79,4±6,2	No Cat	N	N
Dicaprylyl ether	629-82-3	Alkoxy; Ether	L	97,8±4,3	95,2±3,0	No Cat	N	N
Piperonyl butoxide	51-03-6	Alkoxy; Benzodioxole; Benzyl; Ether	L	104,2±4,2	96,5±3,5	No Cat	N	N
Polyethylene glycol (PEG-40) hydrogenated castor oil	61 788-85-0	Acylal; Alcohol; Allyl; Ether	Viscous	77,6±5,4	89,1±2,9	No Cat	N	N
1-(4-Chlorophenyl)-3-(3,4-dichlorophenyl) urea	101-20-2	Aromatic heterocyclic halide; Aryl halide; Urea derivatives	S	106,7±5,3	101,9±6,6	No Cat	N	N
2,2'-Methylenebis-(6-(2H-benzotriazol-2-yl)-4-(1,1,3,3-tetramethylbutyl)-phenol)	1035 97-45-1	Alkane branched with quaternary carbon; Fused carbocyclic aromatic; Fused saturated heterocycles; Precursors quinoid compounds; tert-Butyl	S	102,7±13,4	97,7±5,6	No Cat	N	N

Chemical Name	CASRN	Organic Functional Group ⁽¹⁾	Physical State	VRM1 viability (%) ⁽²⁾	VRM2 viability (%) ⁽³⁾	VRM Prediction	MTT Reducer	Colour interf.
Potassium tetrafluoroborate	14 075-53-7	Inorganic Salt	S	88,6±3,3	92,9±5,1	No Cat	N	N

Abbreviations:

CASRN = Chemical Abstracts Service Registry Number; VRM1 = Validated Reference Method, EpiOcular™ EIT; VRM2 = Validated Reference Method, SkinEthic™ HCE EIT; Colour interf. = colour interference with the standard absorbance (Optical Density (OD)) measurement of MTT formazan.

⁽¹⁾ Organic functional group assigned according to an OECD Toolbox 3,1 nested analysis (8).

⁽²⁾ Based on results obtained with EpiOcular™ EIT in the EURL ECVAM/Cosmetics Europe Eye Irritation Validation Study (EIVS) (8).

⁽³⁾ Based on results obtained with SkinEthic™ HCE EIT in the validation study (10)(11).

⁽⁴⁾ Based on results from the *in vivo* rabbit eye test (TM B.5/OECD TG 405) (2)(14) and using the UN GHS.

⁽⁵⁾ Based on results obtained in the CEFIC Consortium for *in vitro* Eye Irritation testing strategy (CON4EI) Study.

⁽⁶⁾ Classification as 2A or 2B depends on the interpretation of the UN GHS criterion for distinguishing between these two categories, i.e., 1 out of 3 vs 2 out of 3 animals with effects at day 7 necessary to generate a Category 2A classification. The *in vivo* study included 3 animals. All endpoints apart from corneal opacity in one animal recovered to a score of zero by day 7 or earlier. The one animal that did not fully recover by day 7 had a corneal opacity score of 1 (at day 7) that fully recovered at day 9.

23. As part of the proficiency testing, it is recommended that users verify the barrier properties of the tissues after receipt as specified by the RhCE tissue construct producer (see paragraphs 25, 27 and 30). This is particularly important if tissues are shipped over long distance / time periods. Once a test has been successfully established and proficiency in its use has been acquired and demonstrated, such verification will not be necessary on a routine basis. However, when using a test routinely, it is recommended to continue to assess the barrier properties at regular intervals.

PROCEDURE

24. The tests currently covered by this test method are the scientifically valid EpiOcular™ EIT and SkinEthic™ HCE EIT (9)(12)(13), referred to as the Validated Reference Method (VRM1 and VRM2, respectively). The Standard Operating Procedures (SOP) for the RhCE test methods are available and should be employed when implementing and using the test methods in a laboratory (34)(35). The following paragraphs and Appendix 2 describe the main components and procedures of the RhCE tests.

RHCE TEST METHOD COMPONENTS

General conditions

25. Relevant human-derived cells should be used to reconstruct the cornea-like epithelium three-dimensional tissue, which should be composed of progressively stratified but not cornified cells. The RhCE tissue construct is prepared in inserts with a porous synthetic membrane through which nutrients can pass to the cells. Multiple layers of viable, non-keratinised epithelial cells should be present in the reconstructed cornea-like epithelium. The RhCE tissue construct should have the epithelial surface in direct contact with air so as to allow for direct topical exposure of test chemicals in a fashion similar to how the corneal epithelium would be exposed *in vivo*. The RhCE tissue construct should form a functional barrier with sufficient robustness to resist rapid penetration of cytotoxic benchmark substances, e.g. Triton X-100 or sodium dodecyl sulphate (SDS). The barrier function should be demonstrated and may be assessed by determination of either the exposure time required to reduce tissue viability by 50 % (ET₅₀) upon application of a benchmark substance at a specified, fixed concentration (e.g. 100 µl of 0,3 % (v/v) Triton X-100), or the concentration at which a benchmark substance reduces the viability of the tissues by 50 % (IC₅₀) following a fixed exposure time (e.g. 30 minutes treatment with 50 µl SDS) (see paragraph 30). The containment properties of the RhCE tissue construct should prevent the passage of test chemical around the edge of the viable tissue, which could lead to poor modelling of corneal exposure. The human-derived cells used to establish the RhCE tissue construct should be free of contamination by bacteria, viruses, mycoplasma, and fungi. The sterility of the tissue construct should be checked by the supplier for absence of contamination by fungi and bacteria.

Functional conditions*Viability*

26. The assay used for quantifying tissue viability is the MTT assay (16). Viable cells of the RhCE tissue construct reduce the vital dye MTT into a blue MTT formazan precipitate, which is then extracted from the tissue using isopropanol

(or a similar solvent). The extracted MTT formazan may be quantified using either a standard absorbance (Optical Density (OD)) measurement or an HPLC/UPLC-spectrophotometry procedure (36). The OD of the extraction solvent alone should be sufficiently small, i.e. $OD < 0.1$. Users of the RhCE tissue construct should ensure that each batch of the RhCE tissue construct used meets defined criteria for the negative control. Acceptability ranges for the negative control OD values for the VRMs are given in Table 2. An HPLC/UPLC-spectrophotometry user should use the negative control OD ranges provided in Table 2 as the acceptance criterion for the negative control. It should be documented in the test report that the tissues treated with the negative control substance are stable in culture (provide similar tissue viability measurements) for the duration of the test exposure period. A similar procedure should be followed by the tissue producer as part of the quality control tissue batch release, but in this case different acceptance criteria than those specified in Table 2 may apply. An acceptability range (upper and lower limit) for the negative control OD values (in the QC test method conditions) should be established by the RhCE tissue construct developer/supplier.

Table 2

Acceptability ranges for negative control OD values (for the test users)

Test	Lower acceptance limit	Upper acceptance limit
EpiOcular™ EIT (OCL-200) – VRM1 (for both the liquids and the solids protocols)	> 0.8 ⁽¹⁾	< 2,5
SkinEthic™ HCE EIT (HCE/S) – VRM2 (for both the liquids and the solids protocols)	> 1.0	≤ 2,5

⁽¹⁾ This acceptance limit considers the possibility of extended shipping/storage time (e.g. > 4 days), which has been shown not to impact on the performance of the test method (37).

Barrier function

27. The RhCE tissue construct should be sufficiently thick and robust to resist the rapid penetration of cytotoxic benchmark substances, as estimated e.g. by ET_{50} (Triton X-100) or by IC_{50} (SDS) (Table 3). The barrier function of each batch of the RhCE tissue construct used should be demonstrated by the RhCE tissue construct developer/vendor upon supply of the tissues to the end user (see paragraph 30).

Morphology

28. Histological examination of the RhCE tissue construct should demonstrate human cornea-like epithelium structure (including at least 3 layers of viable epithelial cells and a non-keratinised surface). For the VRMs, appropriate morphology has been established by the developer/supplier and therefore does not need to be demonstrated again by a test method user for each tissue batch used.

Reproducibility

29. The results of the positive and negative controls of the test method should demonstrate reproducibility over time.

Quality control (QC)

30. The RhCE tissue construct should only be used if the developer/supplier demonstrates that each batch of the RhCE tissue construct used meets defined production release criteria, among which those for viability (paragraph 26) and barrier function (see paragraph 27) are the most relevant. An acceptability range (upper and lower limits) for the barrier functions as measured by the ET_{50} or IC_{50} (see paragraphs 25 and 26) should be established by the RhCE tissue construct developer/supplier. The ET_{50} and IC_{50} acceptability range used as QC batch release criterion by the developer/supplier of the RhCE tissue constructs (used in the VRMs) is given in Table 3. Data demonstrating compliance with all production release criteria should be provided by the RhCE tissue construct developer/supplier to the test method users so that they are able to include this information in the test report. Only results produced with tissues fulfilling all of these production release criteria can be accepted for reliable prediction of chemicals not requiring classification and labelling for eye irritation or serious eye damage in accordance with UN GHS/CLP.

Table 3

QC batch release criterion

Test	Lower acceptance limit	Upper acceptance limit
EpiOcular™ EIT (OCL-200) – VRM1(100 µl of 0,3 % (v/v) Triton X-100)	ET ₅₀ = 12,2 min	ET ₅₀ = 37,5 min
SkinEthic™ HCE EIT (HCE/S) – VRM2 (30 minutes treatment with 50 µl SDS)	IC ₅₀ = 1 mg/ml	IC ₅₀ = 3,2 mg/ml

Application of the Test Chemical and Control Substances

31. At least two tissue replicates should be used for each test chemical and each control substance in each run. Two different treatment protocols are used, one for liquid test chemicals and one for solid test chemicals (34)(35). For both methods and protocols, the tissue construct surface should be moistened with calcium and magnesium-free Dulbecco's Phosphate Buffered Saline (Ca²⁺/Mg²⁺-free DPBS) before application of test chemicals, to mimic the wet conditions of human eye. The treatment of the tissues is initiated with exposure to the test chemical(s) and control substances. For both treatment protocols in both VRMs, a sufficient amount of test chemical or control substance should be applied to uniformly cover the epithelial surface while avoiding an infinite dose (see paragraphs 32 and 33) (Appendix 2).
32. Test chemicals that can be pipetted at 37°C or lower temperatures (using a positive displacement pipette, if needed) are treated as liquids in the VRMs, otherwise they should be treated as solids (see paragraph 33). In the VRMs, liquid test chemical are evenly spread over the tissue surface (i.e. a minimum of 60 µl/cm² application) (see Appendix 2, (33)(34)). Capillary effects (surface tension effects) that may occur due to the low volumes applied to the insert (on the tissue surface) should be avoided to the extent possible to guarantee the correct dosing of the tissue. Tissues treated with liquid test chemicals are incubated for 30 min at standard culture conditions (37±2°C, 5±1 % CO₂, ≥95 % RH). At the end of the exposure period, the liquid test chemical and the control substances should be carefully removed from the tissue surface by extensive rinsing with Ca²⁺/Mg²⁺-free DPBS at room temperature. This rinsing step is followed by a post-exposure immersion in fresh medium at room temperature (to remove any test chemical absorbed into the tissue) for a pre-defined period of time that varies depending on the VRM used. For VMR1 only, a post-exposure incubation in fresh medium at standard culture conditions is applied prior to performing the MTT assay (see Appendix 2, (34)(35)).
33. Test chemicals that cannot be pipetted at temperatures up to 37°C are treated as solids in the VRMs. The amount of test chemical applied should be sufficient to cover the entire surface of the tissue, i.e. a minimum of 60 mg/cm² application should be used (Appendix 2). Whenever possible, solids should be tested as a fine powder. Tissues treated with solid test chemicals are incubated for a pre-defined period of time (depending on the VRM used) at standard culture conditions (see Appendix 2, (34) (35)). At the end of the exposure period, the solid test chemical and the control substances should be carefully removed from the tissue surface by extensive rinsing with Ca²⁺/Mg²⁺-free DPBS at room temperature. This rinsing step is followed by a post-exposure immersion in fresh medium at room temperature (to remove any test chemical absorbed into the tissue) for a pre-defined period of time that varies depending on the VRM used, and a post-exposure incubation in fresh medium at standard culture conditions, prior to performing the MTT assay (see Appendix 2, (34)(35)).

34. Concurrent negative and positive controls should be included in each run to demonstrate that the viability (determined with the negative control) and the sensitivity (determined with the positive control) of the tissues are within acceptance ranges defined based on historical data. The concurrent negative control also provides the baseline (100 % tissue viability) to calculate the relative percent viability of the tissues treated with the test chemical (%Viability_{test}). The recommended positive control substance to be used with the VRMs is neat methyl acetate (CAS No 79-20-9, commercially available from e.g. Sigma-Aldrich, Cat# 45 997; liquid). The recommended negative control substances to be used with the VRM1 and VRM2 are ultrapure H₂O and Ca²⁺/Mg²⁺-free DPBS, respectively. These were the control substances used in the validation studies of the VRMs and are those for which most historical data exist. The use of suitable alternative positive or negative control substances should be scientifically and adequately justified. Negative and positive controls should be tested with the same protocol(s) as the one(s) used for the test chemicals included in the run (i.e. for liquids and/or solids). This application should be followed by the treatment exposure, rinsing, a post-exposure immersion, and post-exposure incubation where applicable, as described for controls run concurrently to liquid test chemicals (see paragraph 32) or for controls run concurrently to solid test chemicals (see paragraph 33), prior to performing the MTT assay (see paragraph 35) (34)(35). One single set of negative and positive controls is sufficient for all test chemicals of the same physical state (liquids or solids) included in the same run.

Tissue Viability Measurements

35. The MTT assay is a standardised quantitative method (16) that should be used to measure tissue viability under this test method. It is compatible with use in a three-dimensional tissue construct. The MTT assay is performed immediately following the post-exposure incubation period. In the VRMs, the RhCE tissue construct sample is placed in 0,3 ml of MTT solution at 1 mg/ml for 180±15 min at standard culture conditions. The vital dye MTT is reduced into a blue MTT formazan precipitate by the viable cells of the RhCE tissue construct. The precipitated blue MTT formazan product is then extracted from the tissue using an appropriate volume of isopropanol (or a similar solvent) (34)(35). Tissues tested with liquid test chemicals should be extracted from both the top and the bottom of the tissues, while tissues tested with solid test chemicals and coloured liquids should be extracted from the bottom of the tissue only (to minimise any potential contamination of the isopropanol extraction solution with any test chemical that may have remained on the tissue). Tissues tested with liquid test chemicals that are not readily washed off may also be extracted from the bottom of the tissue only. The concurrently tested negative and positive control substances should be treated similarly to the tested chemical. The extracted MTT formazan may be quantified either by a standard absorbance (OD) measurement at 570 nm using a filter band pass of maximum ±30 nm or by using an HPLC/UPLC-spectrophotometry procedure (see paragraph 42) (11)(36).
36. Optical properties of the test chemical or its chemical action on MTT may interfere with the measurement of MTT formazan leading to a false estimate of tissue viability. Test chemicals may interfere with the measurement of MTT formazan by direct reduction of the MTT into blue MTT formazan and/or by colour interference if the test chemical absorbs, naturally or due to treatment procedures, in the same OD range as MTT formazan (i.e., around 570 nm). Pre-checks should be performed before testing to allow identification of potential direct MTT reducers and/or colour interfering chemicals and additional controls should be used to detect and correct for potential interference from such test chemicals (see paragraphs 37-41). This is especially important when a specific test chemical is not completely removed from the RhCE tissue construct by rinsing or when it penetrates the cornea-like epithelium and is therefore present in the RhCE tissue constructs when the MTT assay is performed. For test chemicals absorbing light in the same range as MTT formazan (naturally or after treatment), which are not compatible with the standard absorbance (OD) measurement of MTT formazan due to too strong interference, i.e., strong absorption at 570±30 nm, an HPLC/UPLC-spectrophotometry procedure to measure MTT formazan may be employed (see paragraphs 41 and 42) (11)(36). A detailed description of how to detect and correct for direct MTT reduction and interferences by colouring agents is available in the VRMs SOPs (34)(35). Illustrative flowcharts providing guidance on how to identify and handle direct MTT-reducers and/or colour interfering chemicals for VRM1 and VRM2 are also provided in Appendices III and IV, respectively.

37. To identify potential interference by test chemicals absorbing light in the same range as MTT formazan (naturally or after treatment) and decide on the need for additional controls, the test chemical is added to water and/or isopropanol and incubated for an appropriate time at room temperature (see Appendix 2, (34)(35)). If the test chemical in water and/or isopropanol absorbs sufficient light in the range of 570 ± 20 nm for VRM1 (see Appendix 3), or if a coloured solution is obtained when mixing the test chemical with water for VRM2 (see Appendix 4), the test chemical is presumed to interfere with the standard absorbance (OD) measurement of MTT formazan and further colourant controls should be performed or, alternatively, an HPLC/UPLC-spectrophotometry procedure should be used in which case these controls are not required (see paragraphs 41 and 42 and Appendices III and IV)(34)(35). When performing the standard absorbance (OD) measurement, each interfering test chemical should be applied on at least two viable tissue replicates, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step, to generate a non-specific colour in living tissues (NSC_{living}) control (34)(35). The NSC_{living} control needs to be performed concurrently to the testing of the coloured test chemical and, in case of multiple testing, an independent NSC_{living} control needs to be conducted with each test performed (in each run) due to the inherent biological variability of living tissues. True tissue viability is calculated as: the percent tissue viability obtained with living tissues exposed to the interfering test chemical and incubated with MTT solution (%Viability_{test}) minus the percent non-specific colour obtained with living tissues exposed to the interfering test chemical and incubated with medium without MTT, run concurrently to the test being corrected (%NSC_{living}), i.e., True tissue viability = [%Viability_{test}] - [%NSC_{living}].
38. To identify direct MTT reducers, each test chemical should be added to freshly prepared MTT solution. An appropriate amount of test chemical is added to a MTT solution and the mixture is incubated for approximately 3 hours at standard culture conditions (see Appendices III and IV)(34)(35). If the MTT mixture containing the test chemical (or suspension for insoluble test chemicals) turns blue/purple, the test chemical is presumed to directly reduce MTT and a further functional check on non-viable RhCE tissue constructs should be performed, independently of using the standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure. This additional functional check employs killed tissues that possess only residual metabolic activity but absorb and retain the test chemical in a similar way as viable tissues. Killed tissues of VRM1 are prepared by exposure to low temperature ('freeze-killed'). Killed tissues of VRM2 are prepared by prolonged incubation (e.g. at least 24 ± 1 hours) in water followed by storage to low temperature ('water-killed'). Each MTT reducing test chemical is applied on at least two killed tissue replicates, which undergo the entire testing procedure, to generate a non-specific MTT reduction (NSMTT) control (34)(35). A single NSMTT control is sufficient per test chemical regardless of the number of independent tests/runs performed. True tissue viability is calculated as: the percent tissue viability obtained with living tissues exposed to the MTT reducer (%Viability_{test}) minus the percent non-specific MTT reduction obtained with the killed tissues exposed to the same MTT reducer, calculated relative to the negative control run concurrently to the test being corrected (%NSMTT), i.e., True tissue viability = [%Viability_{test}] - [%NSMTT].
39. Test chemicals that are identified as producing both colour interference (see paragraph 37) and direct MTT reduction (see paragraph 38) will also require a third set of controls when performing the standard absorbance (OD) measurement, apart from the NSMTT and NSC_{living} controls described in the previous paragraphs. This is usually the case with darkly coloured test chemicals absorbing light in the range of 570 ± 30 nm (e.g. blue, purple, black) because their intrinsic colour impedes the assessment of their capacity to directly reduce MTT as described in paragraph 38. This forces the use of NSMTT controls, by default, together with the NSC_{living} controls. Test chemicals for which both NSMTT and NSC_{living} controls are performed may be absorbed and retained by both living and killed tissues. Therefore, in this case, the NSMTT control may not only correct for potential direct MTT reduction by the test chemical, but also for colour interference arising from the absorption and retention of the test chemical by killed tissues. This could lead to double correction for colour interference since the NSC_{living} control already corrects for colour interference arising from the absorption and retention of the test chemical by living tissues. To avoid a possible double correction for colour interference, a third control for non-specific colour inkilled

tissues (NSC_{killed}) needs to be performed (see Appendices III and IV)(34)(35). In this additional control, the test chemical is applied on at least two killed tissue replicates, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step. A single NSC_{killed} control is sufficient per test chemical regardless of the number of independent tests/runs performed, but should be performed concurrently to the NSMTT control and with the same tissue batch. True tissue viability is calculated as: the percent tissue viability obtained with living tissues exposed to the test chemical (%Viability_{test}) minus %NSMTT minus %NSC_{living} plus the percent non-specific colour obtained with killed tissues exposed to the interfering test chemical and incubated with medium without MTT, calculated relative to the negative control ran concurrently to the test being corrected (%NSC_{killed}), i.e., True tissue viability = [%Viability_{test}] - [%NSMTT] - [%NSC_{living}] + [%NSC_{killed}].

40. It is important to note that non-specific MTT reduction and non-specific colour interferences may increase the OD (when performing standard absorbance measurements) of the tissue extract above the linearity range of the spectrophotometer and that non-specific MTT reduction can also increase the MTT formazan peak area (when performing HPLC/UPLC-spectrophotometry measurements) of the tissue extract above the linearity range of the spectrophotometer. On this basis, it is important for each laboratory to determine the OD/peak area linearity range of their spectrophotometer with e.g. MTT formazan (CAS # 57 360-69-7), commercially available from e.g. Sigma-Aldrich (Cat# M2003), before initiating the testing of test chemicals for regulatory purposes.
41. The standard absorbance (OD) measurement using a spectrophotometer is appropriate to assess direct MTT-reducers and colour interfering test chemicals, when the observed interference with the measurement of MTT formazan is not too strong (i.e., the ODs of the tissue extracts obtained with the test chemical without any correction for direct MTT reduction and/or colour interference are within the linear range of the spectrophotometer). Nevertheless, results for test chemicals producing %NSMTT and/or %NSC_{living} $\geq 60\%$ (VRM1, and VRM2 for liquids' protocol) or 50 % (VRM2 for solids' protocol) of the negative control should be taken with caution as this is the established cut-off used in the VRMs to distinguish classified from not classified chemicals (see paragraph 44). Standard absorbance (OD) can however not be measured when the interference with the measurement of MTT formazan is too strong (i.e., leading to uncorrected ODs of the test tissue extracts falling outside of the linear range of the spectrophotometer). Coloured test chemicals or test chemicals that become coloured in contact with water or isopropanol that interfere too strongly with the standard absorbance (OD) measurement of MTT formazan may still be assessed using HPLC/UPLC-spectrophotometry (see Appendices III and IV). This is because the HPLC/UPLC system allows for the separation of the MTT formazan from the chemical before its quantification (36). For this reason, NSC_{living} or NSC_{killed} controls are never required when using HPLC/UPLC-spectrophotometry, independently of the chemical being tested. NSMTT controls should nevertheless be used if the test chemical is suspected to directly reduce MTT (following the procedure described in paragraph 38). NSMTT controls should also be used with test chemicals having a colour (intrinsic or appearing when in water) that impedes the assessment of their capacity to directly reduce MTT as described in paragraph 38. When using HPLC/UPLC-spectrophotometry to measure MTT formazan, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. For test chemicals able to directly reduce MTT, true tissue viability is calculated as: %Viability_{test} minus %NSMTT, as described in the last sentence of paragraph 38. Finally, it should be noted that direct MTT-reducers or direct MTT-reducers that are also colour interfering, which are retained in the tissues after treatment and reduce MTT so strongly that they lead to ODs (using standard OD measurement) or peak areas (using UPLC/HPLC-spectrophotometry) of the tested tissue extracts that fall outside of the linearity range of the spectrophotometer cannot be assessed with RhCE test methods, although these are expected to occur in only very rare situations.

42. HPLC/UPLC-spectrophotometry may be used with all types of test chemicals (coloured, non-coloured, MTT-reducers and non-MTT reducers) for measurement of MTT formazan (11)(36). Due to the diversity of HPLC/UPLC-spectrophotometry systems, it is not feasible for each user to establish the exact same system conditions. As such, qualification of the HPLC/UPLC-spectrophotometry system should be demonstrated before its use to quantify MTT formazan from tissue extracts by meeting the acceptance criteria for a set of standard qualification parameters based on those described in the U.S. Food and Drug Administration guidance for industry on bioanalytical method validation (36)(38). These key parameters and their acceptance criteria are shown in Appendix 5. Once the acceptance criteria defined in Appendix 5 have been met, the HPLC/UPLC-spectrophotometry system is considered qualified and ready to measure MTT formazan under the experimental conditions described in this test method.

Acceptance Criteria

43. For each run using RhCE tissue batches that met the quality control (see paragraph 30), tissues treated with the negative control substance should exhibit OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes and should not be outside the historically established boundaries described in Table 2 (see paragraph 26). Similarly, tissues treated with the positive control substance, i.e., methyl acetate, should show a mean tissue viability < 50 % relative to the negative control in the VRM1 with either the liquids' or the solids' protocols, and ≤ 30 % (liquids' protocol) or ≤ 20 % (solids' protocol) relative to the negative control in the VRM2, thus reflecting the ability of the tissues to respond to an irritant test chemical under the conditions of the test method (34)(35). The variability between tissue replicates of test chemicals and control substances should fall within the accepted limits (i.e., the difference of viability between two tissue replicates should be less than 20 % or the standard deviation (SD) between three tissue replicates should not exceed 18 %). If either the negative control or positive control included in a run is outside of the accepted ranges, the run is considered 'non-qualified' and should be repeated. If the variability between tissue replicates of a test chemical is outside of the accepted range, the test must be considered 'non-qualified' and the test chemical should be re-tested.

Interpretation of Results and Prediction Model

44. The OD values/peak areas obtained with the replicate tissue extracts for each test chemical should be used to calculate the mean percent tissue viability (mean between tissue replicates) normalised to the negative control, which is set at 100 %. The percentage tissue viability cut-off value for identifying test chemicals not requiring classification for eye irritation or serious eye damage (UN GHS and CLP No Category) is given in Table 4. Results should thus be interpreted as follows:

- The test chemical is identified as not requiring classification and labelling according to UN GHS and CLP (No Category) if the mean percent tissue viability after exposure and post-exposure incubation is more than (>) the established percentage tissue viability cut-off value, as shown in Table 4. In this case no further testing in other test methods is required.

- If the mean percent tissue viability after exposure and post-exposure incubation is less than or equal (\leq) to the established percentage tissue viability cut-off value, no prediction can be made, as shown in Table 4. In this case, further testing with other test methods will be required because RhCE test methods show a certain number of false positive results (see paragraphs 14-15) and cannot resolve between UN GHS and CLP Categories 1 and 2 (see paragraph 17).

Table 4

Prediction Models according to UN GHS and CLP classification

VRM	No Category	No prediction can be made
VRM 1 - EpiOcular™ EIT (for both protocols)	Mean tissue viability > 60 %	Mean tissue viability ≤ 60 %
VRM 2 - SkinEthic™ HCE EIT (for the liquids' protocol)	Mean tissue viability > 60 %	Mean tissue viability ≤ 60 %
VRM2 - SkinEthic™ HCE EIT (for the solids' protocol)	Mean tissue viability > 50 %	Mean tissue viability ≤ 50 %

45. A single test composed of at least two tissue replicates should be sufficient for a test chemical when the result is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements and/or mean percent tissue viability equal to $60 \pm 5\%$ (VRM1, and VRM2 for liquids' protocol) or $50 \pm 5\%$ (VRM2 for solids' protocol), a second test should be considered, as well as a third one in case of discordant results between the first two tests.
46. Different percentage tissue viability cut-off values distinguishing classified from non-classified test chemicals may be considered for specific types of mixtures, where appropriate and justifiable, in order to increase the overall performance of the test method for those types of mixtures (see paragraph 14). Benchmark chemicals may be useful for evaluating the serious eye damage/eye irritation potential of unknown test chemicals or product class, or for evaluating the relative ocular toxicity potential of a classified chemical within a specific range of positive responses.

DATA AND REPORTING

Data

47. Data from individual replicate tissues in a run (e.g. OD values/MTT formazan peak areas and calculated percent tissue viability data for the test chemical and controls, and the final RhCE test method prediction) should be reported in tabular form for each test chemical, including data from repeat tests, as appropriate. In addition, mean percent tissue viability and difference of viability between two tissue replicates (if $n=2$ replicate tissues) or SD (if $n \geq 3$ replicate tissues) for each individual test chemical and control should be reported. Any observed interferences of a test chemical with the measurement of MTT formazan through direct MTT reduction and/or coloured interference should be reported for each tested chemical.

Test Report

48. The test report should include the following information:

Test Chemical

Mono-constituent substance

- Chemical identification, such as IUPAC or CAS name(s), CAS registry number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Physical state, volatility, pH, LogP, molecular weight, chemical class, and additional relevant physicochemical properties relevant to the conduct of the study, to the extent available;

- Purity, chemical identity of impurities as appropriate and practically feasible, *etc.*;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Storage conditions and stability to the extent available.

Multi-constituent substance, UVCB and mixture

- Characterisation as far as possible by e.g. chemical identity (see above), purity, quantitative occurrence and relevant physicochemical properties (see above) of the constituents, to the extent available;
- Physical state and additional relevant physicochemical properties relevant to the conduct of the study, to the extent available;
- Purity, chemical identity of impurities as appropriate and practically feasible, *etc.*;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Storage conditions and stability to the extent available.

Positive and Negative Control Substances

- Chemical identification, such as IUPAC or CAS name(s), CAS registry number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Physical state, volatility, molecular weight, chemical class, and additional relevant physicochemical properties relevant to the conduct of the study, to the extent available;
- Purity, chemical identity of impurities as appropriate and practically feasible, *etc.*;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Storage conditions and stability to the extent available;
- Justification for the use of a different negative control than ultrapure H₂O or Ca²⁺/Mg²⁺-free DPBS, if applicable;
- Justification for the use of a different positive control than neat methyl acetate, if applicable;
- Reference to historical positive and negative control results demonstrating suitable run acceptance criteria.

Information Concerning the Sponsor and the Test Facility

- Name and address of the sponsor, test facility and study director.
- *RhCE Tissue Construct and Protocol Used (providing rationale for the choices, if applicable)*

Test Method Conditions

- RhCE tissue construct used, including batch number;
- Wavelength and band pass (if applicable) used for quantifying MTT formazan, and linearity range of measuring device (e.g. spectrophotometer);
- Description of the method used to quantify MTT formazan;
- Description of the HPLC/UPLC-spectrophotometry system used, if applicable;
- Complete supporting information for the specific RhCE tissue construct used including its performance. This should include, but is not limited to:
 - i) Viability quality control (supplier)
 - ii) Viability under test method conditions (user);
 - iii) Barrier function quality control;
 - iv) Morphology, if available;
 - v) Reproducibility and predictive capacity;
 - vi) Other quality controls (QC) of the RhCE tissue construct, if available;
- Reference to historical data of the RhCE tissue construct. This should include, but is not limited to: Acceptability of the QC data with reference to historical batch data;
- Statement that the testing facility has demonstrated proficiency in the use of the test method before routine use by testing of the proficiency chemicals;

Run and Test Acceptance Criteria

- Positive and negative control means and acceptance ranges based on historical data;
- Acceptable variability between tissue replicates for positive and negative controls;
- Acceptable variability between tissue replicates for the test chemical;

Test Procedure

- Details of the test procedure used;
- Doses of test chemical and control substances used;
- Duration and temperature of exposure, post-exposure immersion and post-exposure incubation periods (where applicable);
- Description of any modifications to the test procedure;

- Indication of controls used for direct MTT-reducers and/or colouring test chemicals, if applicable;
- Number of tissue replicates used per test chemical and controls (positive control, negative control, NSMTT, NSCliving and NSKilled, if applicable);

Results

- Tabulation of data from individual test chemicals and control substances for each run (including repeat experiments where applicable) and each replicate measurement, including OD value or MTT formazan peak area, percent tissue viability, mean percent tissue viability, Difference between tissue replicates or SD, and final prediction;
- If applicable, results of controls used for direct MTT-reducers and/or coloured test chemicals, including OD value or MTT formazan peak area, %NSMTT, %NSCliving, %NSKilled, Difference between tissue replicates or SD, final correct percent tissue viability, and final prediction;
- Results obtained with the test chemical(s) and control substances in relation to the define run and test acceptance criteria;
- Description of other effects observed, e.g. coloration of the tissues by a coloured test chemical;

Discussion of the Results

Conclusion

LITERATURE

- (1) UN (2015). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30/Rev.6, Sixth Revised Edition, New York and Geneva: United Nations. Available at: http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev06/English/ST-SG-AC10-30-Rev6e.pdf
- (2) Chapter B.5 of this Annex, Acute Eye Irritation/Corrosion.
- (3) Chapter B.47 of this Annex, Bovine Corneal Opacity and Permeability Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage.
- (4) Chapter B.48 of this Annex, Isolated Chicken Eye Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification.
- (5) Chapter B.61 of this Annex, Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants.
- (6) Chapter B.68 of this Annex, Short Time Exposure *In Vitro* Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage.
- (7) Freeman, S.J., Alépée N., Barroso, J., Cole, T., Compagnoni, A., Rubingh, C., Eskes, C., Lammers, J., McNamee, P., Pfannenbecker, U., Zuang, V. (2010). Prospective Validation Study of Reconstructed Human Tissue Models for Eye Irritation Testing. *ALTEX* 27, Special Issue 2 010,261-266.

- (8) EC EURL ECVAM (2014). The EURL ECVAM - Cosmetics Europe prospective validation study of Reconstructed human Cornea-like Epithelium (RhCE)-based test methods for identifying chemicals not requiring classification and labelling for serious eye damage/eye irritation: Validation Study Report. EUR 28 125 EN; doi:10.2787/41680. Available at: <http://publications.jrc.ec.europa.eu/repository/handle/JRC100280>
- (9) EURL ECVAM Science Advisory Committee (2014). ESAC Opinion on the EURL ECVAM Eye Irritation Validation Study (EIVS) on EpiOcular™ EIT and SkinEthic™ HCE and a related Cosmetics Europe study on HPLC/UPLC-spectrophotometry as an alternative endpoint detection system for MTT-formazan. ESAC Opinion No 2014-03 of 17 November 2014; EUR 28 173 EN; doi: 10.2787/0436 97. Available at: <http://publications.jrc.ec.europa.eu/repository/handle/JRC103702>
- (10) Alépée, N., Leblanc, V., Adriaens, E., Grandidier, M.H., Lelièvre, D., Meloni, M., Nardelli, L., Roper, C.S, Santirocco, E., Toner, F., Van Rompay, A., Vinall, J., Cotovio, J. (2016). Multi-laboratory validation of SkinEthic HCE test method for testing serious eye damage/eye irritation using liquid chemicals. *Toxicol. In Vitro* 31, 43-53.
- (11) Alépée, N., Adriaens, E., Grandidier, M.H., Meloni, M., Nardelli, L., Vinall, C.J., Toner, F., Roper, C.S, Van Rompay, A.R., Leblanc, V., Cotovio, J. (2016). Multi-laboratory evaluation of SkinEthic HCE test method for testing serious eye damage/eye irritation using solid chemicals and overall performance of the test method with regard to solid and liquid chemicals testing. *Toxicol. In Vitro* 34, 55-70.
- (12) URL ECVAM Science Advisory Committee (2016). ESAC Opinion on the SkinEthic™ Human Corneal Epithelium (HCE) Eye Irritation Test (EIT). ESAC Opinion No 2016-02 of 24 June 2016; EUR 28 175 EN; doi: 10.2787/3903 90. Available at: <http://publications.jrc.ec.europa.eu/repository/handle/JRC103704>
- (13) EC EURL ECVAM (2016). Recommendation on the Use of the Reconstructed human Cornea-like Epithelium (RhCE) Test Methods for Identifying Chemicals not Requiring Classification and Labelling for Serious Eye Damage/Eye Irritation According to UN GHS. (Manuscript in Preparation).
- (14) Draize, J.H., Woodard, G., Calvery, H.O. (1944). Methods for the Study of Irritation and Toxicity of Substances Applied Topically to the Skin and Mucous Membranes. *Journal of Pharmacol. and Exp. Therapeutics* 82, 377-390.
- (15) Scott, L., Eskes, C., Hoffmann, S., Adriaens, E., Alépée, N., Bufo, M., Clothier, R., Facchini, D., Faller, C., Guest, R., Harbell, J., Hartung, T., Kamp, H., Le Varlet, B., Meloni, M., McNamee, P., Osborne, R., Pape, W., Pfannenbecker, U., Prinsen, M., Seaman, C., Spielman, H., Stokes, W., Trouba, K., Van den Berghe, C., Van Goethem, F., Vassallo, M., Vinardell, P., Zuang, V. (2010). A Proposed Eye Irritation Testing Strategy to Reduce and Replace *In Vivo* Studies Using Bottom-Up and Top-Down Approaches. *Toxicol. In Vitro* 24, 1-9.
- (16) Mosmann, T. (1983). Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. *J. Immunol. Methods* 65, 55-63.
- (17) OECD (2016). Series on Testing and Assessment No 216: Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Reconstructed Human Cornea-Like Epithelium (RhCE) Test Methods for Identifying Chemicals not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage, Based on the Validated Reference Methods EpiOcular™ EIT and SkinEthic™ HCE EIT described in TG 492. Organisation for Economic Cooperation and Development, Paris.
- (18) OECD (2005). Series on Testing and Assessment No 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Organisation for Economic Cooperation and Development, Paris.

- (19) Kaluzhny, Y., Kandárová, H., Hayden, P., Kubilus, J., d'Argembeau-Thornton, L., Klausner, M. (2011). Development of the EpiOcular™ Eye Irritation Test for Hazard Identification and Labelling of Eye Irritating Chemicals in Response to the Requirements of the EU Cosmetics Directive and REACH Legislation. *Altern. Lab. Anim.* 39, 339-364.
- (20) Nguyen, D.H., Beuerman, R.W., De Wever, B., Rosdy, M. (2003). Three-dimensional construct of the human corneal epithelium for in vitro toxicology. In: Salem, H., Katz, S.A. (Eds), *Alternative Toxicological Methods*, CRC Press, pp. 147-159.
- (21) Pfannenbecker, U., Bessou-Touya, S., Faller, C., Harbell, J., Jacob, T., Raabe, H., Tailhardat, M., Alépée, N., De Smedt, A., De Wever, B., Jones, P., Kaluzhny, Y., Le Varlet, B., McNamee, P., Marrec-Fairley, M., Van Goethem, F. (2013). Cosmetics Europe multi-laboratory pre-validation of the EpiOcular™ reconstituted Human Tissue Test Method for the Prediction of Eye Irritation. *Toxicol. In Vitro* 27, 619-626.
- (22) Alépée, N., Bessou-Touya, S., Cotovio, J., de Smedt, A., de Wever, B., Faller, C., Jones, P., Le Varlet, B., Marrec-Fairley, M., Pfannenbecker, U., Tailhardat, M., van Goethem, F., McNamee, P. (2013). Cosmetics Europe Multi-Laboratory Pre-Validation of the SkinEthic™ Reconstituted Human Corneal Epithelium Test Method for the Prediction of Eye Irritation. *Toxicol. In Vitro* 27, 1 476-1 488.
- (23) Kollé, S.N., Moreno, M.C.R., Mayer, W., van Cott, A., van Ravenzwaay, B., Landsiedel, R. (2015). The EpiOcular™ Eye Irritation Test is the Method of Choice for *In Vitro* Eye Irritation Testing of Agrochemical Formulations: Correlation Analysis of EpiOcular™ Eye Irritation Test and BCOP Test Data to UN GHS, US EPA and Brazil ANIVSA Classifications. *Altern. Lab. Anim.* 43, 1-18.
- (24) Adriaens, E., Barroso, J., Eskes, C., Hoffmann, S., McNamee, P., Alépée, N., Bessou-Touya, S., De Smedt, A., De Wever, B., Pfannenbecker, U., Tailhardat, M., Zuang, V. (2014). Retrospective Analysis of the Draize Test for Serious Eye Damage/Eye Irritation: Importance of Understanding the *In Vivo* Endpoints Under UN GHS/EU CLP for the Development and Evaluation of *In Vitro* Test Methods. *Arch. Toxicol.* 88, 701-723.
- (25) Barroso, J., Pfannenbecker, U., Adriaens, E., Alépée, N., Cluzel, M., De Smedt, A., Hibatallah, J., Klaric, M., Mewes, K.R., Millet, M., Templier, M., McNamee, P. (2017). Cosmetics Europe compilation of historical serious eye damage/eye irritation *in vivo* data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative methods/strategies: the Draize eye test Reference Database (DRD). *Arch. Toxicol.* 91, 521-547.
- (26) Meloni, M., De Servi, B., Marasco, D., Del Prete, S. (2011). Molecular mechanism of ocular surface damage: Application to an *in vitro* dry eye model on human corneal epithelium. *Molecular Vision* 17, 113-126.
- (27) Hackett, R.B., McDonald, T.O. (1991). Eye Irritation. In *Advances in Modern Toxicology: Dermatotoxicology* Marzulli F.N. and Maibach H.I. (Eds.), 4th Edition, pp. 749–815. Washington, DC, USA: Hemisphere Publishing Corporation.
- (28) Fox, D.A., Boyes, W.K. (2008). Toxic Responses of the Ocular and Visual System. In Cassaret and Doull's *Toxicology: The Basic Science of Poisons* Klaassen C.D.(Ed.), 7th Edition, pp. 665–697. Withby, ON, Canada: McGraw-Hill Ryerson.
- (29) Jester, J.V., Li, H.F., Petroll, W.M., Parker, R.D., Cavanagh, H.D., Carr, G.J., Smith, B., Maurer, J.K. (1998). Area and Depth of Surfactant Induced Corneal Injury Correlates with Cell Death. *Invest. Ophthalmol. Vis. Sci.* 39, 922–936.

- (30) Maurer, J.K., Parker, R.D., Jester, J.V. (2002). Extent of Corneal Injury as the Mechanistic Basis for Ocular Irritation: Key Findings and Recommendations for the Development of Alternative Assays. *Reg. Tox. Pharmacol.* 36, 106-117.
- (31) Jester, J.V., Li, L., Molai, A., Maurer, J.K. (2001). Extent of Corneal Injury as a Mechanistic Basis for Alternative Eye Irritation Tests. *Toxicol. In Vitro* 15, 115–130.
- (32) Jester, J.V., Petroll, W.M., Bean, J., Parker, R.D., Carr, G.J., Cavanagh, H.D., Maurer, J.K. (1998). Area and Depth of Surfactant-Induced Corneal Injury Predicts Extent of Subsequent Ocular Responses. *Invest. Ophthalmol. Vis. Sci.* 39, 2 610–2 625.
- (33) Jester, J.V. (2006). Extent of Corneal Injury as a Biomarker for Hazard Assessment and the Development of Alternative Models to the Draize Rabbit Eye Test. *Cutan. Ocul. Toxicol.* 25, 41–54.
- (34) EpiOcular™ EIT SOP, Version 8 (March 05, 2013). EpiOcular™ EIT for the Prediction of Acute Ocular Irritation of Chemicals. Available at: [<https://ecvam-dbalm.jrc.ec.europa.eu/beta/index.cfm/methodsAndProtocols/index>].
- (35) SkinEthic™ HCE EIT SOP, Version 1. (July 20, 2015). SkinEthic™ HCE Eye Irritation Test (EITL for Liquids, EITS for Solids) for the Prediction of Acute Ocular Irritation of Chemicals. Available at: <https://ecvam-dbalm.jrc.ec.europa.eu/beta/index.cfm/methodsAndProtocols/index>
- (36) Alépée, N., Barroso, J., De Smedt, A., De Wever, B., Hibatallah, J., Klaric, M., Mewes, K.R., Millet, M., Pfannenbecker, U., Tailhardat, M., Templier, M., McNamee, P. (2015). Use of HPLC/UPLC-Spectrophotometry for Detection of Formazan in *In Vitro* Reconstructed Human Tissue (RhT)-Based Test Methods Employing the MTT-Reduction Assay to Expand their Applicability to Strongly Coloured Test Chemicals. *Toxicol. In Vitro* 29, 741-761.
- (37) Kaluzhny, Y., Kandárová, H., Handa, Y., DeLuca, J., Truong, T., Hunter, A., Kearney, P., d'Argembeau-Thornton, L., Klausner, M. (2015). EpiOcular™ Eye Irritation Test (EIT) for Hazard Identification and Labeling of Eye Irritating Chemicals: Protocol Optimization for Solid Materials and Extended Shipment Times. *Altern. Lab Anim.* 43, 101-127.
- (38) US FDA (2001). Guidance for Industry: Bioanalytical Method Validation. U.S. Department of Health and Human Services, Food and Drug Administration. May 2001. Available at: <http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf>
- (39) OECD (2017). Guidance Document on an Integrated Approaches on Testing and Assessment for Serious Eye Damage and Eye irritation. Series on Testing and Assessment No 263. ENV Publications, Organisation for Economic Cooperation and Development, Paris.

Appendix 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "concordance", to mean the proportion of correct outcomes of a test method (18).

Benchmark chemical: A chemical used as a standard for comparison to a test chemical. A benchmark chemical should have the following properties: (i) consistent and reliable source(s) for its identification and characterisation; (ii) structural, functional and/or chemical or product class similarity to the chemical(s) being tested; (iii) known physicochemical characteristics; (iv) supporting data on known effects; and (v) known potency in the range of the desired response.

Bottom-Up approach: Step-wise approach used for a test chemical suspected of not requiring classification and labelling for eye irritation or serious eye damage, which starts with the determination of chemicals not requiring classification and labelling (negative outcome) from other chemicals (positive outcome).

Chemical: A substance or mixture.

Concordance: See "Accuracy".

Cornea: The transparent part of the front of the eyeball that covers the iris and pupil and admits light to the interior.

CV: Coefficient of Variation.

Dev: Deviation.

EIT: Eye Irritation Test.

EURL ECVAM: European Union Reference Laboratory for Alternatives to Animal Testing.

Eye irritation: Production of changes in the eye following the application of a test chemical to the anterior surface of the eye, which are fully reversible within 21 days of application. Interchangeable with "Reversible effects on the eye" and with "UN GHS/CLP Category 2".

ET₅₀: Exposure time required to reduce tissue viability by 50 % upon application of a benchmark chemical at a specified, fixed concentration.

False negative rate: The proportion of all positive substances falsely identified by a test method as negative. It is one indicator of test method performance.

False positive rate: The proportion of all negative substances that are falsely identified by a test method as positive. It is one indicator of test method performance.

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

HCE: SkinEthic™ Human Corneal Epithelium.

HPLC: High Performance Liquid Chromatography.

IC₅₀: Concentration at which a benchmark chemical reduces the viability of the tissues by 50 % following a fixed exposure time (e.g. 30 minutes treatment with SDS).

Infinite dose: Amount of test chemical applied to the RhCE tissue construct exceeding the amount required to completely and uniformly cover the epithelial surface.

Irreversible effects on the eye: See “Serious eye damage”.

LLOQ: Lower Limit of Quantification.

LogP: Logarithm of the octanol-water partitioning coefficient

Mixture: A mixture or a solution composed of two or more substances.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration ≥ 10 % (w/w) and < 80 % (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

MTT: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide.

Negative control: A sample containing all components of a test system and treated with a substance known not to induce a positive response in the test system. This sample is processed with test chemical-treated samples and other control samples and is used to determine 100 % tissue viability.

Not Classified: Chemicals that are not classified for Eye irritation (UN GHS/CLP Category 2, UN GHS Category 2A or 2B) or Serious eye damage (UN GHS/CLP Category 1). Interchangeable with “UN GHS/CLP No Category”.

NSC_{killed}: Non-Specific Colour in killed tissues.

NSC_{living}: Non-Specific Colour in living tissues.

NSMTT: Non-Specific MTT reduction.

OD: Optical Density.

Performance standards: Standards, based on a validated test method which was considered scientifically valid, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are: (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (18).

Positive control: A sample containing all components of a test system and treated with a substance known to induce a positive response in the test system. This sample is processed with test chemical-treated samples and other control samples. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (18).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability (18).

Replacement test: A test which is designed to substitute for a test that is in routine use and accepted for hazard identification and/or risk assessment, and which has been determined to provide equivalent or improved protection of human or animal health or the environment, as applicable, compared to the accepted test, for all possible testing situations and chemicals (18).

Reproducibility: The agreement among results obtained from repeated testing of the same test chemical using the same test protocol (See "Reliability") (18).

Reversible effects on the eye: See "Eye irritation".

RhCE: Reconstructed human Cornea-like Epithelium.

Run: A run consists of one or more test chemicals tested concurrently with a negative control and with a positive control.

SD: Standard Deviation.

Sensitivity: The proportion of all positive/active test chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (18).

Serious eye damage: Production of tissue damage in the eye, or serious physical decay of vision, following application of a test substance to the anterior surface of the eye, which is not fully reversible within 21 days of application. Interchangeable with "Irreversible effects on the eye" and with "UN GHS and CLP Category 1".

Standard Operating Procedures (SOP): Formal, written procedures that describe in detail how specific routine, and test-specific, laboratory operations should be performed. They are required by GLP.

Specificity: The proportion of all negative/inactive test chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (18).

Substance: A chemical element and its compounds in the natural state or obtained by any production process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Test: A single test chemical concurrently tested in a minimum of two tissue replicates as defined in the corresponding SOP.

Tissue viability: Parameter measuring total activity of a cell population in a reconstructed tissue as their ability to reduce the vital dye MTT, which, depending on the endpoint measured and the test design used, correlates with the total number and/or vitality of living cells.

Top-Down approach: Step-wise approach used for a chemical suspected of causing serious eye damage, which starts with the determination of chemicals inducing serious eye damage (positive outcome) from other chemicals (negative outcome).

Test chemical: Any substance or mixture tested using this test method.

Tiered testing strategy: A stepwise testing strategy, which uses test methods in a sequential manner. All existing information on a test chemical is reviewed at each tier, using a weight-of-evidence process, to determine if sufficient information is available for a hazard classification decision, prior to progression to the next tier in the strategy. If the hazard potential/potency of a test chemical can be assigned based on the existing information at a given tier, no additional testing is required (18).

ULOQ: Upper Limit of Quantification.

United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS): A system proposing the classification of chemicals (substances and mixtures) according to standardised types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

UN GHS and CLP Category 1: See "Serious eye damage".

UN GHS and CLP Category 2: See "Eye irritation".

UN GHS and CLP No Category: Chemicals that do not meet the requirements for classification as UN GHS/CLP Category 1 or 2 (or UN GHS Category 2A or 2B). Interchangeable with "Not Classified".

UPLC: Ultra-High Performance Liquid Chromatography.

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

Valid test method: A test method considered to have sufficient relevance and reliability for a specific purpose and which is based on scientifically sound principles. A test method is never valid in an absolute sense, but only in relation to a defined purpose (18).

Validated test method: A test method for which validation studies have been completed to determine the relevance (including accuracy) and reliability for a specific purpose. It is important to note that a validated test method may not have sufficient performance in terms of accuracy and reliability to be found acceptable for the proposed purpose (18).

VRM: Validated Reference Method.

VRM1: EpiOcular™ EIT is referred as the Validated Reference Method 1.

VRM2: SkinEthic™ HCE EIT is referred to as the Validated Reference Method 2.

Weight-of-evidence: The process of considering the strengths and weaknesses of various pieces of information in reaching and supporting a conclusion concerning the hazard potential of a test substance.

Appendix 2

MAIN TEST COMPONENTS OF THE RHCE TESTS VALIDATED FOR IDENTIFYING CHEMICALS NOT REQUIRING CLASSIFICATION AND LABELLING FOR EYE IRRITATION OR SEVERE EYE DAMAGE

Test Components	EpiOcular™ EIT (VRM 1)	SkinEthic™ HCE EIT (VRM 2)
Protocols	Liquids (pipetteable at 37±1 °C or lower temperatures for 15 min)	Liquids and viscous (pipetteable)
Model surface	Solids (not pipetteable)	Solids (not pipetteable)
Number of tissue replicates	0,6 cm ²	0,5 cm ²
Pre-check for colour interference	At least 2	At least 2
	50 µl + 1 ml H ₂ O for 60 min at 37±2 °C, 5±1 % CO ₂ , ≥95 % RH (non-coloured test chemicals), or 50 µl + 2 ml isopropanol mixed for 2-3h at RT (coloured test chemicals) →if the OD of the test chemical at 570±20 nm, after subtraction of the OD for isopropanol or water is > 0.08 (which corresponds to approximately 5 % of the mean OD of the negative control), living adapted controls should be performed.	10 µl + 90 µl H ₂ O mixed for 30±2 min at Room Temperature (RT, 18-28 °C) →if test chemical is coloured, living adapted controls should be performed
	50 mg + 1 ml H ₂ O for 60 min at 37±2 °C, 5±1 % CO ₂ , ≥95 % RH(non-coloured test chemicals) and/or 50 mg + 2 ml isopropanol mixed for 2-3h at RT (colored and non-colored test chemicals) →if the OD of the test chemical at 570±20 nm after subtraction of the OD for isopropanol or water is > 0.08 (which corresponds to approximately 5 % of the mean OD of the negative control), living adapted controls should be performed.	10 mg + 90 µl H ₂ O mixed for 30±2 min at RT →if test chemical is coloured, living adapted controls should be performed

Test Components	EpiOcular™ EIT (VRM 1)	EpiOcular™ EIT (VRM 1)	SkinEthic™ HCE EIT (VRM 2)
Pre-check for direct MTT reduction	<p>50 µl + 1 ml MTT 1 mg/ml solution for 180±15 min at 37±2 °C, 5±1 % CO₂, ≥95 % RH</p> <p>→if solution turns blue/purple, freeze-killed adapted controls should be performed (50 µl of sterile deionised water in MTT solution is used as negative control)</p>	<p>50 mg + 1 ml MTT 1 mg/ml solution for 180±15 min at 37±2 °C, 5±1 % CO₂, ≥95 % RH</p> <p>→if solution turns blue/purple, freeze-killed adapted controls should be performed (50 µl of sterile deionised water in MTT solution is used as negative control)</p>	<p>30 µl + 300 µl MTT 1 mg/ml solution for 180±15 min at 37±2 °C, 5±1 % CO₂, ≥95 % RH</p> <p>→if solution turns blue/purple, water-killed adapted controls should be performed (30 µl of sterile deionised water in MTT solution is used as negative control)</p>
Pre-treatment	<p>20 µl Ca²⁺/Mg²⁺-free DPBS for 30 ± 2 min at 37±2 °C, 5±1 % CO₂, ≥95 % RH, protected from light.</p>	<p>20 µl Ca²⁺/Mg²⁺-free DPBS for 30±2 min at 37±2 °C, 5±1 % CO₂, ≥95 % RH, protected from light.</p>	—
Treatment doses and application	<p>50 µl (83,3 µl/cm²)</p>	<p>50 mg (83,3 mg/cm²) using a calibrated tool (e.g. a levelled spoonful calibrated to hold 50 mg of sodium chloride).</p>	<p>30 µl Ca²⁺/Mg²⁺-free DPBS + 30 ± 2 mg (60 mg/cm²)</p> <p>For viscous, use a nylon mesh</p>
Exposure time and temperature	<p>30 min (± 2 min) in culture medium at 37±2 °C, 5±1 % CO₂, ≥95 % RH</p>	<p>6 hours (± 0.25 h) in culture medium at 37±2 °C, 5±1 % CO₂, ≥95 % RH</p>	<p>4 hours (± 0.1 h) in culture medium at 37±2 °C, 5±1 % CO₂, ≥95 % RH</p>

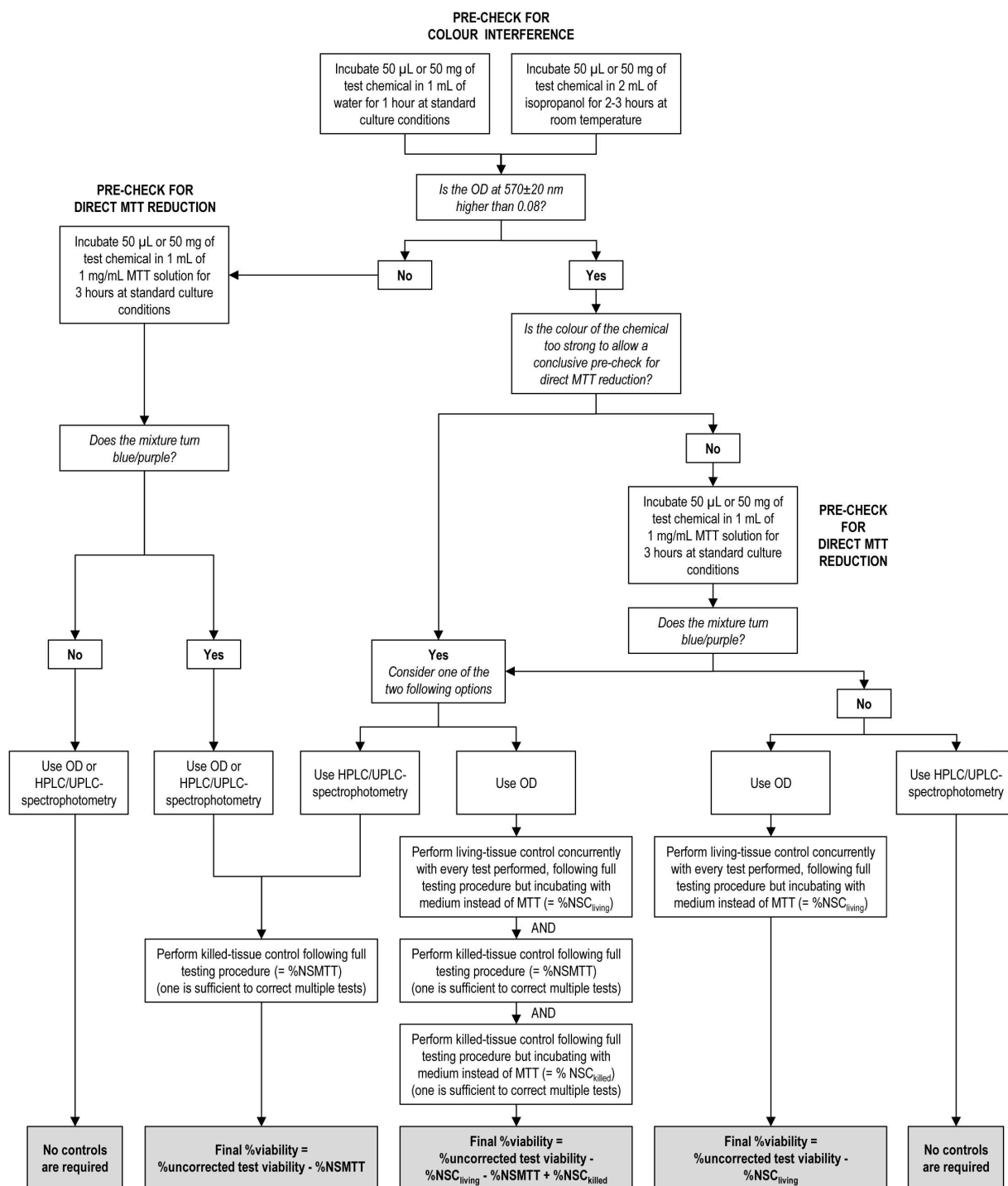
Test Components	EpiOcular™ EIT (VRM 1)		SkinEthic™ HCE EIT (VRM 2)	
Rinsing at room temperature	3 times in 100 ml of Ca ²⁺ /Mg ²⁺ -free DPBS	3 times in 100 ml of Ca ²⁺ /Mg ²⁺ -free DPBS	20 ml Ca ²⁺ /Mg ²⁺ -free DPBS	25 ml Ca ²⁺ /Mg ²⁺ -free DPBS
Post-exposure immersion	12 min (± 2 min) at RT in culture medium	25 min (± 2 min) at RT in culture medium	30 min (± 2 min) at 37 °C, 5 % CO ₂ , 95 % RH in culture medium	30 min (± 2 min) at RT in culture medium
Post-exposure incubation	120 min (± 15 min) in culture medium at 37±2 °C, 5±1 % CO ₂ , ≥95 % RH	18 h (± 0,25 h) in culture medium at 37±2 °C, 5±1 % CO ₂ , ≥95 % RH	none	18 h (± 0,5 h) in culture medium at 37±2 °C, 5±1 % CO ₂ , ≥95 % RH
Negative control	50 µl H ₂ O Tested concurrently	50 µl H ₂ O Tested concurrently	30 ± 2 µl Ca ²⁺ /Mg ²⁺ -free DPBS Tested concurrently	30 ± 2 µl Ca ²⁺ /Mg ²⁺ -free DPBS Tested concurrently
Positive control	50 µl Methyl acetate Tested concurrently	50 µl Methyl acetate Tested concurrently	30 ± 2 µl Methyl acetate Tested concurrently	30 ± 2 µl Methyl acetate Tested concurrently

Test Components	EpiOcular™ EIT (VRM 1)	Epithic™ HCE EIT (VRM 2)
MTT solution	300 µl 1 mg/ml	300 µl 1 mg/ml
MTT incubation time and temperature	180 min (± 15 min) at 37±2 °C, 5±1 % CO ₂ , ≥95 % RH	180 min (± 15 min) at 37±2 °C, 5±1 % CO ₂ , ≥95 % RH
Extraction solvent	2 ml isopropanol (extraction from top and bottom of insert by piercing the tissue)	1.5 ml isopropanol (extraction from bottom of insert)
Extraction time and temperature	2-3 h with shaking (~120 rpm) at RT or overnight at 4-10 °C	4 h with shaking (~120 rpm) at RT or at least overnight without shaking at 4-10 °C
OD reading	570 nm (550 - 590 nm) without reference filter	570 nm (540 - 600 nm) without reference filter
Tissue Quality Control	Treatment with 100 µl of 0,3 % (v/v) Triton X-100 12.2 min ≤ ET ₅₀ ≤ 37,5 min	30 min treatment with SDS (50 µl) 1,0 mg/ml ≤ IC ₅₀ ≤ 3,2 mg/ml

Test Components	EpiOcular™ EIT (VRM 1)	EpiOcular™ EIT (VRM 1)	SkinEthic™ HCE EIT (VRM 2)
Acceptance Criteria	<p>1. Mean OD of the tissue replicates treated with the negative control should be > 0,8 and < 2,5</p> <p>2. Mean viability of the tissue replicates exposed for 30 min with the positive control, expressed as % of the negative control, should be < 50 %</p> <p>3. The difference of viability between two tissue replicates should be less than 20 %.</p>	<p>1. Mean OD of the tissue replicates treated with the negative control should be > 0.8 and < 2.5</p> <p>2. Mean viability of the tissue replicates exposed for 6 hours with the positive control, expressed as % of the negative control, should be < 50 %</p> <p>3. The difference of viability between two tissue replicates should be less than 20 %.</p>	<p>1. Mean OD of the tissue replicates treated with the negative control should be > 1.0 and ≤ 2.5</p> <p>2. Mean viability of the tissue replicates exposed for 4 hours with the positive control, expressed as % of the negative control, should be ≤ 20 %</p> <p>3. The difference of viability between two tissue replicates should be less than 20 %.</p>

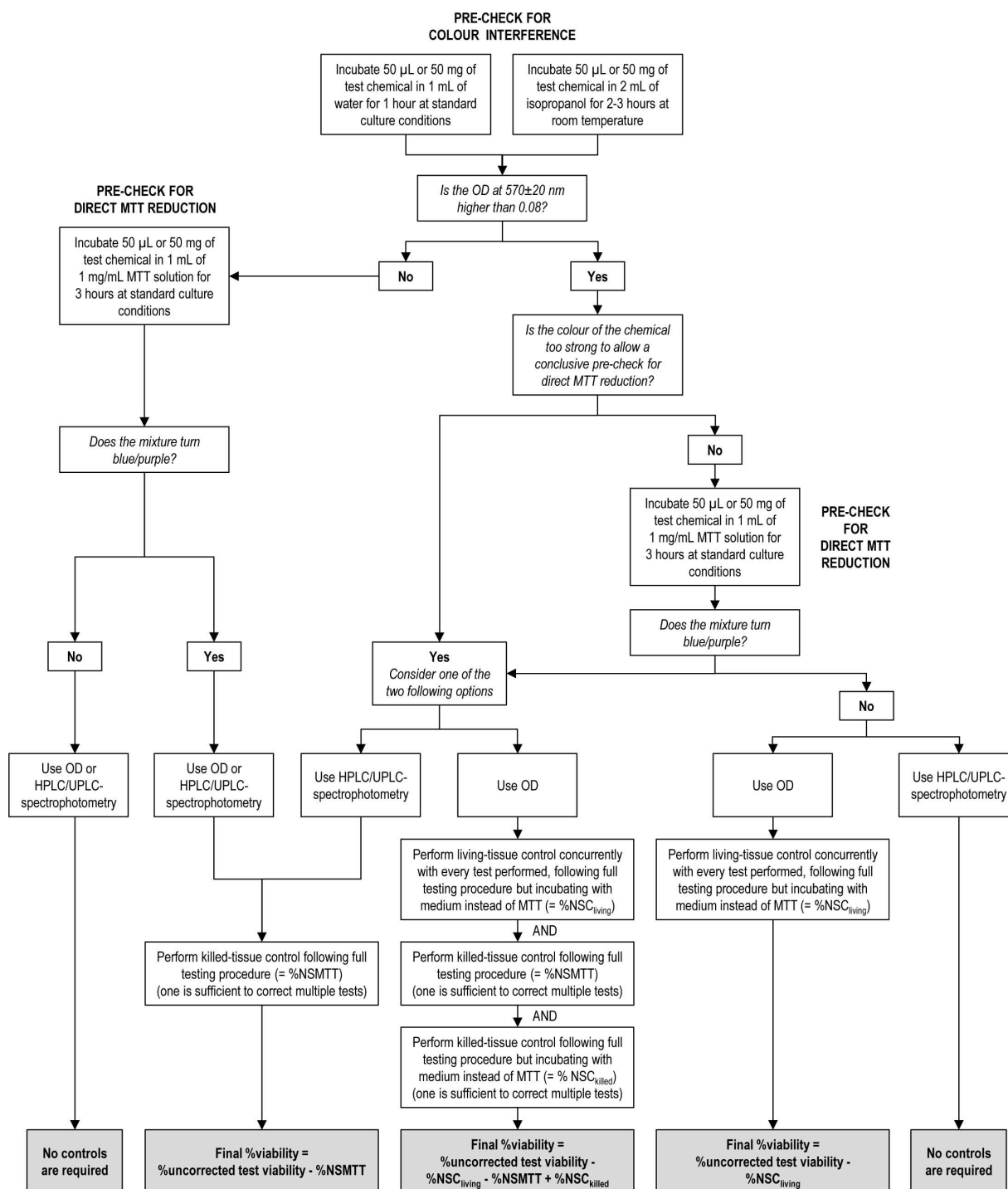
Appendix 3

ILLUSTRATIVE FLOWCHART PROVIDING GUIDANCE ON HOW TO IDENTIFY AND HANDLE DIRECT MTT-REDUCERS AND/OR COLOUR INTERFERING CHEMICALS, BASED ON THE VRM1 SOP



Appendix 4

ILLUSTRATIVE FLOWCHART PROVIDING GUIDANCE ON HOW TO IDENTIFY AND HANDLE DIRECT MTT-REDUCERS AND/OR COLOUR INTERFERING CHEMICALS, BASED ON THE VRM2 SOP



Appendix 5

KEY PARAMETERS AND ACCEPTANCE CRITERIA FOR QUALIFICATION OF AN HPLC/UPLC-SPECTROPHOTOMETRY SYSTEM FOR MEASUREMENT OF MTT FORMAZAN EXTRACTED FROM RHCE TISSUE CONSTRUCTS

Parameter	Protocol Derived from FDA Guidance (36)(38)	Acceptance Criteria
Selectivity	Analysis of isopropanol, living blank (isopropanol extract from living RhCE tissue constructs without any treatment), dead blank (isopropanol extract from killed RhCE tissue constructs without any treatment), and of a dye (e.g. methylene blue)	$\text{Area}_{\text{interference}} \leq 20 \% \text{ of } \text{Area}_{\text{LLOQ}}^{(1)}$
Precision	Quality Controls (i.e., MTT formazan at 1,6 µg/ml, 16 µg/ml and 160 µg/ml) in isopropanol (n=5)	CV ≤ 15 % or ≤ 20 % for the LLOQ
Accuracy	Quality Controls in isopropanol (n=5)	%Dev ≤ 15 % or ≤ 20 % for LLOQ
Matrix Effect	Quality Controls in living blank (n=5)	$85 \% \leq \% \text{Matrix Effect} \leq 115 \%$
Carryover	Analysis of isopropanol after an ULOQ ⁽²⁾ standard	$\text{Area}_{\text{interference}} \leq 20 \% \text{ of } \text{Area}_{\text{LLOQ}}$
Reproducibility (intra-day)	3 independent calibration curves (based on 6 consecutive 1/3 dilutions of MTT formazan in isopropanol starting at ULOQ, i.e., 200 µg/ml); Quality Controls in isopropanol (n=5)	Calibration Curves: %Dev ≤ 15 % or ≤ 20 % for LLOQ Quality Controls: %Dev ≤ 15 % and CV ≤ 15 %
Reproducibility (inter-day)	Day 1: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 2: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 3: 1 calibration curve and Quality Controls in isopropanol (n=3)	
Short Term Stability of MTT Formazan in RhCE Tissue Extract	Quality Controls in living blank (n=3) analysed the day of the preparation and after 24 hours of storage at room temperature	%Dev ≤ 15 %
Long Term Stability of MTT Formazan in RhCE Tissue Extract, if required	Quality Controls in living blank (n=3) analysed the day of the preparation and after several days of storage at -20 °C	%Dev ≤ 15 %

⁽¹⁾ LLOQ: Lower Limit of Quantification, defined to cover 1-2 % tissue viability, i.e., 0,8 µg/ml.

⁽²⁾ ULOQ: Upper Limit of Quantification, defined to be at least two times higher than the highest expected MTT formazan concentration in isopropanol extracts from negative controls (~70 µg/ml in the VRM), i.e., 200 µg/ml.

B.70 HUMAN RECOMBINANT ESTROGEN RECEPTOR (hrER) IN VITRO ASSAYS TO DETECT CHEMICALS WITH ER BINDING AFFINITY

GENERAL INTRODUCTION

OECD Performance-Based Test Guideline

1. This test method is equivalent to OECD test guideline (TG) 493 (2015). TG 493 is a performance-based test guideline (PBTG), describing the methodology for human recombinant *in vitro* assays to detect substances with estrogen receptor binding affinity (hrER binding assays). It comprises two mechanistically and functionally similar assays for the identification of estrogen receptor (i.e. ER α) binders and should facilitate the development of new similar or modified assays in accordance with the principles for validation set forth in the OECD Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (1). The fully validated reference test methods (Appendix 2 and Appendix 3) that provide the basis for this PBTG are:
 - The Freyberger-Wilson (FW) *In Vitro* Estrogen Receptor (ER) Binding Assay Using a Full Length Human Recombinant ER α (2), and
 - The Chemical Evaluation and Research Institute (CERI) *In Vitro* Estrogen Receptor Binding Assay Using a Human Recombinant Ligand Binding Domain Protein (2).

Performance standards (PS) (3) are available to facilitate the development and validation of similar test methods for the same hazard endpoint and allow for timely amendment of PBTG 493 so that new similar assays can be added to an updated PBTG. However, similar test assays will only be added after review and agreement by OECD that performance standards are met. The assays included in TG 493 can be used indiscriminately to address OECD member countries' requirements for test results on estrogen receptor binding while benefiting from the OECD Mutual Acceptance of Data.

Background and principles of the assays included in this test method

2. The OECD initiated a high-priority activity in 1998 to revise existing, and to develop new test guidelines for the screening and testing of potential endocrine disrupting chemicals. The OECD conceptual framework (CF) for testing and assessment of potential endocrine disrupting chemicals was revised in 2012. The original and revised CFs are included as Annexes in the Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption (4). The CF comprises five levels, each level corresponding to a different level of biological complexity. The ER binding assays described in this test method are level 2, which includes “*in vitro* assays providing data about selected endocrine mechanism(s)/pathway(s)”. This test method is for *in vitro* receptor binding assays designed to identify ligands for the human estrogen receptor alpha (ER α).
3. The relevance of the *in vitro* ER binding assay to biological functions has been clearly demonstrated. ER binding assays are designed to identify chemicals that have the potential to disrupt the estrogen hormone pathway, and have been used extensively during the past two decades to characterise ER tissue distribution as well as to identify ER agonists/antagonists. These assays reflect the ligand-receptor interaction which is the initial step of the estrogen signalling pathway and essential for reproduction function in all vertebrates.

4. The interaction of estrogens with ERs can affect transcription of estrogen-controlled genes and induce non-genomic effects, which can lead to the induction or inhibition of cellular processes, including those necessary for cell proliferation, normal foetal development, and reproductive function (5) (6) (7). Perturbation of normal estrogenic systems may have the potential to trigger adverse effects on normal development (ontogenesis), reproductive health and the integrity of the reproductive system. Inappropriate ER signalling can lead to effects such as increased risk of hormone dependent cancer, impaired fertility, and alterations in foetal growth and development (8).
5. *In vitro* binding assays are based on a direct interaction of a substance with a specific receptor ligand binding site that regulates the gene transcription. The key component of the human recombinant estrogen receptor alpha (hrER α) binding assay measures the ability of a radiolabelled ligand ($[^3\text{H}]17\beta$ -estradiol) to bind with the ER in the presence of increasing concentrations of a test chemical (i.e. competitor). Test chemicals that possess a high affinity for the ER compete with the radiolabelled ligand at a lower concentration as compared with those chemicals with lower affinity for the receptor. This assay consists of two major components: a saturation binding experiment to characterise receptor-ligand interaction parameters and document ER specificity, followed by a competitive binding experiment that characterises the competition between a test chemical and a radiolabelled ligand for binding to the ER.
6. Validation studies of the CER1 and the FW binding assays have demonstrated their relevance and reliability for their intended purpose (2).
7. Definitions and abbreviations used in this test method are described in Appendix 1.

Scope and limitations related to the receptor binding assays

8. These assays are being proposed for screening and prioritisation purposes, but can also provide information for a molecular initiation event (MIE) that can be used in a weight of evidence approach. They address chemical binding to the ER α ligand binding domain in an *in vitro* system. Thus, results should not be directly extrapolated to the complex signalling and regulation of the intact endocrine system *in vivo*.
9. Binding of the natural ligand, 17 β -estradiol, is the initial step of a series of molecular events that activates the transcription of target genes and ultimately, culminates with a physiological change (9). Thus binding to the ER α ligand binding domain is considered one of the key mechanisms of ER mediated endocrine disruption (ED), although there are other mechanisms through which ED can occur, including (i) interactions with sites of ER α other than the ligand binding pocket, (ii) interactions with other receptors relevant for estrogen signalling, ER β and G-protein coupled estrogen receptor, other receptors and enzymatic systems within the endocrine system, (iii) hormone synthesis, (iv) metabolic activation and/or inactivation of hormones, (v) distribution of hormones to target tissues, and (vi) clearance of hormones from the body. None of the assays under this test method address these modes of action.

10. This test method addresses the ability of substances to bind to human ER α and does not distinguish between ER α agonists or antagonists. These assays does not address either further downstream events such as gene transcription or physiological changes. Considering that only single mono-constituent substances were used during the validation, the applicability to test mixtures has not been addressed. The assays are nevertheless theoretically applicable to the testing of multi-constituent substances and mixtures. Before use of the test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture.
11. The cell free receptor systems have no intrinsic metabolic capability and they were not validated in combination with metabolic enzyme systems. However, it might be possible to incorporate metabolic activity in a study design but this would require further validation efforts.
12. Chemicals that may denature the protein (i.e. receptor protein), such as surfactant or chemicals that can change the pH of the assay buffer, may not be tested or may only be tested at concentrations devoid of such interactions. Otherwise, the concentration range that can be tested in the assays for a test chemical is limited by its solubility in the assay buffer.
13. For informational purposes, Table 1 provides the test results for the 24 substances that were tested in both of the fully validated assays described in this test method. Of these substances, 17 are classified as ER binders and 6 as non-binders based upon published reports, including *in vitro* assays for ER transcriptional activation and/or the uterotrophic assay (9) (10) (11) (12) (13) (14) (15). In reference to the data summarised in Table 1, there was almost 100 % agreement between the two assays on the classifications of all the substances up to 10^{-4} M, and each substance was correctly classified as an ER binder or non-binder. Supplementary information on this group of substances as well as additional substances tested in the ER binding assays during the validation studies is provided in the Performance Standards for the hrER binding assay (3), Appendix 2 (Tables 1, 2 and 3).

Table 1

Classification of substances as ER binders or non-binders when tested in the FW and CERI hrER Binding Assays with comparison with expected response

	Substance Name	CAS RN	Expected Response	FW Assay		CERI Assay		MESH Chemical Class	Product Class
				Concentration Range (M)	Classification	Concentration Range (M)	Classification		
1	17β-Estradiol	50-28-2	Binder	1x10 ⁻¹¹ – 1x10 ⁻⁶	Binder	1x10 ⁻¹¹ – 1x10 ⁻⁶	Binder	Steroid	Pharmaceutical, Veterinary Agent
2	Norethynodrel	68-23-5	Binder	3x10 ⁻⁹ – 30x10 ⁻⁴	Binder	3x10 ⁻⁹ – 30x10 ⁻⁴	Binder	Steroid	Pharmaceutical, Veterinary Agent
3	Norethindrone	68-22-4	Binder	3x10 ⁻⁹ – 30x10 ⁻⁴	Binder	3x10 ⁻⁹ – 30x10 ⁻⁴	Binder	Steroid	Pharmaceutical, Veterinary Agent
4	Di-n-butyl phthalate	84-74-2	Non-binder (*)	1x10 ⁻¹⁰ – 1x10 ⁻⁴	Non-Binder (**) (†)	1x10 ⁻¹⁰ – 1x10 ⁻⁴	Non-Binder (**) (†)	Hydrocarbon (cyclic), Ester	Plasticiser, Chemical Intermediate
5	DES	56-53-1	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Hydrocarbon (Cyclic), Phenol	Pharmaceutical, Veterinary Agent
6	17α-ethynyles-tradiol	57-63-6	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Steroid	Pharmaceutical, Veterinary Agent
7	Meso-Hexestrol	84-16-2	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Hydrocarbon (Cyclic), Phenol	Pharmaceutical, Veterinary Agent
8	Genistein	446-72-0	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Hydrocarbon (heterocyclic), Flavonoid	Natural Product
9	Equol	531-95-3	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Phytoestrogen Metabolite	Natural Product
10	Butyl paraben (n-butyl-4-hydroxybenzoate)	94-26-8	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Paraben	Preservative

	Substance Name	CAS RN	Expected Response	FW Assay		CERI Assay		MESH Chemical Class	Product Class
				Concentration Range (M)	Classification	Concentration Range (M)	Classification		
11	Nonylphenol (mixture)	84852-15-3	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Alkylphenol	Intermediate Compound
12	<i>o,p'</i> -DDT	789-02-6	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Organochlorine	Insecticide
13	Corticosterone	50-22-6	Non-binder (*)	1x10 ⁻¹⁰ – 1x10 ⁻⁴	Non-binder	1x10 ⁻¹⁰ – 1x10 ⁻⁴	Non-Binder	Steroid	Natural Product
14	Zearalenone	17924-92-4	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Hydrocarbon (heterocyclic), Lactone	Natural Product
15	Tamoxifen	10540-29-1	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Hydrocarbon, (Cyclic)	Pharmaceutical, Veterinary Agent
16	5 α -dihydrotestosterone	521-18-6	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Steroid, Nonphenolic	Natural Product
17	Bisphenol A	80-05-7	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Phenol	Chemical Intermediate
18	4- <i>n</i> -heptylphenol	1987-50-4	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Equivocal (4)	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Alkylphenol	Intermediate
19	Kepone (Chlordane)	143-50-0	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Hydrocarbon, (Halogenated)	Pesticide
20	Benz(a)anthracene	56-55-3	Non-Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Non-Binder (6)	1x10 ⁻¹⁰ – 1x10 ⁻³	Non-Binder (6)	Aromatic Hydrocarbon	Intermediate
21	Enterolactone	78473-71-9	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Binder	Phytoestrogen	Natural Product
22	Progesterone	57-83-0	Non-binder (*)	1x10 ⁻¹⁰ – 1x10 ⁻⁴	Non-Binder	1x10 ⁻¹⁰ – 1x10 ⁻⁴	Non-Binder	Steroid	Natural Product

Substance Name	CAS RN	Expected Response	FW Assay		CERI Assay		MESH Chemical Class	Product Class
			Concentration Range (M)	Classification	Concentration Range (M)	Classification		
23 Octyltriethoxy-silane	2943-75-1	Non-binder	1×10^{-10} – 1×10^{-3}	Non-Binder	1×10^{-10} – 1×10^{-3}	Non-Binder	Silane	Surface Modifier
24 Atrazine	1912-24-9	Non-binder (*)	1×10^{-10} – 1×10^{-4}	Non-Binder	1×10^{-10} – 1×10^{-4}	Non-Binder	Heterocyclic compound	Herbicide

(*) Limit of solubility < 1×10^{-4} M.

(**) The use and classification of di-n-butyl phthalate (DBP) as a non-binder was based on testing up to 10^{-4} M because the substance had been observed to be insoluble at 10^{-3} M (e.g. turbidity) in some laboratories during the pre-validation studies.

(†) During the validation study, di-n-butyl phthalate (DBP) was tested as a coded test substance at concentrations up to 10^{-3} M. Under these conditions, some laboratories observed either a decrease in radioligand binding at the highest concentration (10^{-3} M) and/or an ambiguous curve fit. For these runs, DBP was classified as 'equivocal' or 'binder' in 3/5 laboratories using the CERI assay and 5/6 laboratories using the FW assay (see Reference (2), Sections IV.B.3.a,b and VI.A).

(‡) Classification was not consistent with expected classification. Classification of 4-n-heptylphenol as 'equivocal' or 'non-binder' by 3/5 labs resulted in an average classification of equivocal. Closer inspection revealed that this was due to chemical solubility limitations that prevented the production of a full binding curve.

(§) During the validation study, benz(a)anthracene was reclassified as a non-binder (i.e. negative) based on published literature demonstrating that the *in vitro* estrogenic activity reported for this substance is primarily dependent upon its metabolic activation. Enzymatic metabolic activation of the substance would not be anticipated in the cell free hrER binding assays as used in this inter-laboratory validation study. Thus, the correct classification for this substance is a 'non-binder' when used under the experimental conditions for the FW and CERI assays.

hrER BINDING ASSAY COMPONENTS

Essential Assay Components

14. This test method applies to assays using an ER receptor and a suitably strong ligand to the receptor that can be used as a marker/tracer for the assay and can be displaced with increasing concentrations of a test chemical. Binding assays contain the following two major components: 1) saturation binding and 2) competitive binding. The saturation binding assay is used to confirm the specificity and activity of the receptor preparations, while the competitive binding experiment is used to evaluate the ability of a test chemical to bind to hrER.

Controls

15. The basis for the proposed concurrent reference estrogen and controls should be described. Concurrent controls (solvent (vehicle), positive (ER binder; strong and weak affinity), negative (non-binder)), as appropriate, serve as an indication that the assay is operative under the test conditions and provide a basis for experiment-to-experiment comparisons; they are usually part of the acceptability criteria for a given experiment (1). Full concentration curves for the reference estrogen and controls (i.e. weak binder and non-binder) should be used in one plate during each run. All other plates should contain: 1) a high- (approximately full displacement of radiolabelled ligand) and medium- (approximately the IC_{50}) concentration each of E2 and weak binder in triplicate; 2) solvent control and non-specific binding, each in triplicate.

Standard Quality Control Procedures

16. Standard quality control procedures should be performed as described for each assay to ensure active receptors, the correct chemical concentrations, tolerance bounds remain stable through multiple replications, and retain the ability to provide the expected ER-binding responses over time.

Demonstration of Laboratory Proficiency

17. Prior to testing unknown chemicals with any of the assays under this test method, each laboratory should demonstrate proficiency in using the assay by performing saturation assays to confirm specificity and activity of the ER preparation, and competitive binding assays with the reference estrogen and controls (weak binder and non-binder). A historical database with results for the reference estrogen and controls generated from 3-5 independent experiments conducted on different days should be established by the laboratory. These experiments will be the foundation for the reference estrogen and historical controls for the laboratory and will be used as a partial assessment of assay acceptability for future runs.
18. The responsiveness of the test system will also be confirmed by testing the proficiency substances listed in Table 2. The list of proficiency substances is a subset of the reference substances provided in the Performance Standards for the ER binding assays (3). These substances are commercially available, represent the classes of chemicals commonly associated with ER binding activity, exhibit a suitable range of potency expected for ER binders (i.e. strong to weak) and non-binders (i.e. negatives). For each proficiency substance, concentrations tested should cover the range provided in Table 2. At least three experiments should be performed for each substance and results should be in concordance with expected chemical activity. Each experiment should be conducted independently (i.e. with fresh dilutions of receptor, chemicals, and reagent), with three replicates for each concentration. Proficiency is demonstrated by correct classification (positive/negative) of each proficiency substance. Proficiency testing should be performed by each technician when learning the assays.

Table 2

List of controls and proficiency substances for the hrER competitive binding assays ⁽¹⁾

No	Substance Name	CAS RN ⁽²⁾	Expected Response ⁽³⁾ ⁽⁴⁾	Test concentration range (M)	MeSH chemical class ⁽⁵⁾	Product class ⁽⁶⁾
Controls (Reference estrogen, weak binder, non-binder)						
1	17 β -estradiol	50-28-2	Binder	1x10 ⁻¹¹ – 1x10 ⁻⁶	Steroid	Pharmaceutical, Veterinary agent
2	Norethynodrel (or) Norethindrone	68-23-5 (or)68-22-4	Binder	3x10 ⁻⁹ – 30x10 ⁻⁶	Steroid	Pharmaceutical, Veterinary agent
3	Octyltriethoxysilane	2943-75-1	Non-binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Silane	Surface modifier
Proficiency substances ⁽⁶⁾						
4	Diethylstilbestrol	56-53-1	Binder	1x10 ⁻¹¹ – 1x10 ⁻⁶	Hydrocarbon (cyclic), Phenol	Pharmaceutical, Veterinary agent
5	17 α -ethynylestradiol	57-63-6	Binder	1x10 ⁻¹¹ – 1x10 ⁻⁶	Steroid	Pharmaceutical, Veterinary agent
6	meso-Hexestrol	84-16-2	Binder	1x10 ⁻¹¹ – 1x10 ⁻⁶	Hydrocarbon (cyclic), Phenol	Pharmaceutical, Veterinary agent
7	Tamoxifen	10540-29-1	Binder	1x10 ⁻¹¹ – 1x10 ⁻⁶	Hydrocarbon (cyclic)	Pharmaceutical, Veterinary agent
8	Genistein	446-72-0	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Heterocyclic compound, Flavonoid,	Natural product

No	Substance Name	CAS RN ⁽²⁾	Expected Response ⁽³⁾ ⁽⁴⁾	Test concentration range (M)	MeSH chemical class ⁽⁵⁾	Product class ⁽⁶⁾
9	Bisphenol A	80-05-7	Binder	1x10 ⁻¹⁰ – 1x10 ⁻³	Phenol	Chemical intermediate
10	Zearalnonone	17924-92-4	Binder	1x10 ⁻¹¹ – 1x10 ⁻³	Heterocyclic compound, Lactone	Natural Product
11	Butyl paraben	94-26-8	Binder	1x10 ⁻¹¹ – 1x10 ⁻³	Carboxylic acid, Phenol	Preservative
12	Atrazine	1912-24-9	Non-binder	1x10 ⁻¹¹ – 1x10 ⁻⁶	Heterocyclic compound	Herbicide
13	Di-n-butylphthalate (DBP) ⁽⁷⁾	84-74-2	Non-binder ⁽⁸⁾	1x10 ⁻¹⁰ – 1x10 ⁻⁴	Hydrocarbon (cyclic), Ester	Plasticiser, Chemical intermediate
14	Corticosterone	50-22-6	Non-binder	1x10 ⁻¹¹ – 1x10 ⁻⁴	Steroid	Natural product

⁽¹⁾ If a proficiency substance is no longer commercially available, a substance with the same ER binding classification, comparable potency, and chemical class can be used.

⁽²⁾ Abbreviations: CAS RN = Chemical Abstracts Service Registry Number.

⁽³⁾ Classification as an ER/Binder or Non-binder during the validation study for the CER1 and FW hrER Binding Assays (2).

⁽⁴⁾ ER binding activity was based upon the ICCVAM Background Review Documents (BRD) for ER Binding and TA assays (9) as well as empirical data and other information obtained from referenced studies published and reviewed (10) (11) (12) (13) (14) (15).

⁽⁵⁾ Substances were assigned into one or more chemical classes using the U.S. National Library of Medicine's Medical Subject Headings (MeSH), an internationally recognised standardised classification scheme (available at: <http://www.nlm.nih.gov/mesh>).

⁽⁶⁾ Substances were assigned into one or more product classes using the U.S. National Library of Medicine's Hazardous Substances Database (available at: <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDDB>)

⁽⁷⁾ DBP can be used as an alternate control non-binder tested with maximum concentration of 10⁻⁴ M. The use and classification of di-n-butyl phthalate (DBP) as a non-binder has been based on testing up to 10⁻⁴ M because the substance had been observed to be insoluble at 10⁻³ M (e.g. turbidity) in some laboratories during the pre-validation studies.

Solubility Testing and Concentration Range Finding for Test Chemicals

19. A preliminary test should be conducted to determine the limit of solubility for each test chemical and to identify the appropriate concentration range to use when conducting the test. The limit of solubility of each test chemical is to be initially determined in the solvent and further confirmed under assay conditions. The final concentration tested in the assay should not exceed 1 mM. Range finder testing consists of a solvent control along with eight, log serial dilutions, starting at the maximum acceptable concentration (e.g. 1 mM or lower, based upon the limit of solubility), and the presence of cloudiness or precipitate noted. Concentrations in the second and third experiments should be adjusted as appropriate to better characterise the concentration-response curve.

Test Run Acceptability Criteria

20. Acceptance or rejection of a test run is based on the evaluation of results obtained for the reference estrogen and control used for each experiment. First, for plate 1, the full concentration curves for the reference controls from each experiment should meet the measures of performance with curve-fit parameters (e.g. IC_{50} and Hillslope) based upon the results reported for the respective protocols for the CERI and FW assays (Appendix 2 and 3), and the historical control data from the laboratory conducting the test. All controls (reference estrogen, weak binder, and non-binder) should be correctly classified for each experiment. Secondly, the controls on all subsequent plates need to be assessed for consistency with plate 1. A sufficient range of concentrations of the test chemical should be used to clearly define the top of the competitive binding curve. Variability among replicates at each concentration of the test chemical as well as among the three independent runs should be reasonable and scientifically defensible. The ability to consistently conduct the assay should be demonstrated by the development and maintenance of a historical database for the reference estrogen and controls. Standard deviations (SD) or coefficients of variation (CV) for the means of reference estrogen and control weak binder curves fitting parameters from multiple experiments may be used as a measure of within-laboratory reproducibility. Professional judgment should be applied when reviewing the plate control results from each run as well as for each test chemical.

In addition, the following principles regarding acceptability criteria should be met:

- Data should be sufficient for a quantitative assessment of ER binding
- The concentrations tested should remain within the solubility range of the test chemical.

Analysis of data

21. The defined data analysis procedure for saturation and competitive binding data should adhere to the key principles for characterising receptor-ligand interactions. Typically, saturation binding data are analysed using a non-linear regression model that accounts for total and non-specific binding. A correction for ligand depletion (e.g. Swillens, 1995 (19)) may be needed when determining B_{max} and K_d . Data from competitive binding assays are typically transformed (e.g. percent specific binding and concentration of test chemical (log M)). Estimates of $\log(IC_{50})$ for each test chemical should be determined using an appropriate nonlinear curve fitting software to fit a four parameter Hill equation. Following an initial analysis, the curve fit parameters and a visual review of how well the binding data fit the generated competitive binding curve should be conducted. In some cases, additional analysis may be needed to obtain the best curve fit (e.g. constraining top and/or bottom of curve, use of 10 % rule, see Appendix 4 and Reference 2 (Section III.A.2)).
22. Meeting the acceptability criteria (paragraph 20) indicates the assay system is operating properly, but it does not ensure that any particular test will produce accurate data. Replicating the correct results of the first test is the best indication that accurate data were produced.

General Data Interpretation Criteria

23. There is currently no universally agreed method for interpreting ER binding data. However, both qualitative (e.g. binder/non-binder) and/or quantitative (e.g. log IC₅₀, Relative Binding Affinity (RBA), etc.) assessments of hrER-mediated activity should be based on empirical data and sound scientific judgment.

Test Report

24. The test report should include the following information:

Assay:

- assay used;

Control/Reference/Test chemical

- source, lot number, limit date for use, if available
- stability of the test chemical itself, if known;
- solubility and stability of the test chemical in solvent, if known.
- measurement of pH, osmolality and precipitate in the culture medium to which the test chemical was added, as appropriate.

Mono-constituent substance:

- physical appearance, water solubility, and additional relevant physicochemical properties;
- chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc.

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents.

Solvent/Vehicle:

- characterisation (nature, supplier and lot);
- justification for choice of solvent/vehicle;
- solubility and stability of the test chemical in solvent/vehicle, if known;

Receptors:

- source of receptors (supplier, catalog No, lot, species of receptor, active receptor concentration provided from supplier, certification from supplier)

- characterisation of receptors (including saturation binding results): K_d , B_{max} ,
- storage of receptors
- radiolabelled ligand:
- supplier, catalog No., lot, specific activity

Test conditions:

- solubility limitations under assay conditions;
- composition of binding buffer;
- concentration of receptor;
- concentration of tracer (i.e. radiolabelled ligand);
- concentrations of test chemical;
- percent vehicle in final assay;
- incubation temperature and time;
- method of bound/free separation;
- positive and negative controls/reference substances;
- criteria for considering tests as positive, negative or equivocal;

Acceptability check:

- actual IC_{50} and Hill slope values for concurrent positive controls/reference substances;

Results:

- raw and bound/free data;
- denaturing confirmation check, if appropriate;
- if it exists, the lowest effective concentration (LEC);
- RBA and/or IC_{50} values, as appropriate;
- concentration-response relationship, where possible;
- statistical analyses, if any, together with a measure of error and confidence (e.g. SEM, SD, CV or 95 % CI) and a description of how these values were obtained;

Discussion of the results:

— application of 10 % rule

Conclusion

LITERATURE

- (1) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Environmental, Health and Safety Publications, Series on Testing and Assessment (No 34), Organisation for Economic Cooperation and Development, Paris.
- (2) OECD (2015). Integrated Summary Report: Validation of Two Binding Assays Using Human Recombinant Estrogen Receptor Alpha (hrER α), Health and Safety Publications, Series on Testing and Assessment (No 226), Organisation for Economic Cooperation and Development, Paris.
- (3) OECD (2015). Performance Standards for Binding Assays Using Human Recombinant Estrogen Receptor Alpha (hrER α), Health and Safety Publications, Series on Testing and Assessment (No 222), Organisation for Economic Cooperation and Development, Paris.
- (4) OECD (2012). Guidance Document on Standardized Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Environmental, Health and Safety Publications, Series on Testing and Assessment (No 150), Organisation for Economic Cooperation and Development, Paris.
- (5) Cavailles V. (2002). Estrogens and Receptors: an Evolving Concept, *Climacteric*, 5 Suppl 2: p.20-6.
- (6) Welboren W.J., *et al.* (2009). Genomic Actions of Estrogen Receptor Alpha: What are the Targets and How are they Regulated? *Endocr. Relat. Cancer.*, 16(4): p. 1073-89.
- (7) Younes M. and Honma N. (2011). Estrogen Receptor Beta, *Arch. Pathol. Lab. Med.*, 135(1): p. 63-6.
- (8) Diamanti-Kandarakis *et al.* (2009). Endocrine-Disrupting Chemicals: an Endocrine Society Sci. Statement, *Endo Rev* 30(4):293-342.
- (9) ICCVAM (2002). Background Review Document. Current Status of Test Methods for Detecting Endocrine Disruptors: *In Vitro* Estrogen Receptor Binding Assays. (NIH Publication No 03-4504). National Institute of Environmental Health Sciences, Research Triangle Park, NC.
- (10) ICCVAM (2003). ICCVAM Evaluation of *In Vitro* Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays.
- (11) ICCVAM (2006). ICCVAM Evaluation of *In Vitro* Test Methods for Detecting Potential Endocrine Disruptors: Estrogen Receptor and Androgen Receptor Binding and Transcriptional Activation Assays.
- (12) Akahori Y. *et al.* (2008). Relationship Between the Results of *In Vitro* Receptor Binding Assay to Human Estrogen Receptor Alpha and *In Vivo* Uterotrophic Assay: Comparative Study with 65 Selected Chemicals, *Toxicol. In Vitro*, 22(1): 225-231.

- (13) OECD (2007). Additional Data Supporting the Test Guideline on the Uterotrophic Bioassay in Rodents, Environment, Health and Safety Publications, Series on Testing and Assessment (No 67), Organisation for Economic Cooperation and Development, Paris.
- (14) Takeyoshi, M. (2006). Draft Report of Pre-validation and Inter-laboratory Validation For Stably Transfected Transcriptional Activation (TA) Assay to Detect Estrogenic Activity - The Human Estrogen Receptor Alpha Mediated Reporter Gene Assay Using hER-HeLa-9903 Cell Line, Chemicals Evaluation and Research Institute (CERI): Japan. p. 1-188.
- (15) Yamasaki, K; Noda, S; Imatanaka, N; Yakabe, Y. (2004). Comparative Study of the Uterotrophic Potency of 14 Chemicals in a Uterotrophic Assay and their Receptor-Binding Affinity, *Toxicol. Letters*, 146: 111-120.
- (16) Kummer V; Maskova, J; Zraly, Z; Neca, J; Simeckova, P; Vondracek, J; Machala, M. (2008). Estrogenic Activity of Environmental Polycyclic Aromatic Hydrocarbons in Uterus of Immature Wistar Rats. *Toxicol. Letters*, 180: 213-221.
- (17) Gozgit, JM; Nestor, KM; Fasco, MJ; Pentecost, BT; Arcaro, KF. (2004). Differential Action of Polycyclic Aromatic Hydrocarbons on Endogenous Estrogen-Responsive Genes and on a Transfected Estrogen-Responsive Reporter in MCF-7 Cells. *Toxicol. and Applied Pharmacol.*, 196: 58-67.
- (18) Santodonato, J. (1997). Review of the Estrogenic and Antiestrogenic Activity of Polycyclic Aromatic Hydrocarbons: Relationship to Carcinogenicity. *Chemosphere*, 34: 835-848.
- (19) Swillens S (1995). Interpretation of Binding Curves Obtained with High Receptor Concentrations: Practical Aid for Computer Analysis, *Mol Pharmacol* 47(6):1197-1203.

Appendix 1

DEFINITIONS AND ABBREVIATIONS

10 % Rule: Option to exclude from the analyses data points where the mean of the replicates for the percent [^3H]17 β -estradiol specific bound is 10 % or more above that observed for the mean value at a lower concentration (see Appendix 4).

Acceptability criteria: Minimum standards for the performance of experimental controls and reference standards. All acceptability criteria should be met for an experiment to be considered valid.

Accuracy (concordance): The closeness of agreement between assay results and an accepted reference values. It is a measure of assay performance and one aspect of relevance. The term is often used interchangeably with “concordance” to mean the proportion of correct outcomes of an assay (1).

CF: The OECD Conceptual Framework for the Testing and Evaluation of Endocrine Disrupters.

Chemical: A substance or a mixture.

CV: Coefficient of variation

E2: 17 β -estradiol

ED: Endocrine disruption

hER α : Human estrogen receptor alpha

ER: Estrogen receptor

Estrogenic activity: The capability of a chemical to mimic 17 β -estradiol in its ability to bind estrogen receptors. Binding to the hER α can be detected with this test method.

IC50: The half maximal effective concentration of an inhibitory test chemical.

ICCVAM: The Interagency Coordinating Committee on the Validation of Alternative Methods.

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Interlaboratory reproducibility is determined during the prevalidation and validation processes, and indicates the extent to which an assay can be successfully transferred between laboratories, also referred to as between-laboratory reproducibility (1).

Intra-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Also referred to as “within-laboratory reproducibility” (1).

LEC: Lowest effective concentration is the lowest concentration of test chemical that produces a response (i.e. the lowest test chemical concentration at which the fold induction is statistically different from the concurrent vehicle control).

Me-too test: A colloquial expression for an assay that is structurally and functionally similar to a validated and accepted reference test method. Interchangeably used with similar test method

PBTG: Performance-Based Test Guideline

Performance standards: Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed assay that is mechanistically and functionally similar. Included are (1) essential assay components; (2) a minimum list of reference chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (3) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed assay should demonstrate when evaluated using the minimum list of reference chemicals (1).

Proficiency substances: A subset of the Reference substances included in the Performance Standards that can be used by laboratories to demonstrate technical competence with a standardised assay. Selection criteria for these substances typically include that they represent the range of responses, are commercially available, and have high quality reference data available.

Proficiency: The demonstrated ability to properly conduct an assay prior to testing unknown substances.

Reference estrogen: 17 β -estradiol (E2, CAS 50-28-2).

Reference test methods: The assays upon which PBTG 493 is based.

RBA: Relative Binding Affinity. The RBA of a substance is calculated as a percent of the log (IC₅₀) for the substance relative to the log (IC₅₀) for 17 β -estradiol

Relevance: Description of relationship of an assay to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the assay correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of an assay (1).

Reliability: Measure of the extent that an assay can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility.

SD: Standard deviation.

Test chemical: Any substance or mixture tested using this test method.

Validated test method: An assay for which validation studies have been completed to determine the relevance (including accuracy) and reliability for a specific purpose. It is important to note that a validated test method may not have sufficient performance in terms of accuracy and reliability to be found acceptable for the proposed purpose (1).

Validation: The process by which the reliability and relevance of a particular approach, method, assay, process or assessment is established for a defined purpose (1).

Appendix 2

THE FREYBERGER-WILSON IN VITRO ESTROGEN RECEPTOR (ER α) SATURATION AND COMPETITIVE BINDING ASSAYS USING FULL LENGTH RECOMBINANT ER α

INITIAL CONSIDERATIONS AND LIMITATIONS (SEE ALSO GENERAL INTRODUCTION)

1. This *in vitro* Estrogen Receptor (ER α) saturation and competitive binding assay uses full length human receptor ER α (hrER α) that is produced in and isolated from baculovirus-infected insect cells. The protocol, developed by Freyberger and Wilson, underwent an international multi-laboratory validation study (2) which has demonstrated its relevance and reliability for the intended purpose of the assay.
2. This assay is a screening procedure for identifying substances that can bind to the full length hrER α . It is used to determine the ability of a test chemical to compete with 17 β -estradiol for binding to hrER α . Quantitative assay results may include the IC₅₀ (a measure of the concentration of test chemical needed to displace half of the [³H]-17 β -estradiol from the hrER α) and the relative binding affinities of test chemicals for the hrER α compared to 17 β -estradiol. For chemical screening purposes, acceptable qualitative assay results may include classifications of test chemicals as either hrER α binders, non-binders, or equivocal based upon criteria described for the binding curves.
3. The assay uses a radioactive ligand that requires a radioactive materials license for the laboratory. All procedures with radioisotopes and hazardous chemicals should follow the regulations and procedures as described by national legislation.
4. The “GENERAL INTRODUCTION” and “hrER BINDING ASSAY COMPONENTS” should be read before using this assay for regulatory purposes. Definitions and abbreviations used in this TG are described in Appendix 1.

PRINCIPLES OF THE ASSAY (SEE ALSO GENERAL INTRODUCTION)

5. The hrER α binding assay measures the ability of a radiolabelled ligand ([³H]17 β -estradiol) to bind with the ER in the presence of increasing concentrations of a test chemical (i.e. competitor). Test chemicals that possess a high affinity for the ER compete with the radiolabelled ligand at a lower concentration as compared with those chemicals with lower affinity for the receptor.
6. This assay consists of two major components: a saturation binding experiment to characterise receptor-ligand interaction parameters, followed by a competitive binding experiment that characterises the competition between a test chemical and a radiolabelled ligand for binding to the ER.
7. The purpose of the saturation binding experiment is to characterise a particular batch of receptors for binding affinity and number in preparation for the competitive binding experiment. The saturation binding experiment measures, under equilibrium conditions, the affinity of a fixed concentration of the estrogen receptor for its natural ligand (represented by the dissociation constant, K_d), and the concentration of active receptor sites (B_{max}).
8. The competitive binding experiment measures the affinity of a substance to compete with [³H]17 β -estradiol for binding to the ER. The affinity is quantified by the concentration of test chemical that, at equilibrium, inhibits 50 % of the specific binding of the [³H]17 β -estradiol (termed the “inhibitory concentration 50 %” or IC₅₀). This can also be evaluated using the relative binding affinity (RBA, relative to the IC₅₀ of estradiol measured separately in the same run). The competitive binding experiment measures the binding of [³H]17 β -estradiol at a fixed concentration in the presence of a wide range (eight orders of magnitude) of test chemical concentrations. The data are then fit, where possible, to a form of the Hill equation (Hill, 1910) that describes the displacement of the radioligand by a one-site competitive binder. The extent of displacement of the radiolabelled estradiol at equilibrium is used to characterise the test chemical as a binder, non-binder, or generating an equivocal response.

PROCEDURE

Demonstration of Acceptable hrER α Protein Performance

9. Prior to routinely conducting the saturation and competitive binding assays, each new batch of hrER α should be shown to be performing correctly in the laboratory in which it will be used. A two-step process should be used to demonstrate performance. These steps are the following:
- Conduct a saturation [^3H]-17 β -estradiol binding assay to demonstrate hrER α specificity and saturation. Nonlinear regression analysis of these data (e.g. BioSoft; McPherson, 1985; Motulsky, 1995) and the subsequent Scatchard plot should document hrER α binding affinity of the [^3H]-17 β -estradiol (Kd) and the number of receptors (Bmax) for each batch of hrER α .
 - Conduct a competitive binding assay using the control substances (reference estrogen (17 β -estradiol)), a weak binder (e.g. norethynodrel or norethindrone), and a non-binder (octyltriethoxysilane, OTES). Each laboratory should establish an historical database to document the consistency of IC $_{50}$ and other relevant values for the reference estrogen and weak binder among experiments and different batches of hrER α . The parameters of the competitive binding curves for the control substances should be within the limits of the 95 % confidence interval (see Table 1) that were developed using data from laboratories that participated in the validation study for this assay (2).

Table 1

Performance criteria developed for the reference estrogen and weak binder, FW hrER Binding Assay

Substance	Parameter	Mean ^(a)	Standard Deviation (n)	95 % Confidence Intervals ^(b)	
				Lower Limit	Upper Limit
17 β -estradiol	Top (%)	100,44	10,84 (67)	97,8	103,1
	Bottom (%)	0,29	1,25 (67)	-0,01	0,60
	Hill Slope	-1,06	0,20 (67)	-1,11	-1,02
	LogIC $_{50}$ (M)	-8,92 ^(c)	0,18 (67)	-8,97	-8,88
Norethynodrel	Top (%)	99,42	8,90 (68)	97,27	101,60
	Bottom (%)	2,02	3,42 (68)	1,19	2,84
	Hill Slope	-1,01	0,38 (68)	-1,10	-0,92
	Log IC $_{50}$ (M)	-6,39	0,27 (68)	-6,46	-6,33
Norethindrone ^(c)	Top (%)	96,14	8,44 (27)	92,80	99,48
	Bottom (%)	2,38	5,02 (27)	0,40	4,37
	Hill Slope	-1,41	0,32 (27)	-1,53	-1,28
	LogIC $_{50}$ (M)	-5,73	0,27 (27)	-5,84	-5,62

^(a) Mean (n) \pm Standard Deviation (SD) were calculated using curve fit parameter estimates (4-parameter Hill Equation) for control runs conducted in four laboratories during the validation study (see Annex N of Reference 2).

^(b) The 95 % confidence intervals are provided as a guide for acceptability criteria.

^(c) Testing of norethindrone was optional for Subtask 4 during validation study (see Reference 2, see Subtask 4). Thus, the mean \pm SD (n) were calculated using curve fit estimates (4-parameter Hill equation) for control runs conducted in two laboratories. The range for the IC $_{50}$ will be dependent upon the Kd of the receptor preparation and concentration of radiolabelled ligand used within each laboratory. Appropriate adjustment for the range of the IC $_{50}$ based upon the conditions used to conduct the assay will be acceptable.

Demonstration of laboratory proficiency

10. See paragraphs 17 and 18 and Table 2 in “**hrER BINDING ASSAY COMPONENTS**” of this test method. Each assay (saturation and competitive binding) should consist of three independent runs (i.e. with fresh dilutions of receptor, chemicals, and reagents) on different days, and each run should contain three replicates.

Determination of Receptor (hrERa) Concentration

11. The concentration of active receptor varies slightly by batch and storage conditions. For this reason, the concentration of active receptor as received from the supplier should be determined. This will yield the appropriate concentration of active receptor at the time of the run.
12. Under conditions corresponding to competitive binding (i.e. 1 nM [³H]-estradiol), nominal concentrations of 0,25, 0,5, 0,75, and 1 nM receptor should be incubated in the absence (total binding) and presence (non-specific binding) of 1 μM unlabelled estradiol. Specific binding, calculated as the difference of total and non-specific binding, is plotted against the nominal receptor concentration. The concentration of receptor that gives specific binding values corresponding to 20 % of added radiolabel is related to the corresponding nominal receptor concentration, and this receptor concentration should be used for saturation and competitive binding experiments. Frequently, a final hrER concentration of 0,5 nM will comply with this condition.
13. If the 20 % criterion repeatedly cannot be met, the experimental set up should be checked for potential errors. Failure to achieve the 20 % criterion may indicate that there is very little active receptor in the recombinant batch, and the use of another receptor batch should then be considered.

Saturation assay

14. Eight increasing concentrations of [³H]17β-estradiol should be evaluated in triplicate, under the following three conditions (see Table 2):
 - In the absence of unlabelled 17β-estradiol and presence of ER. This is the determination of total binding by measure of the radioactivity in the wells that have only [³H]17β-estradiol.
 - In the presence of a 1 000- fold excess concentration of unlabelled 17β-estradiol over labelled 17β-estradiol and presence of ER. The intent of this condition is to saturate the active binding sites with unlabelled 17β-estradiol, and by measuring the radioactivity in the wells, determine the non-specific binding. Any remaining hot estradiol that can bind to the receptor is considered to be binding at a non-specific site as the cold estradiol should be at such a high concentration that it is bound to all of the available specific sites on the receptor.
 - In the absence of unlabelled 17β-estradiol and absence of ER (determination of total radioactivity)

Preparation of [³H]-17β-estradiol and unlabelled 17β-estradiol solutions

15. Dilutions of [³H]-17β-estradiol should be prepared by adding assay buffer to a 12 nM stock solution of [³H]-17β-estradiol to obtain concentrations initially ranging from 0,12nM to 12 nM. By adding 40 μl of these solutions to the respective assay wells of a 96-well microtiter plate (in a final volume of 160 μl), the final assay concentrations, ranging from 0,03 to 3,0 nM, will be obtained. Preparation of assay buffer, [³H]-17β-estradiol stock solution and dilutions and determination of the concentrations are described in depth in the FW protocol (2).
16. Dilutions of ethanolic 17β-estradiol solutions should be prepared by adding assay buffer to achieve eight increasing concentrations initially ranging from 0,06 μM to 6 μM. By adding 80 μl of these solutions to the respective assay wells of a 96-well microtiter plate (in a final volume of 160 μl), the final assay concentrations, ranging from 0,03μM to 3μM, will be obtained. The final concentration of unlabelled 17β-estradiol in the individual non-specific binding assay wells should be 1 000-fold of the labelled [³H]-17β- estradiol concentration. Preparation of unlabelled 17β-estradiol dilutions is described in depth in the FW protocol (2).

17. The nominal concentration of receptor that gives specific binding of $20 \pm 5\%$ should be used (see paragraphs 12-13). The hrER α solution should be prepared immediately prior to use.
18. The 96-well microtiter plates are prepared as illustrated in Table 2, with 3 replicates per concentration. Example of plate concentration and volume assignment of [3 H]-17 β -estradiol, unlabelled 17 β -estradiol, buffer and receptor are provided in Appendix 2.2.

Table 2

Saturation Binding Assay Microtiter Plate Layout

	1	2	3	4	5	6	7	8	9	10	11	12	
A	0,03 nM [3 H] E2 + ER			0,06 nM [3 H] E2 + ER			0,08 nM [3 H] E2 + ER			0,10 nM [3 H] E2 + ER			Total Binding (Solvent)
B	0,30 nM [3 H] E2 + ER			0,60 nM [3 H] E2 + ER			1,0 nM [3 H] E2 + ER			3,0 nM [3 H] E2 + ER			
C													
D	0,03 nM [3 H] E2 + ER + 0,03 μ M E2			0,06 nM [3 H] E2 + ER + 0,06 μ M E2			0,08 nM [3 H] E2 + ER + 0,08 μ M E2			0,10 nM [3 H] E2 + ER + 0,10 μ M E2			Non-Specific Binding
E	0,30 nM [3 H] E2 + ER + 0,30 μ M E2			0,60 nM [3 H] E2 + ER + 0,60 μ M E2			1,0 nM [3 H] E2 + ER + 1,0 μ M E2			3,0 nM [3 H] E2 + ER + 3,0 μ M E2			
F													
G													
H													

[3 H] E2: [3 H]-17 β -estradiol

ER: estrogen receptor

E2: unlabelled 17 β -estradiol

19. Assay microtiter plates should be incubated at 2° to 8°C for 16 to 20 hours and placed on a rotator during the incubation period.

Measurement of [3 H]-17 β -Estradiol bound to hrER α

20. [3 H]-17 β -Estradiol bound to hrER α should be separated from free [3 H]-17 β -Estradiol by adding 80 μ l of cold DCC suspension to each well, shaking the microtiter plates for 10 minutes and centrifugating for 10 minutes at about 2 500 RPM. To minimise dissociation of bound [3 H]-17 β -estradiol from the hrER α during this process, it is extremely important that the buffers and assay wells be kept between 2° and 8°C and that each step be conducted quickly. A shaker for microtiter plates is necessary to process plates efficiently and quickly.
21. 50 μ l of supernatant containing the hrER α -bound [3 H]-17 β -estradiol should then be taken with extreme care, to avoid any contamination of the wells by touching DCC, and should be placed on a second microtiter plate.
22. 200 μ l of scintillation fluid, capable of converting the kinetic energy of nuclear emissions into light energy, should then be added to each well (A1-B12 and D1 to E12). Wells G1-H12 (identified as total dpms) represent serial dilutions of the [3 H]-17 β -estradiol (40 μ l) that should be delivered directly into the scintillation fluid in the wells of the measurement plate as indicated in Table 3, i.e. these wells contain only 200 μ l of scintillation fluid and the appropriate dilution of [3 H]-17 β -estradiol. These measures demonstrate how much [3 H]-17 β -estradiol in dpms was added to each set of wells for the total binding and non-specific binding.

Table 3

Saturation Binding Assay Microtiter Plate Layout, Radioactivity Measurement

	1	2	3	4	5	6	7	8	9	10	11	12	
A	0,03 nM [3H] E2 + ER			0,06 nM [3H] E2 + ER			0,08 nM [3H] E2 + ER			0,10 nM [³ H] E2 + ER			Total Binding (Solvent)
B	0,30 nM [3H] E2 + ER			0,60 nM [3H] E2 + ER			1,0 nM [3H] E2 + ER			3,0 nM [3H] E2 + ER			
C													
D	0,03 nM [3H] E2 + ER + 0,03 µM E2			0,06 nM [3H] E2 + ER + 0,06 µM E2			0,08 nM [3H] E2 + ER + 0,08 µM E2			0,10 nM [3H] E2 + ER + 0,10 µM E2			Non-Specific Binding
E	0,30 nM [3H] E2 + ER + 0,30 µM E2			0,60 nM [3H] E2 + ER + 0,60 µM E2			1,0 nM [3H] E2 + ER + 1,0 µM E2			3,0 nM [3H] E2 + ER + 3,0 µM E2			
F													
G	0,03 nM [3H] E2 (total dpms)			0,06 nM [3H] E2			0,08 nM [3H] E2			0,10 nM [3H] E2			Total dpms (*)
H	0,30 nM [3H] E2			0,60 nM [3H] E2			1,0 nM [3H] E2			3,0 nM [3H] E2			

[3H] E2: [3H]-17β-estradiol

ER: estrogen receptor

E2: unlabelled 17β-estradiol

dpms: disintegrations per minute

(*) The hot serial dilutions of [³H]-labelled estradiol here should be directly added into 200 µl of scintillation fluid in wells G1 – H12.

23. Measurement should start with a delay of at least 2 hours and counting time should be 40 minutes per well. A microtiter plate scintillation counter should be used for determination of dpm/well with quench correction. Alternatively, if a scintillation counter for a microtiter plate is not available, samples may be measured in a conventional counter. Under these conditions, a reduction of counting time may be considered.

Competitive binding assay

24. The competitive binding assay measures the binding of a single concentration of [³H]-17β-estradiol in the presence of increasing concentrations of a test chemical. Three concurrent replicates should be used at each concentration within one run. In addition, three non-concurrent runs should be performed for each chemical tested. The assay should be set up in one or more 96-well microtiter plates

Controls

25. When performing the assay, concurrent solvent and controls (i.e. reference estrogen, weak binder, and non-binder) should be included in each experiment. Full concentration curves for the reference estrogen and controls (i.e. weak binder and non-binder) should be used in one plate during each run. All other plates should contain (i) a high- (maximum displacement) and medium- (approximately the IC₅₀) concentration each of E2 and weak binder in triplicate; (ii) solvent control and non-specific binding, each at least in triplicate. Procedures for the preparation of assay buffer, controls, [³H]-17β-estradiol, hrERα and test chemical solutions are described in Reference 2 (Annex K, see FW Assay Protocol).

Solvent control:

26. The solvent control indicates that the solvent does not interact with the test system and also measures total binding (TB). Ethanol is the preferred solvent. Alternatively, if the highest concentration of the test chemical is not soluble in ethanol, DMSO may be used. The concentration of ethanol or DMSO, if used, in the final assay wells is 1,5 % and may not exceed 2 %.

Buffer control:

27. The buffer control (BC) should contain neither solvent nor test chemical, but all of the other components of the assay. The results of the buffer control are compared to the solvent control to verify that the solvent used does not affect the assay system.

Strong binder (reference estrogen)

28. 17 β -estradiol (CAS 50-28-2) is the endogenous ligand and binds with high affinity to the ER, alpha subtype. A standard curve using unlabelled 17 β -estradiol should be prepared for each hrER α competitive binding assay, to allow for an assessment of variability when conducting the assay over time within the same laboratory. Eight solutions of unlabelled 17 β -estradiol should be prepared in ethanol, with concentrations in the assay wells ranging from 100 nM – 10 pM (-7[logM] to -11[logM]), spaced as follows: (-7[logM], -8[logM], -8.5[logM], -9[logM], -9.5[logM], -10[logM], -11[logM]). The highest concentration of unlabelled 17 β -estradiol (1 μ M) also serves as the non-specific binding indicator. This concentration is distinguished by the label “NSB” in Table 4 even though it is also part of the standard curve.

Weak binder

29. A weak binder (norethynodrel (CAS68-23-5) or norethindrone (CAS 68-22-4)) should be included to demonstrate the sensitivity of each experiment and to allow an assessment of variability when conducting the assay over time. Eight solutions of the weak binder should be prepared in ethanol, with concentrations in the assay wells ranging from 3 nM to 30 μ M (-8.5[logM] to -4.5[logM]), spaced as follows: -4.5[logM], -5[logM], -5.5[logM], -6[logM], -6.5[logM], -7[logM], -7.5[logM], -8.5[logM].

Non binder

30. Octyltriethoxysilane (OTES, CAS 2 943-75-1) should be used as the negative control (non-binder). It provides assurance that the assay as run, will detect when test chemicals do not bind to the hrER α . Eight solutions of the non-binder should be prepared in ethanol, with concentrations in the assay wells ranging from 0,1nM to 1 000 μ M (-10[logM] to -3[logM]), in log increments. Di-*n*-butyl phthalate (DBP) can be used as an alternate control non-binder. Its maximum solubility has been shown to be -4[logM].

hrER α concentration

31. The amount of receptor that gives specific binding of 20 \pm 5 % of 1 nM radioligand should be used (see paragraphs 12-13 of Appendix 2). The hrER α solution should be prepared immediately prior to use.

[3H]-17 β -estradiol

32. The concentration of [³H]-17 β -estradiol in the assay wells should be of 1,0 nM.

Test Chemicals

33. In the first instance, it is necessary to conduct a solubility test to determine the limit of solubility for each test chemical and to identify the appropriate concentration range to use when conducting the test protocol. The limit of solubility of each test chemical is to be initially determined in the solvent and further confirmed under assay conditions. The final concentration tested in the assay should not exceed 1 mM. Range finder testing consists of a solvent control along with 8 log serial dilutions, starting at the maximum acceptable concentration (e.g. 1 mM or lower, based upon the limit of solubility), and the presence of cloudiness or precipitate noted (see also paragraph 35). The test chemical should be tested using 8 log concentration spaced curves as defined by the preceding range finding test. Concentrations in the second and third experiments should be adjusted as appropriate to better characterise the concentration-response curve.
34. Dilutions of the test chemical should be prepared in the appropriate solvent (see paragraph 26 of Appendix 2). If the highest concentration of the test chemical is not soluble in either ethanol or DMSO, and adding more solvent would cause the solvent concentration in the final tube to be greater than the acceptable limit, the highest concentration may be reduced to the next lower concentration. In this case, an additional concentration may be added at the low end of the concentration series. Other concentrations in the series should remain unchanged.

35. The test chemical solutions should be closely monitored when added to the assay well, as the test chemical may precipitate upon addition to the assay well. The data for all wells that contain precipitate should be excluded from curve-fitting, and the reason for exclusion of the data noted.
36. If there is prior existing information from other sources that provide a $\log(\text{IC}_{50})$ of a test chemical, it may be appropriate to geometrically space the dilutions (i.e. 0,5 log units around the expected $\log(\text{IC}_{50})$). The final result should reflect sufficient spread of concentrations on either side of the $\log(\text{IC}_{50})$, including the "top" and "bottom", such that the binding curve can be adequately characterised.

Assay plate organisation

37. Labelled microtiter plates should be prepared considering sextuple incubations with codes for the solvent control, the highest concentration of the reference estrogen which also serves as the non-specific binding (NSB) indicator, and the buffer control and considering triplicate incubations with codes for each of the eight concentrations of the non-binding control (octyltriethoxysilane), the 7 lower concentrations for the reference estrogen, the eight concentrations dose levels of the weak binder, and the 8 concentrations of each test chemical (TC). An example layout of the plate diagram for the full concentration curves for the reference estrogen and control is given below in Table 4. Additional microtiter plates are used for the test chemicals and should include plate controls, i.e. 1) a high- (maximum displacement) and medium- (approximately the IC_{50}) concentration each of E2 and weak binder in triplicate; 2) solvent control and non-specific binding, each in sextuple (Table 5). An example of a competitive assay microtiter plate layout worksheet using three unknown test chemicals is provided in Appendix 2.3. The concentrations indicated in Tables 4 and 5 are the final concentrations of the assay. The maximum concentration for E2 should be 1×10^{-7} M and for the weak binder, the highest concentration used for the weak binder on plate 1 should be used. The IC_{50} concentration has to be determined by the laboratory based on their historical control database. It is expected that this value would be similar to that observed in the validation studies (see Table 1).

Table 4

Competitive Binding Assay Microtiter Plate Layout, Full Concentration Curves for Reference Estrogen and Controls (Plate 1)

	1	2	3	4	5	6	7	8	9	10	11	12
A	TB (Solvent only)			TB (Solvent only)			NSB			NSB		
B	E2 (1×10^{-7})			E2 (1×10^{-8})			E2 ($1 \times 10^{-8,5}$)			E2 (1×10^{-9})		
C	E2 ($1 \times 10^{-9,5}$)			E2 (1×10^{-10})			E2 (1×10^{-11})			Blank (*)		
D	NE ($1 \times 10^{-4,5}$)			NE (1×10^{-5})			NE ($1 \times 10^{-5,5}$)			NE (1×10^{-6})		
E	NE ($1 \times 10^{-6,5}$)			NE (1×10^{-7})			NE ($1 \times 10^{-7,5}$)			NE ($1 \times 10^{-8,5}$)		
F	OTES (1×10^{-3})			OTES (1×10^{-4})			OTES (1×10^{-5})			OTES (1×10^{-6})		
G	OTES (1×10^{-7})			OTES (1×10^{-8})			OTES (1×10^{-9})			OTES (1×10^{-10})		
H	Blank (for hot) (**)			Blank (for hot) (**)			Buffer control			Buffer control		

In this example, the weak binder is norethinodrel (NE)

(*) real blank, well not used

(**) blank not used during the incubation, but used to confirm the total radioactivity added.

Table 5

Competitive Binding Assay Microtiter Plate Layout, Full Concentration Curves for Test Chemicals and Plate Controls

	1	2	3	4	5	6	7	8	9	10	11	12
A	TB (Solvent only)			TB (Solvent only)			NSB			NSB		
B	TC1 (1×10 ⁻³)			TC1 (1×10 ⁻⁴)			TC1 (1×10 ⁻⁵)			TC1 (1×10 ⁻⁶)		
C	TC1 (1×10 ⁻⁷)			TC1 (1×10 ⁻⁸)			TC1 (1×10 ⁻⁹)			TC1 (1×10 ⁻¹⁰)		
D	TC2 (1×10 ⁻³)			TC2 (1×10 ⁻⁴)			TC2 (1×10 ⁻⁵)			TC2 (1×10 ⁻⁶)		
E	TC2 (1×10 ⁻⁷)			TC2 (1×10 ⁻⁸)			TC2 (1×10 ⁻⁹)			TC2 (1×10 ⁻¹⁰)		
F	TC3 (1×10 ⁻³)			TC3 (1×10 ⁻⁴)			TC3 (1×10 ⁻⁵)			TC3 (1×10 ⁻⁶)		
G	TC3 (1×10 ⁻⁷)			TC3 (1×10 ⁻⁸)			TC3 (1×10 ⁻⁹)			TC3 (1×10 ⁻¹⁰)		
H	NE (IC ₅₀)			NE (1×10 ^{-4,5})			E2 (IC ₅₀)			E2 (1×10 ⁻⁷)		

In this example, the weak binder is norethinodrel (NE)

Completion of competitive binding assay

38. As shown in Table 6, 80 µl of the solvent control, buffer control, reference estrogen, weak binder, non-binder, and test chemicals prepared in assay buffer should be added to the wells. Then, 40 µl of a 4 nM [³H]-17β-estradiol solution should be added to each well. After gentle rotation for 10 to 15 minutes between 2° to 8°C, 40 µl of hrERα solution should be added. Assay microtiter plates should be incubated at 2° to 8°C for 16 to 20 hours, and placed on a rotator during the incubation period.

Table 6

Volume of Assay Components for hrER Competitive Binding Assay, Microtiter Plates

Volume (µl)	Constituent
80	Unlabelled 17β-estradiol, norethynodrel, OTES, test chemicals, solvent or buffer
40	4 nM [³ H]-17β-estradiol solution
40	hrERα solution, concentration as determined
160	Total volume in each assay well

39. The quantification of [³H]-17β-Estradiol bound to hrERα, following separation of [³H]-17β-Estradiol bound to hrERα from free [³H]-17β-Estradiol by adding 80 µl of cold DCC suspension to each well, should then be performed as described in paragraphs 20-23 for the saturation binding assay.
40. Wells H1-6 (identified as blank (for hot) in table 4) represent the dpms of the [³H]-labelled-estradiol in 40 µl. The 40 µl aliquot should be delivered directly into the scintillation fluid in wells H1 – H6.

Acceptability criteria

Saturation binding assay

41. The specific binding curve should reach a plateau as increasing concentrations of [³H]-17β-estradiol were used, indicating saturation of hrERα with ligand.
42. The specific binding at 1 nM of [3H]-17β-estradiol should be inside the acceptable range 15 % to 25 % of the average measured total radioactivity added across runs. Occasional slight excursions outside of this range are acceptable, but if runs are consistently outside this range or a particular run is significantly outside this range, the protein concentration should be adjusted and the saturation assay repeated.
43. The data should produce a linear Scatchard plot.
44. The non-specific binding should not be excessive. The value for non-specific binding should typically be <35 % of the total binding. However, the ratio might occasionally exceed this limit when measuring very low dpm for the lowest concentration of radiolabelled 17β-Estradiol tested.

Competitive binding assay

45. Increasing concentrations of unlabelled 17β-estradiol should displace [³H]-17β- estradiol from the receptor in a manner consistent with a one-site competitive binding.
46. The IC₅₀ value for the reference estrogen (i.e. 17β-estradiol) should be approximately equal to the molar concentration of [³H]-17β-estradiol plus the K_d determined from the saturation binding assay.
47. The total specific binding should be consistently within the acceptable range of 20 ± 5 % when the average measured concentration of total radioactivity added to each well was 1 nM across runs. Occasional slight excursions outside of this range are acceptable, but if runs are consistently outside this range or a particular run is significantly outside this range, the protein concentration should be adjusted.
48. The solvent should not alter the sensitivity or reproducibility of the assay. The results of the solvent control (TB wells) are compared to the buffer control to verify that the solvent used does not affect the assay system. The results of the TB and Buffer control should be comparable if there is no effect of the solvent on the assay.
49. The non-binder should not displace more than 25 % of the [³H]-17β-estradiol from the hrERα when tested up to 10⁻³ M (OTES) or 10⁻⁴ M (DBP).
50. Performance criteria were developed for the reference estrogen and two weak binders (e.g. norethynodrel, norethin-drone) using data from the validation study of the FW hrER Binding Assay (Annex N of Reference 2). 95 % confidence intervals are provided for the mean (n) +/- SD for all control runs across the laboratories participating in the validation study. 95 % confidence intervals were calculated for the curve fit parameters (i.e. top, bottom, Hillslope, logIC₅₀) for the reference estrogen and weak binders and for the log₁₀RBA of the weak binders relative to the reference estrogen and are provided as performance criteria for the positive controls. Table 1 provides expected ranges for the curve fit parameters that can be used as performance criteria. In practice, the range of the IC₅₀ may vary slightly based upon the K_d of receptor preparation and ligand concentration.

51. No performance criteria was developed for curve fit parameters for the test chemicals because of the wide array of existing potential test chemicals and variation in potential affinities and outcomes (e.g. Full curve, partial curve, no curve fit). However, professional judgment should be applied when reviewing results from each run for a test chemical. A sufficient range of concentrations of the test chemical should be used to clearly define the top (e.g. 90 - 100 % of binding) of the competitive curve. Variability among replicates at each concentration of test chemical as well as among the 3 non-concurrent runs should be reasonable and scientifically defensible. Controls from each run for a test chemical should approach the measures of performance reported for this FW assay and be consistent historical control data from each respective laboratory.

ANALYSIS OF DATA

Saturation binding assay

52. Both total and non-specific binding are measured. From these values, specific binding of increasing concentrations of [³H]-17β-estradiol under equilibrium conditions is calculated by subtracting non-specific from total. A graph of specific binding versus [³H]-17β-estradiol concentration should reach a plateau for maximum specific binding indicative of saturation of the hrERα with the [³H]-17β-estradiol. In addition, analysis of the data should document the binding of the [³H]-17β-estradiol to a single, high-affinity binding site. Non-specific, total, and specific binding should be displayed on a saturation binding curve. Further analysis of these data should use a non-linear regression analysis (e.g. BioSoft; McPherson, 1985; Motulsky, 1995) with a final display of the data as a Scatchard plot.
53. The data analysis should determine Bmax and Kd from the total binding data alone, using the assumption that non-specific binding is linear, unless justification is given for using a different method. In addition, robust regression should be used when determining the best fit unless justification is given. The method chosen for robust regression should be stated. Correction for ligand depletion (e.g. using the method of Swillens 1995) should always be used when determining Bmax and Kd from saturation binding data.

Competitive binding assay

54. The competitive binding curve is plotted as specific [³H]-17β-estradiol binding versus the concentration (log10 units) of the competitor. The concentration of the test chemical that inhibits 50 % of the maximum specific [³H]-17β-estradiol binding is the IC₅₀ value.
55. Estimates of log(IC₅₀) values for the positive controls (e.g. reference estrogen and weak binder) should be determined using an appropriate nonlinear curve fitting software to fit a four parameter Hill equation (e.g. BioSoft; McPherson, 1985; Motulsky, 1995). The top, bottom, slope, and log(IC₅₀) should generally be left unconstrained when fitting these curves. Robust regression should be used when determining the best fit unless justification is given. Correction for ligand depletion should not be used. Following the initial analysis, each binding curve should be reviewed to ensure appropriate fit to the model. The relative binding affinity (RBA) for the weak binder should be calculated as a percent of the log (IC₅₀) for the weak binder relative to the log (IC₅₀) for 17β-estradiol. Results from the positive controls and the non-binder control should be evaluated using the measures of the assay performance in paragraphs 45-50 in this Appendix 2.
56. Data for all test chemicals should be analysed using a step-wise approach to ensure that data are appropriately analysed and that each competitive binding curve is properly classified. It is recommended that each run for a test chemical initially undergo a standardised data analysis that is identical to that used for the reference estrogen and weak binder controls (see paragraph 55 above). Once completed, a technical review of the curve fit parameters as well as a visual review of how well the data fit the generated competitive binding curve for each run should be conducted. During this technical review, the observations of a concentration dependent decrease in the percent [³H]-17β-estradiol specifically bound, low variability among the technical replicates at each chemical concentration, and consistency in fit parameters among the three runs are a good indication that the assay and data analyses were conducted appropriately.

Data interpretation

57. Providing that all acceptability criteria are fulfilled, a test chemical is considered to be a binder for the hrERa if a binding curve can be fit and the lowest point on the response curve within the range of the data is less than 50 % (Figure 1).
58. Providing that all acceptability criteria are fulfilled, a test chemical is considered to be a non-binder for the hrERa if:
- A binding curve can be fit and the lowest point on the fitted response curve within the range of the data is above 75 %, or
 - A binding curve cannot be fit and the lowest unsmoothed average percent binding among the concentration groups in the data is above 75 %.
59. Test chemicals are considered equivocal if none of the above conditions are met (e.g. the lowest point on the fitted response curve is between 76 – 51 %).

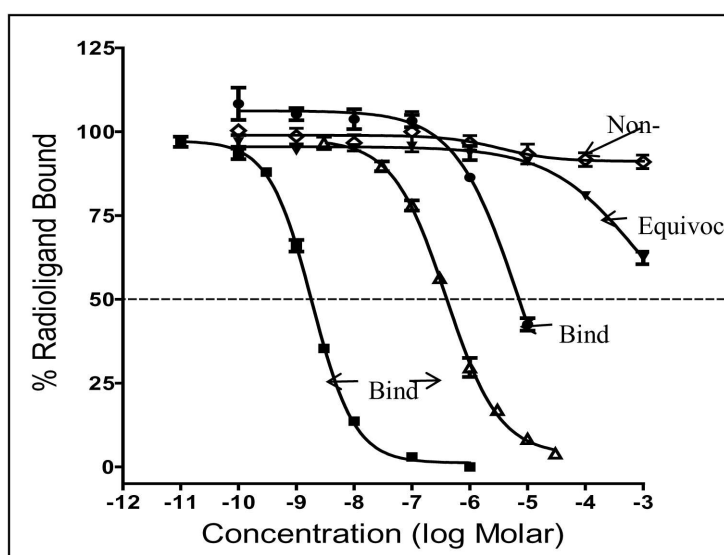
Table 7

Criteria for assigning classification based upon competitive binding curve for a test chemical

Classification	Criteria
Binder ^a	A binding curve can be fit. The lowest point on the response curve within the range of the data is less than 50 % .
Non-binder ^b	If a binding curve can be fit, the lowest point on the fitted response curve within the range of the data is above 75 % . If a binding curve cannot be fit, the lowest unsmoothed average percent binding among the concentration groups in the data is above 75 % .
Equivocal ^c	Any testable run that is neither a binder nor a non-binder (e.g. The lowest point on the fitted response curve is between 76 – 51 %).

Figure 1

Examples of test chemical classification using competitive binding curve



60. Multiple runs conducted within a laboratory for a test chemical are combined by assigning numeric values to each run and averaging across the runs as shown in Table 8. Results for the combined runs within each laboratory are compared with the expected classification for each test chemical.

Table 8

Method for classification of test chemical using multiple runs within a laboratory

To assign value to each run:	
Classification	Numeric Value
Binder	2
Equivocal	1
Non-binder	0

To classify average of numeric value across runs:	
Classification	Numeric Value
Binder	Average $\geq 1,5$
Equivocal	$0,5 \leq \text{Average} < 1,5$
Non-binder	Average $< 0,5$

TEST REPORT

61. See paragraph 24 of “**hrER BINDING ASSAY COMPONENTS**” of this test method.

Appendix 2.1

LIST OF TERMS

[³H]E2: 17β-Estradiol radiolabelled with tritium

DCC: Dextran-coated charcoal

E2: Unlabelled 17β-estradiol (inert)

Assay buffer: 10 mM Tris, 10 mg Bovine Serum Albumin /ml, 2 mM DTT, 10 % glycerol, 0,2 mM leupeptin, pH 7.5

hrERα: Human recombinant estrogen receptor alpha

Replicate: One of multiple wells that contain the same contents at the same concentrations and are assayed concurrently within a single run. In this protocol, each concentration of test chemical is tested in triplicate; that is, there are three replicates that are assayed simultaneously at each concentration of test chemical.

Run: A complete set of concurrently-run microtiter plate assay wells that provides all the information necessary to characterise binding of a test chemical to the hrERα (viz., total [³H]-17β-estradiol added to the assay well, maximum binding of [³H]-17β-estradiol to the hrERα, nonspecific binding, and total binding at various concentrations of test chemical). A run could consist of as few as one assay well (i.e. replicate) per concentration, but since this protocol requires assaying in triplicate, one run consists of three assay wells per concentration. In addition, this protocol requires three independent (i.e. non-concurrent) runs per chemical.

Appendix 2.2

TYPICAL [³H]-17 β -ESTRADIOL SATURATION ASSAY WITH THREE REPLICATE WELLS

Typical [³ H]-17 β -Estradiol Saturation Assay with Three Replicate Wells											
Position	Replicate	Well Type Code	Hot E2 Initial Concentration (nM)	Hot E2 Volume (μ l)	Hot E2 Final Concentration (nM)	Cold E2 Initial Concentration (μ M)	Cold E2 Volume (μ l)	Cold E2 Final Concentration (μ M)	Buffer Volume (μ l)	Receptor Volume (μ l)	Total volume in wells
A1	1	H	0,12	40	0,03	—	—	—	80	40	160
A2	2	H	0,12	40	0,03	—	—	—	80	40	160
A3	3	H	0,12	40	0,03	—	—	—	80	40	160
A4	1	H	0,24	40	0,06	—	—	—	80	40	160
A5	2	H	0,24	40	0,06	—	—	—	80	40	160
A6	3	H	0,24	40	0,06	—	—	—	80	40	160
A7	1	H	0,32	40	0,08	—	—	—	80	40	160
A8	2	H	0,32	40	0,08	—	—	—	80	40	160
A9	3	H	0,32	40	0,08	—	—	—	80	40	160
A10	1	H	0,40	40	0,10	—	—	—	80	40	160
A11	2	H	0,40	40	0,10	—	—	—	80	40	160
A12	3	H	0,40	40	0,10	—	—	—	80	40	160
B1	1	H	1,20	40	0,30	—	—	—	80	40	160
B2	2	H	1,20	40	0,30	—	—	—	80	40	160
B3	3	H	1,20	40	0,30	—	—	—	80	40	160
B4	1	H	2,40	40	0,60	—	—	—	80	40	160
B5	2	H	2,40	40	0,60	—	—	—	80	40	160
B6	3	H	2,40	40	0,60	—	—	—	80	40	160
B7	1	H	4,00	40	1,00	—	—	—	80	40	160
B8	2	H	4,00	40	1,00	—	—	—	80	40	160

Typical [3H]-17 β -Estradiol Saturation Assay with Three Replicate Wells

Position	Replicate	Well Type Code	Hot E2 Initial Concentration (nM)	Hot E2 Volume (μ l)	Hot E2 Final Concentration (nM)	Cold E2 Initial Concentration (μ M)	Cold E2 Volume (μ l)	Cold E2 Final Concentration (μ M)	Buffer Volume (μ l)	Receptor Volume (μ l)	Total volume in wells
B9	3	H	4,00	40	1,00	—	—	—	80	40	160
B10	1	H	12,00	40	3,00	—	—	—	80	40	160
B11	2	H	12,00	40	3,00	—	—	—	80	40	160
B12	3	H	12,00	40	3,00	—	—	—	80	40	160
D1	1	HC	0,12	40	0,03	0,06	80	0,03	—	40	160
D2	2	HC	0,12	40	0,03	0,06	80	0,03	—	40	160
D3	3	HC	0,12	40	0,03	0,06	80	0,03	—	40	160
D4	1	HC	0,24	40	0,06	0,12	80	0,06	—	40	160
D5	2	HC	0,24	40	0,06	0,12	80	0,06	—	40	160
D6	3	HC	0,24	40	0,06	0,12	80	0,06	—	40	160
D7	1	HC	0,32	40	0,08	0,16	80	0,08	—	40	160
D8	2	HC	0,32	40	0,08	0,16	80	0,08	—	40	160
D9	3	HC	0,32	40	0,08	0,16	80	0,08	—	40	160
D10	1	HC	0,40	40	0,10	0,2	80	0,1	—	40	160
D11	2	HC	0,40	40	0,10	0,2	80	0,1	—	40	160
D12	3	HC	0,40	40	0,10	0,2	80	0,1	—	40	160
E1	1	HC	1,20	40	0,30	0,6	80	0,3	—	40	160
E2	2	HC	1,20	40	0,30	0,6	80	0,3	—	40	160
E3	3	HC	1,20	40	0,30	0,6	80	0,3	—	40	160
E4	1	HC	2,40	40	0,60	1,2	80	0,6	—	40	160
E5	2	HC	2,40	40	0,60	1,2	80	0,6	—	40	160

Typical [3H]-17 β -Estradiol Saturation Assay with Three Replicate Wells

Position	Replicate	Well Type Code	Hot E2 Initial Concentration (nM)	Hot E2 Volume (μ l)	Hot E2 Final Concentration (nM)	Cold E2 Initial Concentration (μ M)	Cold E2 Volume (μ l)	Cold E2 Final Concentration (μ M)	Buffer Volume (μ l)	Receptor Volume (μ l)	Total volume in wells
E6	3	HC	2,40	40	0,60	1,2	80	0,6	—	40	160
E7	1	HC	4,00	40	1,00	2	80	1	—	40	160
E8	2	HC	4,00	40	1,00	2	80	1	—	40	160
E9	3	HC	4,00	40	1,00	2	80	1	—	40	160
E10	1	HC	12,00	40	3,00	6	80	3	—	40	160
E11	2	HC	12,00	40	3,00	6	80	3	—	40	160
E12	3	HC	12,00	40	3,00	6	80	3	—	40	160
G1	1	Hot	0,12	40	0,03	—	—	—	—	—	40
G2	2	Hot	0,12	40	0,03	—	—	—	—	—	40
G3	3	Hot	0,12	40	0,03	—	—	—	—	—	40
G4	1	Hot	0,24	40	0,06	—	—	—	—	—	40
G5	2	Hot	0,24	40	0,06	—	—	—	—	—	40
G6	3	Hot	0,24	40	0,06	—	—	—	—	—	40
G7	1	Hot	0,32	40	0,08	—	—	—	—	—	40
G8	2	Hot	0,32	40	0,08	—	—	—	—	—	40
G9	3	Hot	0,32	40	0,08	—	—	—	—	—	40
G10	1	Hot	0,40	40	0,10	—	—	—	—	—	40
G11	2	Hot	0,40	40	0,10	—	—	—	—	—	40
G12	3	Hot	0,40	40	0,10	—	—	—	—	—	40
H1	1	Hot	1,20	40	0,30	—	—	—	—	—	40
H2	2	Hot	1,20	40	0,30	—	—	—	—	—	40
H3	3	Hot	1,20	40	0,30	—	—	—	—	—	40
H4	1	Hot	2,40	40	0,60	—	—	—	—	—	40

Typical [3H]-17 β -Estradiol Saturation Assay with Three Replicate Wells

Position	Replicate	Well Type Code	Hot E2 Initial Concentration (nM)	Hot E2 Volume (μ l)	Hot E2 Final Concentration (nM)	Cold E2 Initial Concentration (μ M)	Cold E2 Volume (μ l)	Cold E2 Final Concentration (μ M)	Buffer Volume (μ l)	Receptor Volume (μ l)	Total volume in wells
H5	2	Hot	2,40	40	0,60	—	—	—	—	—	40
H6	3	Hot	2,40	40	0,60	—	—	—	—	—	40
H7	1	Hot	4,00	40	1,00	—	—	—	—	—	40
H8	2	Hot	4,00	40	1,00	—	—	—	—	—	40
H9	3	Hot	4,00	40	1,00	—	—	—	—	—	40
H10	1	Hot	12,00	40	3,00	—	—	—	—	—	40
H11	2	Hot	12,00	40	3,00	—	—	—	—	—	40
H12	3	Hot	12,00	40	3,00	—	—	—	—	—	40

Note that the "hot" wells are empty during incubation. The 40 μ l are added only for scintillation counting.

Appendix 2.3

COMPETITIVE BINDING ASSAY WELL LAYOUT

Plate	Position	Replicate	Well type	Well code	Concentration code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	A1	1	total binding	TB	TB1	—	40		40	80	160	—
S	A2	2	total binding	TB	TB2	—	40		40	80	160	—
S	A3	3	total binding	TB	TB3	—	40		40	80	160	—
S	A4	1	total binding	TB	TB4	—	40		40	80	160	—
S	A5	2	total binding	TB	TB5	—	40		40	80	160	—
S	A6	3	total binding	TB	TB6	—	40		40	80	160	—
S	A7	1	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
S	A8	2	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
S	A9	3	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
S	A10	1	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
S	A11	2	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
S	A12	3	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
S	B1	1	cold E2	S	S1	2,00E-07	40	—	40	80	160	1,0E-07
S	B2	2	cold E2	S	S1	2,00E-07	40	—	40	80	160	1,0E-07
S	B3	3	cold E2	S	S1	2,00E-07	40	—	40	80	160	1,0E-07
S	B4	1	cold E2	S	S2	2,00E-08	40	—	40	80	160	1,0E-08
S	B5	2	cold E2	S	S2	2,00E-08	40	—	40	80	160	1,0E-08
S	B6	3	cold E2	S	S2	2,00E-08	40	—	40	80	160	1,0E-08
S	B7	1	cold E2	S	S3	6,00E-09	40	—	40	80	160	3,0E-09
S	B8	2	cold E2	S	S3	6,00E-09	40	—	40	80	160	3,0E-09
S	B9	3	cold E2	S	S3	6,00E-09	40	—	40	80	160	3,0E-09

Plate	Position	Replicate	Well type	Well code	Concentration code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	B10	1	cold E2	S	S4	2,00E-09	40	—	40	80	160	1,0E-09
S	B11	2	cold E2	S	S4	2,00E-09	40	—	40	80	160	1,0E-09
S	B12	3	cold E2	S	S4	2,00E-09	40	—	40	80	160	1,0E-09
S	C1	1	cold E2	S	S5	6,00E-10	40	—	40	80	160	3,0E-10
S	C2	2	cold E2	S	S5	6,00E-10	40	—	40	80	160	3,0E-10
S	C3	3	cold E2	S	S5	6,00E-10	40	—	40	80	160	3,0E-10
S	C4	1	cold E2	S	S6	2,00E-10	40	—	40	80	160	1,0E-10
S	C5	2	cold E2	S	S6	2,00E-10	40	—	40	80	160	1,0E-10
S	C6	3	cold E2	S	S6	2,00E-10	40	—	40	80	160	1,0E-10
S	C7	1	cold E2	S	S7	2,00E-11	40	—	40	80	160	1,0E-11
S	C8	2	cold E2	S	S7	2,00E-11	40	—	40	80	160	1,0E-11
S	C9	3	cold E2	S	S7	2,00E-11	40	—	40	80	160	1,0E-11
S	C10	1	blank	blank	B1	—	—	160	—	—	160	—
S	C11	2	blank	blank	B2	—	—	160	—	—	160	—
S	C12	3	blank	blank	B3	—	—	160	—	—	160	—
S	D1	1	norethy- nodrel	NE	WP1	6,00E-05	40	—	40	80	160	3,0E-05
S	D2	1	norethy- nodrel	NE	WP1	6,00E-05	40	—	40	80	160	3,0E-05
S	D3	1	norethy- nodrel	NE	WP1	6,00E-05	40	—	40	80	160	3,0E-05
S	D4	1	norethy- nodrel	NE	WP2	2,00E-05	40	—	40	80	160	1,0E-05
S	D5	1	norethy- nodrel	NE	WP2	2,00E-05	40	—	40	80	160	1,0E-05
S	D6	1	norethy- nodrel	NE	WP2	2,00E-05	40	—	40	80	160	1,0E-05

Plate	Position	Replicate	Well type	Well code	Concentration code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	D7	1	norethy-nodrel	NE	WP3	6,00E-06	40	—	40	80	160	3,0E-06
S	D8	1	norethy-nodrel	NE	WP3	6,00E-06	40	—	40	80	160	3,0E-06
S	D9	1	norethy-nodrel	NE	WP3	6,00E-06	40	—	40	80	160	3,0E-06
S	D10	1	norethy-nodrel	NE	WP4	2,00E-06	40	—	40	80	160	1,0E-06
S	D11	1	norethy-nodrel	NE	WP4	2,00E-06	40	—	40	80	160	1,0E-06
S	D12	1	norethy-nodrel	NE	WP4	2,00E-06	40	—	40	80	160	1,0E-06
S	E1	1	norethy-nodrel	NE	WP	6,00E-07	40		40	80	160	3,0E-07
S	E2	2	norethy-nodrel	NE	WP	6,00E-07	40		40	80	160	3,0E-07
S	E3	3	norethy-nodrel	NE	WP	6,00E-07	40		40	80	160	3,0E-07
S	E4	1	norethy-nodrel	NE	WP	2,00E-07	40		40	80	160	1,0E-07
S	E5	2	norethy-nodrel	NE	WP	2,00E-07	40		40	80	160	1,0E-07
S	E6	3	norethy-nodrel	NE	WP	2,00E-07	40		40	80	160	1,0E-07
S	E7	1	norethy-nodrel	NE	WP	6,00E-08	40	—	40	80	160	3,0E-08
S	E8	2	norethy-nodrel	NE	WP	6,00E-08	40	—	40	80	160	3,0E-08
S	E9	3	norethy-nodrel	NE	WP	6,00E-08	40	—	40	80	160	3,0E-08
S	E10	1	norethy-nodrel	NE	WP	6,00E-09	40	—	40	80	160	3,0E-09

Plate	Position	Replicate	Well type	Well code	Concentration code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	E11	2	norethy- nodrel	NE	WP	6,00E-09	40	—	40	80	160	3,0E-09
S	E12	3	norethy- nodrel	NE	WP	6,00E-09	40	—	40	80	160	3,0E-09
S	F1	1	OTES	N	OTES	2,00E-03	40	—	40	80	160	1,0E-03
S	F2	2	OTES	N	OTES	2,00E-03	40	—	40	80	160	1,0E-03
S	F3	3	OTES	N	OTES	2,00E-03	40	—	40	80	160	1,0E-03
S	F4	1	OTES	N	OTES	2,00E-04	40	—	40	80	160	1,0E-04
S	F5	2	OTES	N	OTES	2,00E-04	40	—	40	80	160	1,0E-04
S	F6	3	OTES	N	OTES	2,00E-04	40	—	40	80	160	1,0E-04
S	F7	1	OTES	N	OTES	2,00E-05	40	—	40	80	160	3,0E-05
S	F8	2	OTES	N	OTES	2,00E-05	40	—	40	80	160	3,0E-05
S	F9	3	OTES	N	OTES	2,00E-05	40	—	40	80	160	3,0E-05
S	F10	1	OTES	N	OTES	2,00E-06	40	—	40	80	160	1,0E-06
S	F11	2	OTES	N	OTES	2,00E-06	40	—	40	80	160	1,0E-06
S	F12	3	OTES	N	OTES	2,00E-06	40	—	40	80	160	1,0E-06
S	G1	1	OTES	N	OTES	2,00E-07	40	—	40	80	160	3,0E-07
S	G2	2	OTES	N	OTES	2,00E-07	40	—	40	80	160	3,0E-07
S	G3	3	OTES	N	OTES	2,00E-07	40	—	40	80	160	3,0E-07
S	G4	1	OTES	N	OTES	2,00E-08	40	—	40	80	160	1,0E-08
S	G5	2	OTES	N	OTES	2,00E-08	40	—	40	80	160	1,0E-08
S	G6	3	OTES	N	OTES	2,00E-08	40	—	40	80	160	1,0E-08
S	G7	1	OTES	N	OTES	2,00E-09	40	—	40	80	160	1,0E-09

Plate	Position	Replicate	Well type	Well code	Concentration code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	G8	2	OTES	N	OTES	2,00E-09	40	—	40	80	160	1,0E-09
S	G9	3	OTES	N	OTES	2,00E-09	40	—	40	80	160	1,0E-09
S	G10	1	OTES	N	OTES	2,00E-10	40	—	40	—	160	1,0E-10
S	G11	2	OTES	N	OTES	2,00E-10	40	—	40	—	160	1,0E-10
S	G12	3	OTES	N	OTES	2,00E-10	40	—	40	—	160	1,0E-10
S	H1	1	hot	H	H	—	—	—	40	—	40	—
S	H2	1	hot	H	H	—	—	—	40	—	40	—
S	H3	1	hot	H	H	—	—	—	40	—	40	—
S	H4	1	hot	H	H	—	—	—	40	—	40	—
S	H5	1	hot	H	H	—	—	—	40	—	40	—
S	H6	1	hot	H	H	—	—	—	40	—	40	—
S	H7	1	buffer control	BC	BC	—	40	80	40	—	160	—
S	H8	1	buffer control	BC	BC	—	40	80	40	—	160	—
S	H9	1	buffer control	BC	BC	—	40	80	40	—	160	—
S	H10	1	buffer control	BC	BC	—	40	80	40	—	160	—
S	H11	1	buffer control	BC	BC	—	40	80	40	—	160	—
S	H12	1	buffer control	BC	BC	—	40	80	40	—	160	—

Note that the "hot" wells are empty during incubation. The 40 µl are added only for scintillation counting.

Competitive Binding Assay Well Layout												
Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µL)	Buffer Volume (µL)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µL)	Competitor Final Concentration (M)
P1	A1	1	total binding	TB	TBB1B1	—	40	—	40	80	160	—
P1	A2	2	total binding	TB	TB2	—	40	—	40	80	160	—
P1	A3	3	total binding	TB	TB3	—	40	—	40	80	160	—
P1	A4	1	total binding	TB	TB4	—	40	—	40	80	160	—
P1	A5	2	total binding	TB	TB5	—	40	—	40	80	160	—
P1	A6	3	total binding	TB	TB6	—	40	—	40	80	160	—
P1	A7	1	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
P1	A8	2	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
P1	A9	3	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
P1	A10	1	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
P1	A11	2	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
P1	A12	3	cold E2 (high)	NSB	S0	2,00E-06	40	—	40	80	160	1,0E-06
P1	B1	1	Test Chemical 1	TC1	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	B2	2	Test Chemical 1	TC1	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	B3	3	Test Chemical 1	TC1	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	B4	1	Test Chemical 1	TC1	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	B5	2	Test Chemical 1	TC1	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	B6	3	Test Chemical 1	TC1	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	B7	1	Test Chemical 1	TC1	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	B8	2	Test Chemical 1	TC1	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	B9	3	Test Chemical 1	TC1	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	B10	1	Test Chemical 1	TC1	4	2,00E-06	40	0	40	80	160	1,0E-06
P1	B11	2	Test Chemical 1	TC1	4	2,00E-06	40	0	40	80	160	1,0E-06
P1	B12	3	Test Chemical 1	TC1	4	2,00E-06	40	0	40	80	160	1,0E-06

Competitive Binding Assay Well Layout												
Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µL)	Buffer Volume (µL)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µL)	Competitor Final Concentration (M)
P1	C1	1	Test Chemical 1	TC1	5	2,00E-07	40	0	40	80	160	1,0E-07
P1	C2	2	Test Chemical 1	TC1	5	2,00E-07	40	0	40	80	160	1,0E-07
P1	C3	3	Test Chemical 1	TC1	5	2,00E-07	40	0	40	80	160	1,0E-07
P1	C4	1	Test Chemical 1	TC1	6	2,00E-08	40	0	40	80	160	1,0E-08
P1	C5	2	Test Chemical 1	TC1	6	2,00E-08	40	0	40	80	160	1,0E-08
P1	C6	3	Test Chemical 1	TC1	6	2,00E-08	40	0	40	80	160	1,0E-08
P1	C7	1	Test Chemical 1	TC1	7	2,00E-09	40	0	40	80	160	1,0E-09
P1	C8	2	Test Chemical 1	TC1	7	2,00E-09	40	0	40	80	160	1,0E-09
P1	C9	3	Test Chemical 1	TC1	7	2,00E-09	40	0	40	80	160	1,0E-09
P1	C10	1	Test Chemical 1	TC1	8	2,00E-10	40	0	40	80	160	1,0E-10
P1	C11	2	Test Chemical 1	TC1	8	2,00E-10	40	0	40	80	160	1,0E-10
P1	C12	3	Test Chemical 1	TC1	8	2,00E-10	40	0	40	80	160	1,0E-10
P1	D1	1	Test Chemical 2	TC2	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	D2	2	Test Chemical 2	TC2	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	D3	3	Test Chemical 2	TC2	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	D4	1	Test Chemical 2	TC2	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	D5	2	Test Chemical 2	TC2	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	D6	3	Test Chemical 2	TC2	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	D7	1	Test Chemical 2	TC2	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	D8	2	Test Chemical 2	TC2	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	D9	3	Test Chemical 2	TC2	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	D10	1	Test Chemical 2	TC2	4	2,00E-06	40	0	40	80	160	1,0E-06
P1	D11	2	Test Chemical 2	TC2	4	2,00E-06	40	0	40	80	160	1,0E-06
P1	D12	3	Test Chemical 2	TC2	4	2,00E-06	40	0	40	80	160	1,0E-06

Competitive Binding Assay Well Layout

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µL)	Buffer Volume (µL)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µL)	Competitor Final Concentration (M)
P1	E1	1	Test Chemical 2	TC2	5	2,00E-07	40	0	40	80	160	1,0E-07
P1	E2	2	Test Chemical 2	TC2	5	2,00E-07	40	0	40	80	160	1,0E-07
P1	E3	3	Test Chemical 2	TC2	5	2,00E-07	40	0	40	80	160	1,0E-07
P1	E4	1	Test Chemical 2	TC2	6	—	40	0	40	80	160	1,0E-08
P1	E5	2	Test Chemical 2	TC2	6	—	40	0	40	80	160	1,0E-08
P1	E6	3	Test Chemical 2	TC2	6	—	40	0	40	80	160	1,0E-08
P1	E7	1	Test Chemical 2	TC2	7	2,00E-06	40	0	40	80	160	1,0E-09
P1	E8	2	Test Chemical 2	TC2	7	2,00E-06	40	0	40	80	160	1,0E-09
P1	E9	3	Test Chemical 2	TC2	7	2,00E-06	40	0	40	80	160	1,0E-09
P1	E10	1	Test Chemical 2	TC2	8	2,00E-06	40	0	40	80	160	1,0E-10
P1	E11	2	Test Chemical 2	TC2	8	2,00E-06	40	0	40	80	160	1,0E-10
P1	E12	3	Test Chemical 2	TC2	8	2,00E-06	40	0	40	80	160	1,0E-10
P1	F1	1	Test Chemical 3	TC3	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	F2	2	Test Chemical 3	TC3	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	F3	3	Test Chemical 3	TC3	1	2,00E-03	40	0	40	80	160	1,0E-03
P1	F4	1	Test Chemical 3	TC3	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	F5	2	Test Chemical 3	TC3	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	F6	3	Test Chemical 3	TC3	2	2,00E-04	40	0	40	80	160	1,0E-04
P1	F7	1	Test Chemical 3	TC3	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	F8	2	Test Chemical 3	TC3	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	F9	3	Test Chemical 3	TC3	3	2,00E-05	40	0	40	80	160	1,0E-05
P1	F10	1	Test Chemical 3	TC3	4	2,00E-06	40	0	40	80	160	1,0E-06
P1	F11	2	Test Chemical 3	TC3	4	2,00E-06	40	0	40	80	160	1,0E-06
P1	F12	3	Test Chemical 3	TC3	4	2,00E-06	40	0	40	80	160	1,0E-06

Competitive Binding Assay Well Layout													
Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µL)	Buffer Volume (µL)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µL)	Competitor Final Concentration (M)	
P1	G1	1	Test Chemical 3	TC3	5	2,00E-07	40	0	40	80	160	1,0E-07	
P1	G2	2	Test Chemical 3	TC3	5	2,00E-07	40	0	40	80	160	1,0E-07	
P1	G3	3	Test Chemical 3	TC3	5	2,00E-07	40	0	40	80	160	1,0E-07	
P1	G4	1	Test Chemical 3	TC3	6	2,00E-08	40	0	40	80	160	1,0E-08	
P1	G5	2	Test Chemical 3	TC3	6	2,00E-08	40	0	40	80	160	1,0E-08	
P1	G6	3	Test Chemical 3	TC3	6	2,00E-08	40	0	40	80	160	1,0E-08	
P1	G7	1	Test Chemical 3	TC3	7	2,00E-09	40	0	40	80	160	1,0E-09	
P1	G8	2	Test Chemical 3	TC3	7	2,00E-09	40	0	40	80	160	1,0E-09	
P1	G9	3	Test Chemical 3	TC3	7	2,00E-09	40	0	40	80	160	1,0E-09	
P1	G10	1	Test Chemical 3	TC3	8	2,00E-10	40	0	40	80	160	1,0E-10	
P1	G11	2	Test Chemical 3	TC3	8	2,00E-10	40	0	40	80	160	1,0E-10	
P1	G12	3	Test Chemical 3	TC3	8	2,00E-10	40	0	40	80	160	1,0E-10	
P1	H1	1	norethynodrel	NE		IC50	40	0	40	80	160		
P1	H2	2	norethynodrel	NE		IC50	40	0	40	80	160		
P1	H3	3	norethynodrel	NE		IC50	40	0	40	80	160		
P1	H4	1	norethynodrel	NE		1,00E-4,5	40	0	40	80	160		
P1	H5	2	norethynodrel	NE		1,00E-4,5	40	0	40	80	160		
P1	H6	3	norethynodrel	NE		1,00E-4,5	40	0	40	80	160		
P1	H7	1	cold E2 S			IC50	40	0	40	80	160		
P1	H8	2	cold E2 S			IC50	40	0	40	80	160		
P1	H9	3	cold E2 S			IC50	40	0	40	80	160		
P1	H10	1	cold E2 S			1,00E-7	40	0	40	80	160		
P1	H11	2	cold E2 S			1,00E-7	40	0	40	80	160		
P1	H12	3	cold E2 S			1,00E-7	40	0	40	80	160		

Appendix 3

THE CHEMICAL EVALUATION AND RESEARCH INSTITUTE (CERI) IN VITRO ESTROGEN RECEPTOR BINDING ASSAY USING A HUMAN RECOMBINANT ERA LIGAND BINDING DOMAIN PROTEIN

INITIAL CONSIDERATIONS AND LIMITATIONS (SEE ALSO GENERAL INTRODUCTION)

1. This *in vitro* Estrogen Receptor (ER α) saturation and competitive binding assay uses a ligand binding domain (LBD) of the human ER α (hrER α). This protein construct was produced by the Chemicals Evaluation Research Institute (CERI), Japan, and exists as a glutathione-S-transferase (GST) fusion protein, and is expressed in *E. coli*. The CERI protocol underwent an international multi-laboratory validation study (2) which has demonstrated its relevance and reliability for the intended purpose of the assay.
2. This assay is a screening procedure for identifying substances that can bind to the hrER α . It is used to determine the ability of a test chemical to compete with 17 β -estradiol for binding to hrER α -LBD. Quantitative assay results may include the IC₅₀ (a measure of the concentration of test chemical needed to displace half of the [³H]-17 β -estradiol from the hrER α) and the relative binding affinities of test chemicals for the hrER α compared to 17 β -estradiol. For chemical screening purposes, acceptable qualitative assay results may include classifications of test chemicals as either hrER α binders, non-binders, or equivocal based upon criteria described for the binding curves.
3. The assay uses a radioactive ligand that requires a radioactive materials license for the laboratory. All procedures with radioisotopes and hazardous chemicals should follow the regulations and procedures as described by national legislation.
4. The “**GENERAL INTRODUCTION**” and “**hrER BINDING ASSAY COMPONENTS**” should be read before using this assay for regulatory purposes. Definitions and abbreviations used in this TG are described in Appendix 1.

PRINCIPLES OF THE ASSAY (SEE ALSO GENERAL INTRODUCTION)

5. The hrER α binding assay measures the ability of a radiolabelled ligand ([³H]17 β -estradiol) to bind with the ER in the presence of increasing concentrations of a test chemical (i.e. competitor). Test chemicals that possess a high affinity for the ER compete with the radiolabelled ligand at a lower concentration as compared with those chemicals with lower affinity for the receptor.
6. This assay consists of two major components: a saturation binding experiment to characterise receptor-ligand interaction parameters, followed by a competitive binding experiment that characterises the competition between a test chemical and a radiolabelled ligand for binding to the ER.
7. The purpose of the saturation binding experiment is to characterise a particular batch of receptors for binding affinity and number in preparation for the competitive binding experiment. The saturation binding experiment measures, under equilibrium conditions, the affinity of a fixed concentration of the estrogen receptor for its natural ligand (represented by the dissociation constant, K_d), and the concentration of active receptor sites (B_{max}).
8. The competitive binding experiment measures the affinity of a substance to compete with [³H]17 β -estradiol for binding to the ER. The affinity is quantified by the concentration of test chemical that, at equilibrium, inhibits 50 % of the specific binding of the [³H]17 β -estradiol (termed the “inhibitory concentration 50 %” or IC₅₀). This can also be evaluated using the relative binding affinity (RBA, relative to the IC₅₀ of estradiol measured separately in the same run). The competitive binding experiment measures the binding of [³H]17 β -estradiol at a fixed concentration in the presence of a wide range (eight orders of magnitude) of test chemical concentrations. The data are then fit, where possible, to a form of the Hill equation (Hill, 1910) that describes the displacement of the radioligand by a one-site competitive binder. The extent of displacement of the radiolabelled estradiol at equilibrium is used to characterise the test chemical as a binder, non-binder, or generating an equivocal response.

PROCEDURE

Demonstration of Acceptable hrER α Protein Performance

9. Prior to routinely conducting the saturation and competitive binding assays, each new batch of hrER α should be shown to be performing correctly in the laboratory in which it will be used. A two-step process should be used to demonstrate performance. These steps are the following:

- Conduct a saturation [3H]-17 β -estradiol binding assay to demonstrate hrER α specificity and saturation. Nonlinear regression analysis of these data (e.g. BioSoft; McPherson, 1985; Motulsky, 1995) and the subsequent Scatchard plot should document hrER α binding affinity of the [3H]-17 β -estradiol (Kd) and the number of receptors (Bmax) for a particular batch of hrER α .
- Conduct a competitive binding assay using the control substances (reference estrogen (17 β -estradiol), a weak binder (e.g. norethynodrel or norethindrone), and a non-binder (octyltriethoxysilane, OTES). Each laboratory should establish an historical database to document the consistency of IC₅₀ and the relevant values for the reference estrogen and weak binder among experiments and different batches of hrER α . In addition, the parameters of the competitive binding curves for the control substances should be within the limits of the 95 % confidence interval (see Table 1) that were developed using data from laboratories that participated in the validation study for this assay (2).

Table 1

Performance criteria developed for the reference estrogen and weak binder, CERi hrER Binding Assay

Substance	Parameter	Mean ^(a)	Standard Deviation(n)	95 % Confidence Intervals ^(b)	
				Lower Limit	Upper Limit
17β-estradiol	Top	104,74	13,12 (70)	101,6	107,9
	Bottom	0,85	2,41 (70)	0,28	1,43
	HillSlope	-1,22	0,20 (70)	-1,27	-1,17
	LogIC ₅₀	-8,93	0,23 (70)	-8,98	-8,87
Norethynodrel	Top	101,31	10,55 (68)	98,76	103,90
	Bottom	2,39	5,01 (68)	1,18	3,60
	HillSlope	-1,04	0,21 (68)	-1,09	-0,99
	LogIC ₅₀	-6,19	0,40 (68)	-6,29	-6,10
Norethindrone ^(c)	Top	92,27	7,79 (23)	88,90	95,63
	Bottom	16,52	10,59 (23)	11,94	21,10
	Hill Slope	-1,18	0,32 (23)	-1,31	-1,04
	LogIC ₅₀	-6,01	0,54 (23)	-6,25	-5,78

^(a) Mean \pm Standard Deviation (SD) with (sample size (n) were calculated using curve fit estimates (4-parameter Hill equation) for control runs conducted in four laboratories during the validation study (see Annex N of reference 2).

^(b) The 95 % confidence are provided as a guide for acceptability criteria.

^(c) Testing of norethindrone was optional for Subtask 4 during validation study (see Reference 2, see Subtask 4). Thus, the mean \pm SD (n) were calculated using curve fit estimates (4-parameter Hill equation) for control runs conducted in two laboratories. The range for the IC₅₀ will be dependent upon the Kd of the receptor preparation and concentration of radiolabelled ligand used within each laboratory. Appropriate adjustment for the range of the IC₅₀ based upon the conditions used to conduct the assay will be acceptable.

Demonstration of laboratory proficiency

10. See paragraphs 17 and 18 and Table 2 in “**hrER BINDING ASSAY COMPONENTS**” of this test method. Each assay (saturation and competitive binding) should consist of three independent runs (i.e. with fresh dilutions of receptor, chemicals, and reagents) on different days, and each run should contain three replicates.

Determination of Receptor (hrERa) Concentration

11. The concentration of active receptor varies slightly by batch and storage conditions. For this reason, the concentration of active receptor as received from the supplier should be determined. This will yield the appropriate concentration of active receptor at the time of the run.
12. Under conditions corresponding to competitive binding (i.e. 0,5 nM [³H]-estradiol), nominal concentrations of 0,1, 0,2, 0,4 and 0,6 nM receptor should be incubated in the absence (total binding) and presence (non-specific binding) of 1 µM unlabelled estradiol. Specific binding, calculated as the difference of total and non-specific binding, is plotted against the nominal receptor concentration. The concentration of receptor that gives specific binding values corresponding to 40 % of added radiolabel is related to the corresponding receptor concentration, and this receptor concentration should be used for saturation and competitive binding experiments. Frequently, a final hrER concentration of 0,2 nM will comply with this condition.
13. If the 40 % criterion repeatedly cannot be met, the experimental set up should be checked for potential errors. Failure to achieve the 40 % criterion may indicate that there is very little active receptor in the recombinant batch, and the use of another receptor batch should then be considered.

Saturation assay

14. Eight increasing concentrations of [³H]17β-estradiol should be evaluated in triplicate, under the following three conditions (see Table 2):
 - a. In the absence of unlabelled 17β-estradiol and presence of ER. This is the determination of total binding by measure of the radioactivity in the wells that have only [³H]17β-estradiol.
 - b. In the presence of a 2000- fold excess concentration of unlabelled 17β-estradiol over labelled 17β-estradiol and presence of ER. The intent of this condition is to saturate the active binding sites with unlabelled 17β-estradiol, and by measuring the radioactivity in the wells, determine the non-specific binding. Any remaining hot estradiol that can bind to the receptor is considered to be binding at a non-specific site as the cold estradiol should be at such a high concentration that it is bound to all of the available specific sites on the receptor.
 - c. In the absence of unlabelled 17β-estradiol and absence of ER (determination of total radioactivity)

Preparation of [³H]-17β-estradiol, unlabelled 17β-estradiol solutions and hrERa

15. A 40 nM solution of [³H]-17β-estradiol should be prepared from a 1 µM stock solution of [³H]-17β-estradiol in DMSO, by adding DMSO (to prepare 200 nM) and assay buffer at room temperature (to prepare 40 nM). Using this 40 nM solution, the series of [³H]-17β-estradiol dilutions prepared, ranging from 0,313 nM to 40 nM with assay buffer at room temperature (as represented in lane 12 of Table 2). The final assay concentrations, ranging from 0,0313 to 4,0 nM, will be obtained by adding 10 µl of these solutions to the respective assay wells of a 96-well microtiter plate (see Tables 2 and 3). Preparation of assay buffer, calculation of the original [³H]-17β-estradiol stock solution based on its specific activity, preparation of dilutions and determination of the concentrations are described in depth in the CERI protocol (2).

16. Dilutions of unlabelled 17 β -estradiol solutions should be prepared from a 1 nM 17 β -estradiol stock solution by adding assay buffer to achieve eight increasing concentrations initially ranging from 0,625 μ M to 80 μ M. The final assay concentrations, ranging from 0,0625 to 8 μ M, will be obtained by adding 10 μ l of these solutions to the respective assay wells of a 96-well microtiter plate dedicated to the measurement of non-specific binding (see Tables 2 and 3). Preparation of unlabelled 17 β -estradiol dilutions is described in depth in the CERI protocol (2).
17. The concentration of receptor that gives 40 \pm 10 % specific binding should be used (see paragraphs 12-13). The hrER α solution should be prepared with ice-cold assay buffer immediately prior to use, i.e. after all wells for total binding, non-specific binding and hot ligand alone have been prepared.
18. The 96-well microtiter plates are prepared as illustrated in Table 2, with 3 replicates per [3 H]-17 β -estradiol concentration. Volume assignment of [3 H]-17 β -estradiol, unlabelled 17 β -estradiol, buffer and receptor are provided in Table 3.

Table 2

Saturation Binding Assay Microtiter Plate Layout

	1 (*)	2 (*)	3 (*)	4 (*)	5 (*)	6 (*)	7 (*)	8 (*)	9 (*)	10	11 (**)	12 (**)
	For measurement of TB			For measurement of NSB			For determination of hot ligand alone			/	Unlabelled E2 dilutions for plate column 4-6	[3 H]E2 dilutions for plate column 1-9
A	0,0313 nM [3 H] E2+ ER			0,0313 nM [3 H] E2+ 0,0625 μ M E2+ ER			0,0313 nM			/	0,625 μ M	0,313 nM
B	0,0625 nM [3 H] E2+ ER			0,0625 nM [3 H] E2+ 0,125 μ M E2+ ER			0,0625 nM			/	1,25 μ M	0,625 nM
C	0,125 nM [3 H] E2+ ER			0,125 nM [3 H] E2+ 0,25 μ M E2+ ER			0,125 nM			/	2,5 μ M	1,25 nM
D	0,250 nM [3 H] E2+ ER			0,250 nM [3 H] E2+ 0,5 μ M E2+ ER			0,250 nM			/	5 μ M	2,5 nM
E	0,50 nM [3 H] E2+ ER			0,50 nM [3 H] E2+ 1 μ M E2+ ER			0,50 nM			/	10 μ M	5 nM
F	1,00 nM [3 H] E2+ ER			1,00 nM [3 H] E2+ 2 μ M E2+ ER			1,00 nM			/	20 μ M	10 nM
G	2,00 nM [3 H] E2+ ER			2,00 nM [3 H] E2+ 4 μ M E2+ ER			2,00 nM			/	40 μ M	20 nM
H	4,00 nM [3 H] E2+ ER			4,00 nM [3 H] E2+ 8 μ M E2+ ER			4,00 nM			/	80 μ M	40 nM

TB- total binding,

NSB- non-specific binding

[3 H] E2- [3 H]17 β -estradiolE2- unlabelled 17 β -estradiol

(*) The indicated concentrations here are the final concentrations in each well

(**) The dilutions of unlabelled E2 and [3 H]E2 can be prepared in a different plate.

Table 3

Reagent Volumes for Saturation Microtiter Plate

Lane Number		1	2	3	4	5	6	7 (*)	8 (*)	9 (*)
Preparation Steps		TB Wells			NSB Wells			Hot Ligand Alone		
Volume of components for reaction wells above and order to add	Buffer	60 µl			50 µl			90 µl		
	unlabelled E2 from lane 11 in Table2	—			10 µl			—		
	[3H]E2 from lane12 in Table2	10 µl			10 µl			10 µl		
	hrERα	30 µl			30 µl			—		
Total reaction volume		100 µl			100 µl			100 µl		
Incubation		FOLLOWING 2 HOUR INCUBATION REACTION						Quantification of the radioactivity just after the preparation. No incubation		
Treatment with 0,4 % DCC		Yes			Yes			No		
Volume of 0,4 % DCC		100 µl			100 µl			—		
Filtration		Yes			Yes			No		
MEASURING THE DPMS										
Quantification volume added to scintillation cocktail		100 µl (**)			100 µl (**)			50 µl		

(*) If an LSC for microplates is used for measuring dpms, the preparation of hot ligand alone in the same assay plate of TB and NSB wells is not appropriate. The hot ligand alone should be prepared in a different plate.

(**) If centrifugation is used to separate DCC, the 50 µl of supernatant should be measured by LSC in order to avoid contamination of DCC.

19. Assay microtiter plates for the determination of total binding and non-specific binding should be incubated at room temperature (22°C to 28°C) for two hours.

Measurement of [³H]-17β-Estradiol bound to hrERα

20. Following the two hour incubation period, [³H]-17β-Estradiol bound to hrERα should be separated from free [³H]-17β-Estradiol by adding 100µl an ice cold 0,4 % DCC suspension to the wells. The plates should then be placed on ice for 10 minutes and the reaction mixture and DCC suspension should be filtered, by transfer to a microtiter plate filter, to remove DCC. A 100 µl of the filtrate should then be added to scintillation fluid in LSC vials for determination of disintegration per minute (dpms) per vial by liquid scintillation counting.
21. Alternatively, if a microplate filter is not available, removal of DCC can be obtained by centrifugation. A 50 µl of supernatant containing the hrERα-bound [³H]-17β-estradiol should then be taken with extreme care, to avoid any contamination of the wells by touching DCC, and should be used for scintillation counting.

22. The hot ligand alone condition is used for determining the disintegration per minute (dpm) of [³H]-17β-estradiol added to the assay wells. The radioactivity should be quantified just after preparation. These wells should not be incubated and should not be treated with DCC suspension but their content should be delivered directly into the scintillation fluid. These measures demonstrate how much [³H]-17β-estradiol in dpms was added to each set of wells for the total binding and non-specific binding.

Competitive binding assay

23. The competitive binding assay measures the binding of a single concentration of [³H]-17β-estradiol in the presence of increasing concentrations of a test chemical. Three concurrent replicates should be used at each concentration within one run. In addition, three non-concurrent runs should be performed for each chemical tested. The assay should be set up in one or more 96-well microtiter plates.

Controls

24. When performing the assay, concurrent solvent and controls (i.e. reference estrogen, weak binder, and non-binder) should be included in each experiment. Full concentration curves for the reference estrogen and controls (i.e. weak binder and non-binder) should be used in one plate during each run. All other plates should contain (i) a high- (maximum displacement i.e. approximately full displacement of radiolabelled ligand) and medium- (approximately, the IC₅₀) concentration of E2 and weak binder in triplicate; (ii) solvent control and non-specific binding, each in triplicate. Procedures for the preparation of assay buffer, [³H]-17β-estradiol, hrERα and test chemical solutions are described in depth in the CERI protocol (2).

Solvent control:

25. The solvent control indicates that the solvent does not interact with the test system and also measures total binding (TB). DMSO is the preferred solvent. Alternatively, if the highest concentration of the test chemical is not soluble in DMSO, ethanol may be used. The concentration of DMSO in the final assay wells should be 2,05 % and could be increased up to 2,5 % in case of lack of solubility of the test chemical. Concentrations of DMSO above 2,5 % should not be used because of interference of higher solvent concentrations with the assay. For test chemicals that are not soluble in DMSO, but are soluble in ethanol, a maximum of 2 % ethanol may be used in the assay without interference.

Buffer control:

26. The buffer control (BC) should contain neither solvent nor test chemical, but all of the other components of the assay. The results of the buffer control are compared to the solvent control to verify that the solvent used does not affect the assay system.

Strong binder (reference estrogen)

27. 17β-estradiol (CAS 50-28-2) is the endogenous ligand and binds with high affinity to the ER, alpha subtype. A standard curve using unlabelled 17β-estradiol should be prepared for each hrERα competitive binding assay, to allow for an assessment of variability when conducting the assay over time within the same laboratory. Eight solutions of unlabelled 17β-estradiol should be prepared in DMSO and assay buffer, with final concentrations in the assay wells to be used for the standard curve spaced as follows: 10⁻⁶, 10⁻⁷, 10⁻⁸, 10^{-8.5}, 10⁻⁹, 10^{-9.5}, 10⁻¹⁰, 10⁻¹¹ M. The highest concentration of unlabelled 17β-estradiol (1 μM) should serve as the non-specific binding indicator. This concentration is distinguished by the label "NSB" in Table 4 even though it is also part of the standard curve.

Weak binder

28. A weak binder (norethynodrel (CAS68-23-5), or alternate, norethindrone (CAS 68-22-4)) should be included to demonstrate the sensitivity of each experiment and to allow an assessment of variability when conducting the assay over time. Eight solutions of the weak binder should be prepared in DMSO and assay buffer, with final concentrations in the assay wells as follows: 10^{-4.5}, 10^{-5.5}, 10⁻⁶, 10^{-6.5}, 10⁻⁷, 10^{-7.5}, 10⁻⁸ and 10⁻⁹ M.

Non binder

29. Octytriethoxysilane (OTES, CAS 2943-75-1) should be used as the negative control (non-binder). It provides assurance that the assay as run, will detect test chemicals that do not bind to the hrER α . Eight solutions of the non-binder should be prepared in DMSO and assay buffer, with final concentrations in the assay wells as follows: 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} , 10^{-7} , 10^{-8} , 10^{-9} , 10^{-10} M. Di-n-butyl phthalate (DBP, CAS 84-72-2) can be used as an alternative non-binder, but only tested up to 10^{-4} M. The maximum solubility of DBP in the assay has been demonstrated to be 10^{-4} M.

hrER α concentration

30. The amount of receptor that gives specific binding of 40 ± 10 % should be used (see paragraphs 12-13 of Appendix 3). The hrER α solution should be prepared by dilution of the functional hrER α into ice cold assay buffer, immediately prior to use.

[3 H]-17 β -estradiol

31. The final concentration of [3 H]-17 β -estradiol in the assay wells should be of 0,5 nM.

Test Chemicals

32. In the first instance, it is necessary to conduct a solubility test to determine the limit of solubility for each test chemical and to identify the appropriate concentration range to use when conducting the test protocol. The limit of solubility of each test chemical is to be initially determined in the solvent and then further confirmed under assay conditions. The final concentration tested in the assay should not exceed 1mM. Range finder testing includes a solvent control along with at least 8 log serial dilutions, starting at maximum acceptable concentration (e.g. 1 mM or lower, based upon the limit of solubility), and the presence of cloudiness or precipitate noted (see also paragraph 35 of Appendix 3). Once the concentration range for testing has been determined, a test chemical should be tested using 8 log concentrations spaced appropriately as defined in the preceding range finding test. Concentrations tested in the second and third experiments should be further adjusted as appropriate to better characterise the concentration response curve, if necessary.
33. Dilutions of the test chemical should be prepared in the appropriate solvent (see paragraph 25 of Appendix 3). If the highest concentration of the test chemical is not soluble in either DMSO or ethanol, and adding more solvent would cause the solvent concentration in the final tube to be greater than the acceptable limit, the highest concentration may be reduced to the next lower concentration. In this case, an additional concentration may be added at the low end of the concentration series. Other concentrations in the series should remain unchanged.
34. The test chemical solutions should be closely monitored when added to the assay well, as the test chemical may precipitate upon addition to the assay well. The data for all wells that contain precipitate should be excluded from curve-fitting, and the reason for exclusion of the data noted.
35. If there is prior existing information from other sources that provide a $\log(\text{IC}_{50})$ of a test chemical, it may be appropriate to geometrically space the dilutions more closely around the expected $\log(\text{IC}_{50})$ (i.e. 0,5 log units). The final results should show enough sufficient spread of concentrations on either side of the $\log(\text{IC}_{50})$, including the "top" and "bottom", such that the binding curve can be adequately characterised.

Assay plate organisation

36. Labelled microtiter plates should be prepared using sextuple incubations for the solvent control, the highest concentration the reference estrogen (E2) which also serves as the non-specific binding (NSB) indicator, the buffer control, and triplicate incubations for each of the eight concentrations of the non-binding control (octyltriethoxysilane), the seven lower concentrations for the reference estrogen (E2), the eight concentrations of the weak binder (norethynodrel or norethindrone), and the eight concentrations of each test chemical (TC). An example layout of the plate layout diagram for the full concentration curves for the reference estrogen and controls is give below in Table 4. Additional microtiter plates are used for the test chemical and should contain plate controls (i.e. (i) a high- (maximum displacement) and medium- (approximately, the IC_{50}) concentration of E2 and weak binder in triplicate; (ii) solvent control (as total binding) and non-specific binding, each in sextuple (Table 5). An example of a competitive assay microtiter plate layout worksheet using three unknown test chemicals is provided in Appendix 3.3. The concentrations indicated in the worksheet as well as in Tables 4 and 5 refer to the final concentrations used in each assay well. The maximum concentration for E2 should be 1×10^{-7} M and for the weak binder, the highest concentration used for the weak binder on plate 1 should be used. The IC_{50} concentration has to be determined by the laboratory based on their historical control database. The expectation is that this value would be similar to that observed in the validation studies (see table 1).

Table 4

Competitive Binding Assay Microtiter Plate Layout ⁽¹⁾, ⁽²⁾, Full Concentration Curves for Reference Estrogen and Controls (Plate 1)

	1	2	3	4	5	6	7	8	9	10	11	12
	Buffer Control and Positive Control (E2)			Weak Positive (Norethynodrel)			Negative Control (OTES)			TB and NSB		
A	Blank (*)			1×10^{-9} M			1×10^{-10} M			TB (solvent control) (2,05 % DMSO)		
B	E2 (1×10^{-11} M)			1×10^{-8} M			1×10^{-9} M					
C	E2 (1×10^{-10} M)			$1 \times 10^{-7.5}$ M			1×10^{-8} M			NSB (10^{-6} M E2)		
D	E2 ($1 \times 10^{-9.5}$ M)			1×10^{-7} M			1×10^{-7} M					
E	E2 (1×10^{-9} M)			$1 \times 10^{-6.5}$ M			1×10^{-6} M			Buffer control		
F	E2 ($1 \times 10^{-8.5}$ M)			1×10^{-6} M			1×10^{-5} M					
G	E2 (1×10^{-8} M)			$1 \times 10^{-5.5}$ M			1×10^{-4} M			Blank (for hot) (**)		
H	E2 (1×10^{-7} M)			$1 \times 10^{-4.5}$ M			1×10^{-3} M					

⁽¹⁾ Sample set up for the standards microtiter plate to be run with each experiment.

⁽²⁾ Note that this microtiter plate is made using the dilutions made in the dilution plate described for the standards in the previous sections.

In this example, the weak binder is norethinodrel (NE)

(*) real blank, well not used

(**) blank, not used during the incubation, but used to confirm the total radioactivity added.

Table 5

Competitive Binding Assay Microtiter Plate Layout, Additional Plates for Test Chemicals (TC) and Plate Controls

	1	2	3	4	5	6	7	8	9	10	11	12
	Test Chemical-1 (TC-1)			Test Chemical-2 (TC-2)			Test Chemical-3 (TC-3)			Controls		
A	TC-1 (1×10^{-10} M)			TC-2 (1×10^{-10} M)			TC-3 (1×10^{-10} M)			E2 (1×10^{-7} M)		
B	TC-1 (1×10^{-9} M)			TC-2 (1×10^{-9} M)			TC-3 (1×10^{-9} M)			E2 (IC_{50})		
C	TC-1 (1×10^{-8} M)			TC-2 (1×10^{-8} M)			TC-3 (1×10^{-8} M)			NE ($1 \times 10^{-4.5}$ M)		
D	TC-1 (1×10^{-7} M)			TC-2 (1×10^{-7} M)			TC-3 (1×10^{-7} M)			NE (IC_{50})		
E	TC-1 (1×10^{-6} M)			TC-2 (1×10^{-6} M)			TC-3 (1×10^{-6} M)			NSB (10^{-6} M E2)		
F	TC-1 (1×10^{-5} M)			TC-2 (1×10^{-5} M)			TC-3 (1×10^{-5} M)					
G	TC-1 (1×10^{-4} M)			TC-2 (1×10^{-4} M)			TC-3 (1×10^{-4} M)			TB (Solvent control)		
H	TC-1 (1×10^{-3} M)			TC-2 (1×10^{-3} M)			TC-3 (1×10^{-3} M)					

In this example, the weak binder is norethinodrel (NE)

Completion of competitive binding assay

37. Excepting wells for total binding and blanks (for hot), as shown in Table 6, 50 μ l of the assay buffer should be placed in each well, and should be mixed with 10 μ l of the solvent control, reference estrogen (E2), weak binder, non-binder, and test chemicals, respectively, 10 μ l of a 5 nM [3H]-17 β -estradiol solution. Then, 30 μ l of ice cold receptor solution was added to each plate and mixed gently. The hrER α solution should be the last reagent to be added. Assay microtiter plates should be incubated at room temperature (22° to 28°C) for 2 hours.

Table 6

Volume of Assay Components for hrER Competitive Binding Assay, Microtiter Plates

Preparation Steps		Other than TB wells	TB wells	Blank (for hot)
Volume of components for reaction wells above and order to add	Room Temperature assay Buffer	50 μ l	60 μ l	90 μ l
	Unlabelled E2, weak binder, non-binder, solvent and test chemicals (*)	10 μ l	—	—
	[3H]-17 β -estradiol to yield final concentration of 0,5 nM (i.e. 5 nM)	10 μ l	10 μ l	10 μ l
	hrER α concentration as determined (see paragraphs 12-13)	30 μ l	30 μ l	—
Total volume in each assay well		100 μ l	100 μ l	100 μ l

(*) properly prepared to obtain final concentration within the acceptable solvent concentration

38. The quantification of [³H]-17β-Estradiol bound to hrERα, following separation of [³H]-17β-Estradiol bound to hrERα from free [³H]-17β-Estradiol by adding 100 µl of ice-cold DCC suspension to each well, should then be performed as described in paragraphs 21-23 of Appendix 3 for the saturation binding assay.
39. Wells G10-12 and H10-12 (identified as blank (for hot) in Table 4) represent the dpms of the [³H]-labelled-estradiol in 10 µl. The 10 µl aliquot should be delivered directly into the scintillation fluid.

Acceptability criteria

Saturation binding assay

40. The specific binding curve should reach a plateau as increasing concentrations of [³H]-17β-estradiol were used, indicating saturation of hrERα with ligand.
41. The specific binding at 0,5 nM of [³H]-17β-estradiol should be inside the acceptable range 30 % to 50 % of the average measured total radioactivity added across runs. Occasional slight excursions outside of this range are acceptable, but if runs are consistently outside this range or a particular run is significantly outside this range, the protein concentration should be adjusted and the saturation assay repeated.
42. The data should produce a linear Scatchard plot.
43. The non-specific binding should not be excessive. The value for non-specific binding should typically be <35 % of the total binding. However, the ratio might occasionally exceed this limit when measuring very low dpm for the lowest concentration of radiolabelled 17β-estradiol tested.

Competitive binding assay

44. Increasing concentrations of unlabelled 17β-estradiol should displace [³H]-17β-estradiol from the receptor in a manner consistent with a one-site competitive binding.
45. The IC₅₀ value for the reference estrogen (i.e. 17β-estradiol) should be approximately equal to the molar concentration of [³H]-17β-estradiol plus the K_d determined from the saturation binding assay.
46. The total specific binding should be consistently within the acceptable range of 40 ± 10 % when the average measured concentration of total radioactivity added to each well was 0,5 nM across runs. Occasional slight excursions outside of this range are acceptable, but if runs are consistently outside this range or a particular run is significantly outside this range, the protein concentration should be adjusted.
47. The solvent should not alter the sensitivity or reproducibility of the assay. The results of the solvent control (TB wells) are compared to the buffer control to verify that the solvent used does not affect the assay system. The results of the TB and Buffer control should be comparable if there is no effect of the solvent on the assay.
48. The non-binder should not displace more than 25 % of the [³H]-17β-estradiol from the hrERα when tested up to 10⁻³ M (OTES) or 10⁻⁴ M (DBP).

49. Performance criteria were developed for the reference estrogen and two weak binders (e.g. norethynodrel, norethindrone) using data from the validation study for the CERI hrER Binding Assay (Annex N of reference 2). 95 % confidence intervals are provided for the mean \pm SD (n) of all control runs across four laboratories that participated in the validation study. 95 % confidence intervals were calculated for the curve fit parameters (i.e. top, bottom, Hillslope and Log IC₅₀) for the reference estrogen and weak binders, and the Log₁₀RBA of the weak binders relative to the reference estrogen. Table 1 provides expected ranges for the curve fit parameters that can be used as performance criteria. In practice, the range of the IC₅₀ may vary slightly based upon the experimentally derived Kd of the receptor preparation and ligand concentration used for the assay.
50. No performance criteria were developed for curve fit parameters for the test chemicals because of the wide array of existing potential test chemicals and variation in potential affinities and outcomes (e.g. Full curve, partial curve, no curve fit). However, professional judgment should be applied when reviewing results from each run for a test chemical. A sufficient range of concentrations of the test chemical should be used to clearly define the top (e.g. 90 - 100 % of binding) of the competitive curve. Variability among replicates at each concentration of test chemical as well as among the 3 non-concurrent runs should be reasonable and scientifically defensible. Controls from each run for a test chemical should approach the measures of performance reported for this CERI assay and be consistent historical control data from each respective laboratory.

ANALYSIS OF DATA

Saturation binding assay

51. Both total and non-specific binding are measured. From these values, specific binding of increasing concentrations of [³H]-17 β -estradiol under equilibrium conditions is calculated by subtracting non-specific from total. A graph of specific binding versus [³H]-17 β -estradiol concentration should reach a plateau for maximum specific binding indicative of saturation of the hrER α with the [³H]-17 β -estradiol. In addition, analysis of the data should document the binding of the [³H]-17 β -estradiol to a single, high-affinity binding site. Non-specific, total, and specific binding should be displayed on a saturation binding curve. Further analysis of these data should use a non-linear regression analysis (e.g. BioSoft; McPherson, 1985; Motulsky, 1995) with a final display of the data as a Scatchard plot.
52. The data analysis should determine Bmax and Kd from the total binding data alone, using the assumption that non-specific binding is linear, unless justification is given for using a different method. In addition, robust regression should be used when determining the best fit unless justification is given. The method chosen for robust regression should be stated. Correction for ligand depletion (e.g. using the method of Swillens 1995) should always be used when determining Bmax and Kd from saturation binding data.

Competitive binding assay

53. The competitive binding curve is plotted as specific [³H]-17 β -estradiol binding versus the concentration (log₁₀ units) of the competitor. The concentration of the test chemical that inhibits 50 % of the maximum specific [³H]-17 β -estradiol binding is the IC₅₀ value.
54. Estimates of log(IC₅₀) values for the positive controls (e.g. reference estrogen and weak binder) should be determined using an appropriate nonlinear curve fitting software to fit a four parameter Hill equation (e.g. BioSoft; McPherson, 1985; Motulsky, 1995). The top, bottom, slope, and log(IC₅₀) should generally be left unconstrained when fitting these curves. Robust regression should be used when determining the best fit unless justification is given. Correction for ligand depletion should not be used. Following the initial analysis, each binding curve should be reviewed to ensure appropriate fit to the model. The relative binding affinity (RBA) for the weak binder should be calculated as a percent of the log (IC₅₀) for the weak binder relative to the log (IC₅₀) for 17 β -estradiol. Results from the positive controls and the non-binder control should be evaluated using the measures of the assay performance in paragraphs 44-49 of this Appendix 3.

55. Data for all test chemicals should be analysed using a step-wise approach to ensure that data are appropriately analysed and that each competitive binding curve is properly classified. It is recommended that each run for a test chemical initially undergo a standardised data analysis that is identical to that used for the reference estrogen and weak binder controls (see paragraph 54 of this Appendix 3). Once completed, a technical review of the curve fit parameters as well as a visual review of how well the data fit the generated competitive binding curve for each run should be conducted. During this technical review, the observations of a concentration dependent decrease in the percent [³H]-17β-estradiol specifically bound, low variability among the technical replicates at each test chemical concentration, and consistency in fit parameters among the three runs are a good indication that the assay and data analyses were conducted appropriately.

Data interpretation

56. Providing that all acceptability criteria are fulfilled, a test chemical is considered to be a binder for the hrERα if a binding curve can be fit and the lowest point on the response curve within the range of the data is less than 50 % (Figure 1).

57. Providing that all acceptability criteria are fulfilled, a test chemical is considered to be a non-binder for the hrERα if:

— A binding curve can be fit and the lowest point on the fitted response curve within the range of the data is above 75 %, or

— A binding curve cannot be fit and the lowest unsmoothed average percent binding among the concentration groups in the data is above 75 %.

58. Test chemicals are considered equivocal if none of the above conditions are met (e.g. the lowest point on the fitted response curve is between 76 – 51 %).

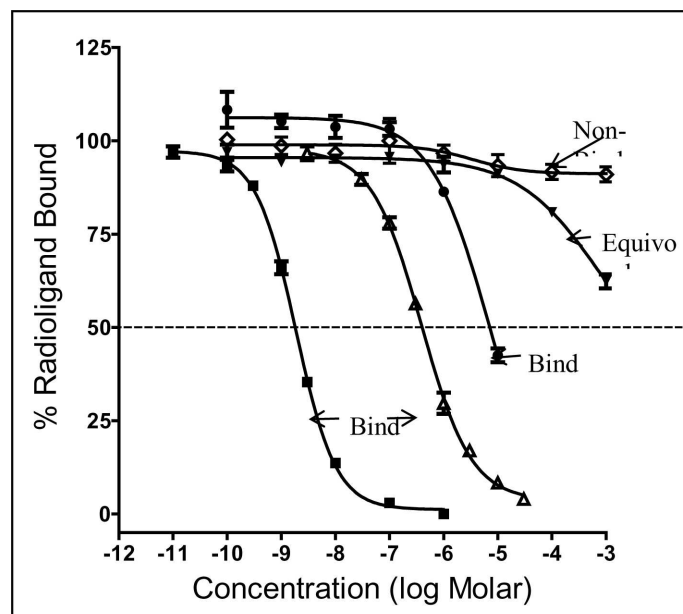
Table 7

Criteria for assigning classification based upon competitive binding curve for a test chemical

Classification	Criteria
Binder ^a	A binding curve can be fit. The lowest point on the response curve within the range of the data is less than 50 % .
Non-binder ^b	If a binding curve can be fit, the lowest point on the fitted response curve within the range of the data is above 75 % . If a binding curve cannot be fit, the lowest unsmoothed average percent binding among the concentration groups in the data is above 75 % .
Equivocal ^c	Any testable run that is neither a binder nor a non-binder (e.g. The lowest point on the fitted response curve is between 76 – 51 %).

Figure 1

Examples of test chemical classification using competitive binding curve.



59. Multiple runs conducted within a laboratory for a test chemical are combined by assigning numeric values to each run and averaging across the runs as shown in Table 8. Results for the combined runs within each laboratory are compared with the expected classification for each test chemical.

Table 8

Method for classification of test chemical using multiple runs within a laboratory

To assign value to each run:	
Classification	Numeric Value
Binder	2
Equivocal	1
Non-binder	0
To classify average of numeric value across runs:	
Classification	Numeric Value
Binder	Average $\geq 1,5$
Equivocal	$0,5 \leq \text{Average} < 1,5$
Non-binder	Average $< 0,5$

TEST REPORT

60. See paragraph 24 of "hrER BINDING ASSAY COMPONENTS" of this test method.

Appendix 3.1

LIST OF TERMS

[³H]E2: 17β-Estradiol radiolabelled with tritium

DCC: Dextran-coated charcoal

E2: Unlabelled 17β-estradiol (inert)

Assay buffer: 10 mM Tris-HCl, pH 7.4, containing 1 mM EDTA, 1mM EGTA, 1 mM NaVO₃, 10 % Glycerol, 0,2 mM Leupeptin, 1 mM Dithiothreitol and 10 mg/ml Bovine Serum Albumin

hrERα: Human recombinant estrogen receptor alpha (ligand binding domain)

Replicate: One of multiple wells that contain the same contents at the same concentrations and are assayed concurrently within a single run. In this protocol, each concentration of test chemical is tested in triplicate; that is, there are three replicates that are assayed simultaneously at each concentration of test chemical.

Run: A complete set of concurrently-run microtiter plate assay wells that provides all the information necessary to characterise binding of a test chemical to the hrERα (viz., total [³H]-17β-estradiol added to the assay well, maximum binding of [³H]-17β-estradiol to the hrERα, nonspecific binding, and total binding at various concentrations of test chemical). A run could consist of as few as one assay well (i.e. replicate) per concentration, but since this protocol requires assaying in triplicate, one run consists of three assay wells per concentration. In addition, this protocol requires three independent (i.e. non-concurrent) runs per chemical.

Appendix 3.2

COMPETITIVE BINDING ASSAY WELL LAYOUT

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	A1	1	Blank	BK	BK1	—	—	—	—	—	—	—
S	A2	2	Blank	BK	BK2	—	—	—	—	—	—	—
S	A3	3	Blank	BK	BK3	—	—	—	—	—	—	—
S	B1	1	cold E2	S	S1	1,00E-10	30	50	10	10	100	1,0E-11
S	B2	2	cold E2	S	S1	1,00E-10	30	50	10	10	100	1,0E-11
S	B3	3	cold E2	S	S1	1,00E-10	30	50	10	10	100	1,0E-11
S	C1	1	cold E2	S	S2	1,00E-09	30	50	10	10	100	1,0E-10
S	C2	2	cold E2	S	S2	1,00E-09	30	50	10	10	100	1,0E-10
S	C3	3	cold E2	S	S2	1,00E-09	30	50	10	10	100	1,0E-10
S	D1	1	cold E2	S	S3	3,16E-09	30	50	10	10	100	3,2E-10
S	D2	2	cold E2	S	S3	3,16E-09	30	50	10	10	100	3,2E-10

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	D3	3	cold E2	S	S3	3,16E-09	30	50	10	10	100	3,2E-10
S	E1	1	cold E2	S	S4	1,00E-08	30	50	10	10	100	1,0E-09
S	E2	2	cold E2	S	S4	1,00E-08	30	50	10	10	100	1,0E-09
S	E3	3	cold E2	S	S4	1,00E-08	30	50	10	10	100	1,0E-09
S	F1	1	cold E2	S	S5	3,16E-08	30	50	10	10	100	3,2E-09
S	F2	2	cold E2	S	S5	3,16E-08	30	50	10	10	100	3,2E-09
S	F3	3	cold E2	S	S5	3,16E-08	30	50	10	10	100	3,2E-09
S	G1	1	cold E2	S	S6	1,00E-07	30	50	10	10	100	1,0E-08
S	G2	2	cold E2	S	S6	1,00E-07	30	50	10	10	100	1,0E-08
S	G3	3	cold E2	S	S6	1,00E-07	30	50	10	10	100	1,0E-08
S	H1	1	cold E2	S	S7	1,00E-06	30	50	10	10	100	1,0E-07
S	H2	2	cold E2	S	S7	1,00E-06	30	50	10	10	100	1,0E-07
S	H3	3	cold E2	S	S7	1,00E-06	30	50	10	10	100	1,0E-07

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	A4	1	norethynodrel	NE	WP1	1,00E-08	30	50	10	10	100	1,0E-09
S	A5	2	norethynodrel	NE	WP1	1,00E-08	30	50	10	10	100	1,0E-09
S	A6	3	norethynodrel	NE	WP1	1,00E-08	30	50	10	10	100	1,0E-09
S	B4	1	norethynodrel	NE	WP2	1,00E-07	30	50	10	10	100	1,0E-08
S	B5	2	norethynodrel	NE	WP2	1,00E-07	30	50	10	10	100	1,0E-08
S	B6	3	norethynodrel	NE	WP2	1,00E-07	30	50	10	10	100	1,0E-08
S	C4	1	norethynodrel	NE	WP3	3,16E-07	30	50	10	10	100	3,2E-08
S	C5	2	norethynodrel	NE	WP3	3,16E-07	30	50	10	10	100	3,2E-08
S	C6	3	norethynodrel	NE	WP3	3,16E-07	30	50	10	10	100	3,2E-08
S	D4	1	norethynodrel	NE	WP4	1,00E-06	30	50	10	10	100	1,0E-07
S	D5	2	norethynodrel	NE	WP4	1,00E-06	30	50	10	10	100	1,0E-07
S	D6	3	norethynodrel	NE	WP4	1,00E-06	30	50	10	10	100	1,0E-07
S	E4	1	norethynodrel	NE	WP5	3,16E-06	30	50	10	10	100	3,2E-07

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	E5	2	norethynodrel	NE	WP5	3,16E-06	30	50	10	10	100	3,2E-07
S	E6	3	norethynodrel	NE	WP5	3,16E-06	30	50	10	10	100	3,2E-07
S	F4	1	norethynodrel	NE	WP6	1,00E-05	30	50	10	10	100	1,0E-06
S	F5	2	norethynodrel	NE	WP6	1,00E-05	30	50	10	10	100	1,0E-06
S	F6	3	norethynodrel	NE	WP6	1,00E-05	30	50	10	10	100	1,0E-06
S	G4	1	norethynodrel	NE	WP7	3,16E-05	30	50	10	10	100	3,2E-06
S	G5	2	norethynodrel	NE	WP7	3,16E-05	30	50	10	10	100	3,2E-06
S	G6	3	norethynodrel	NE	WP7	3,16E-05	30	50	10	10	100	3,2E-06
S	H4	1	norethynodrel	NE	WP8	3,16E-04	30	50	10	10	100	3,2E-05
S	H5	2	norethynodrel	NE	WP8	3,16E-04	30	50	10	10	100	3,2E-05
S	H6	3	norethynodrel	NE	WP8	3,16E-04	30	50	10	10	100	3,2E-05
S	A7	1	OTES	N	OTES1	1,00E-09	30	50	10	10	100	1,0E-10

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	A8	2	OTES	N	OTES1	1,00E-09	30	50	10	10	100	1,0E-10
S	A9	3	OTES	N	OTES1	1,00E-09	30	50	10	10	100	1,0E-10
S	B7	1	OTES	N	OTES2	1,00E-08	30	50	10	10	100	1,0E-09
S	B8	2	OTES	N	OTES2	1,00E-08	30	50	10	10	100	1,0E-09
S	B9	3	OTES	N	OTES2	1,00E-08	30	50	10	10	100	1,0E-09
S	C7	1	OTES	N	OTES3	1,00E-07	30	50	10	10	100	1,0E-08
S	C8	2	OTES	N	OTES3	1,00E-07	30	50	10	10	100	1,0E-08
S	C9	3	OTES	N	OTES3	1,00E-07	30	50	10	10	100	1,0E-08
S	D7	1	OTES	N	OTES4	1,00E-06	30	50	10	10	100	1,0E-07
S	D8	2	OTES	N	OTES4	1,00E-06	30	50	10	10	100	1,0E-07
S	D9	3	OTES	N	OTES4	1,00E-06	30	50	10	10	100	1,0E-07
S	E7	1	OTES	N	OTES5	1,00E-05	30	50	10	10	100	1,0E-06
S	E8	2	OTES	N	OTES5	1,00E-05	30	50	10	10	100	1,0E-06

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	E9	3	OTES	N	OTES5	1,00E-05	30	50	10	10	100	1,0E-06
S	F7	1	OTES	N	OTES6	1,00E-04	30	50	10	10	100	1,0E-05
S	F8	2	OTES	N	OTES6	1,00E-04	30	50	10	10	100	1,0E-05
S	F9	3	OTES	N	OTES6	1,00E-04	30	50	10	10	100	1,0E-05
S	G7	1	OTES	N	OTES7	1,00E-03	30	50	10	10	100	1,0E-04
S	G8	2	OTES	N	OTES7	1,00E-03	30	50	10	10	100	1,0E-04
S	G9	3	OTES	N	OTES7	1,00E-03	30	50	10	10	100	1,0E-04
S	H7	1	OTES	N	OTES8D-BP7	1,00E-02	30	50	10	10	100	1,0E-03
S	H8	2	OTES	N	OTES88	1,00E-02	30	50	10	10	100	1,0E-03
S	H9	3	OTES	N	OTES8	1,00E-02	30	50	10	10	100	1,0E-03
S	A10	1	total binding	TB	TB1	—	30	60	10	—	100	—
S	A11	2	total binding	TB	TB2	—	30	60	10	—	100	—
S	A12	3	total binding	TB	TB3	—	30	60	10	—	100	—
S	B10	4	total binding	TB	TB4	—	30	60	10	—	100	-

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	B11	5	total binding	TB	TB5	—	30	60	10	—	100	—
S	B12	6	total binding	TB	TB6	—	30	60	10	—	100	—
S	C10	1	cold E2 (high)	NSB	S1	1,00E-05	30	50	10	10	100	1,0E-06
S	C11	2	cold E2 (high)	NSB	S2	1,00E-05	30	50	10	10	100	1,0E-06
S	C12	3	cold E2 (high)	NSB	S3	1,00E-05	30	50	10	10	100	1,0E-06
S	D10	4	cold E2 (high)	NSB	S4	1,00E-05	30	50	10	10	100	1,0E-06
S	D11	5	cold E2 (high)	NSB	S5	1,00E-05	30	50	10	10	100	1,0E-06
S	D12	6	cold E2 (high)	NSB	S6	1,00E-05	30	50	10	10	100	1,0E-06
S	E10	1	Buffer control	BC	BC1	—	—	100	—	—	100	—
S	E11	2	Buffer control	BC	BC2	—	—	100	—	—	100	—
S	E12	3	Buffer control	BC	BC3	—	—	100	—	—	100	—
S	F10	4	Buffer control	BC	BC4	—	—	100	—	—	100	—
S	F11	5	Buffer control	BC	BC5	—	—	100	—	—	100	—
S	F12	6	Buffer control	BC	BC6	—	—	100	—	—	100	—
S	G10 (*)	1	Blank (for hot)	Hot	H1	—	90	—	10	—	100	—

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
S	G11 (*)	2	Blank (for hot)	Hot	H2	—	90	—	10	—	100	—
S	G12 (*)	3	Blank (for hot)	Hot	H3	—	90	—	10	—	100	—
S	H10 (*)	4	Blank (for hot)	Hot	H4	—	90	—	10	—	100	—
S	H11 (*)	5	Blank (for hot)	Hot	H5	—	90	—	10	—	100	—
S	H12	6	Blank (for hot)	Hot	H6	—	90	—	10	—	100	—
P1	A1	1	Unknown 1	U1	1	1,00E-09	30	50	10	10	100	1,0E-10
P1	A2	2	Unknown 1	U1	1	1,00E-09	30	50	10	10	100	1,0E-10
P1	A3	3	Unknown 1	U1	1	1,00E-09	30	50	10	10	100	1,0E-10
P1	B1	1	Unknown 1	U1	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	B2	2	Unknown 1	U1	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	B3	3	Unknown 1	U1	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	C1	1	Unknown 1	U1	3	1,00E-07	30	50	10	10	100	1,0E-08
P1	C2	2	Unknown 1	U1	3	1,00E-07	30	50	10	10	100	1,0E-08
P1	C3	3	Unknown 1	U1	3	1,00E-07	30	50	10	10	100	1,0E-08

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
P1	D1	1	Unknown 1	U1	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	D2	2	Unknown 1	U1	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	D3	3	Unknown 1	U1	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	E1	1	Unknown 1	U1	5	1,00E-05	30	50	10	10	100	1,0E-06
P1	E2	2	Unknown 1	U1	5	1,00E-05	30	50	10	10	100	1,0E-06
P1	E3	3	Unknown 1	U1	5	1,00E-05	30	50	10	10	100	1,0E-06
P1	F1	1	Unknown 1	U1	6	1,00E-04	30	50	10	10	100	1,0E-05
P1	F2	2	Unknown 1	U1	6	1,00E-04	30	50	10	10	100	1,0E-05
P1	F3	3	Unknown 1	U1	6	1,00E-04	30	50	10	10	100	1,0E-05
P1	G1	1	Unknown 1	U1	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	G2	2	Unknown 1	U1	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	G3	3	Unknown 1	U1	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	H1	1	Unknown 1	U1	8	1,00E-02	30	50	10	10	100	1,0E-03
P1	H2	2	Unknown 1	U1	8	1,00E-02	30	50	10	10	100	1,0E-03

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
P1	H3	3	Unknown 1	U1	8	1,00E-02	30	50	10	10	100	1,0E-03
P1	A4	1	Unknown 2	U2	1	1,00E-09	30	50	10	10	100	1,0E-10
P1	A5	2	Unknown 2	U2	1	1,00E-09	30	50	10	10	100	1,0E-10
P1	A6	3	Unknown 2	U2	1	1,00E-09	30	50	10	10	100	1,0E-10
P1	B4	1	Unknown 2	U2	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	B5	2	Unknown 2	U2	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	B6	3	Unknown 2	U2	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	C4	1	Unknown 2	U2	3	1,00E-07	30	50	10	10	100	1,0E-08
P1	C5	2	Unknown 2	U2	3	1,00E-07	30	50	10	10	100	1,0E-08
P1	C6	3	Unknown 2	U2	3	1,00E-07	30	50	10	10	100	1,0E-08
P1	D4	1	Unknown 2	U2	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	D5	2	Unknown 2	U2	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	D6	3	Unknown 2	U2	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	E4	1	Unknown 2	U2	5	1,00E-05	30	50	10	10	100	1,0E-06

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
P1	E5	2	Unknown 2	U2	5	1,00E-05	30	50	10	10	100	1,0E-06
P1	E6	3	Unknown 2	U2	5	1,00E-05	30	50	10	10	100	1,0E-06
P1	F4	1	Unknown 2	U2	6	1,00E-04	30	50	10	10	100	1,0E-05
P1	F5	2	Unknown 2	U2	6	1,00E-04	30	50	10	10	100	1,0E-05
P1	F6	3	Unknown 2	U2	6	1,00E-04	30	50	10	10	100	1,0E-05
P1	G4	1	Unknown 2	U2	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	G5	2	Unknown 2	U2	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	G6	3	Unknown 2	U2	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	H4	1	Unknown 2	U2	8	1,00E-02	30	50	10	10	100	1,0E-03
P1	H5	2	Unknown 2	U2	8	1,00E-02	30	50	10	10	100	1,0E-03
P1	H6	3	Unknown 2	U2	8	1,00E-02	30	50	10	10	100	1,0E-03
P1	A7	1	Unknown 3	U3	1	1,00E-09	30	50	10	10	100	1,0E-10
P1	A8	2	Unknown 3	U3	1	1,00E-09	30	50	10	10	100	1,0E-10
P1	A9	3	Unknown 3	U3	1	1,00E-09	30	50	10	10	100	1,0E-10

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
P1	B7	1	Unknown 3	U3	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	B8	2	Unknown 3	U3	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	B9	3	Unknown 3	U3	2	1,00E-08	30	50	10	10	100	1,0E-09
P1	C7	1	Unknown 3	U3	3	1,00E-07	30	50	10	10	100	1,0E-08
P1	C8	2	Unknown 3	U3	3	1,00E-07	30	50	10	10	100	1,0E-08
P1	C9	3	Unknown 3	U3	3	1,00E-07	30	50	10	10	100	1,0E-08
P1	D7	1	Unknown 3	U3	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	D8	2	Unknown 3	U3	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	D9	3	Unknown 3	U3	4	1,00E-06	30	50	10	10	100	1,0E-07
P1	E7	1	Unknown 3	U3	5	1,00E-05	30	50	10	10	100	1,0E-06
P1	E8	2	Unknown 3	U3	5	1,00E-05	30	50	10	10	100	1,0E-06
P1	E9	3	Unknown 3	U3	5	1,00E-05	30	50	10	10	100	1,0E-06
P1	F7	1	Unknown 3	U3	6	1,00E-04	30	50	10	10	100	1,0E-05
P1	F8	2	Unknown 3	U3	6	1,00E-04	30	50	10	10	100	1,0E-05

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
P1	F9	3	Unknown 3	U3	6	1,00E-04	30	50	10	10	100	1,0E-05
P1	G7	1	Unknown 3	U3	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	G8	2	Unknown 3	U3	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	G9	3	Unknown 3	U3	7	1,00E-03	30	50	10	10	100	1,0E-04
P1	H7	1	Unknown 3	U3	8	1,00E-02	30	50	10	10	100	1,0E-03
P1	H8	2	Unknown 3	U3	8	1,00E-02	30	50	10	10	100	1,0E-03
P1	H9	3	Unknown 3	U3	8	1,00E-02	30	50	10	10	100	1,0E-03
P1	A10	1	Control E2 (max)	S	E2max1	1,00E-06	30	50	10	10	100	1,00E-07
P1	A11	2	Control E2 (max)	S	E2max2	1,00E-06	30	50	10	10	100	1,00E-07
P1	A12	3	Control E2 (max)	S	E2max3	1,00E-06	30	50	10	10	100	1,00E-07
P1	B10	1	Control E2 (IC50)	S	E2IC501	E2IC50x-10	30	50	10	10	100	E2IC50
P1	B11	2	Control E2 (IC50)	S	E2IC502	E2IC50x-10	30	50	10	10	100	E2IC50
P1	B12	3	Control E2 (IC50)	S	E2IC503	E2IC50x-10	30	50	10	10	100	E2IC50

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
P1	C10	1	Control NE (max)	S	Nemax1	1,00E-3,5	30	50	10	10	100	1,00E-4,5
P1	C11	2	Control NE (max)	S	Nemax2	1,00E-3,5	30	50	10	10	100	1,00E-4,5
P1	C12	3	Control NE (max)	S	Nemax3	1,00E-3,5	30	50	10	10	100	1,00E-4,5
P1	D10	1	Control NE (IC50)	S	NEIC501	NEIC50 x10	30	50	10	10	100	NEIC50
P1	D11	2	Control NE (IC50)	S	NEIC502	NEIC50 x10	30	50	10	10	100	NEIC50
P1	D12	3	Control NE (IC50)	S	NEIC503	NEIC50 x10	30	50	10	10	100	NEIC50
P1	E10	1	cold E2 (high)	NSB	S1	1,00E-05	30	50	10	10	100	1,0E-06
P1	E11	2	cold E2 (high)	NSB	S2	1,00E-05	30	50	10	10	100	1,0E-06
P1	E12	3	cold E2 (high)	NSB	S3	1,00E-05	30	50	10	10	100	1,0E-06
P1	F10	4	cold E2 (high)	NSB	S4	1,00E-05	30	50	10	10	100	1,0E-06
P1	F11	5	cold E2 (high)	NSB	S5	1,00E-05	30	50	10	10	100	1,0E-06
P1	F12	6	cold E2 (high)	NSB	S6	1,00E-05	30	50	10	10	100	1,0E-06

Plate	Position	Replicate	Well type	Well Code	Concentration Code	Competitor Initial Concentration (M)	hrER stock (µl)	Buffer Volume (µl)	Tracer (Hot E2) Volume (µL)	Volume from dilution plate (µL)	Final Volume (µl)	Competitor Final Concentration (M)
P1	G10	1	total binding	TB	TB1	—	30	60	10	—	100	—
P1	G11	2	total binding	TB	TB2	—	30	60	10	—	100	—
P1	G12	3	total binding	TB	TB3	—	30	60	10	—	100	—
P1	H10	4	total binding	TB	TB4	—	30	60	10	—	100	—
P1	H11	5	total binding	TB	TB5	—	30	60	10	—	100	—
P1	H12	6	total binding	TB	TB6	—	30	60	10	—	100	—

(*) Note that the "hotspace" wells are empty during incubation. The 10 µl are added only for scintillation counting.

Appendix 4

CONSIDERATIONS FOR THE ANALYSIS OF DATA FROM THE HRER COMPETITIVE BINDING ASSAY

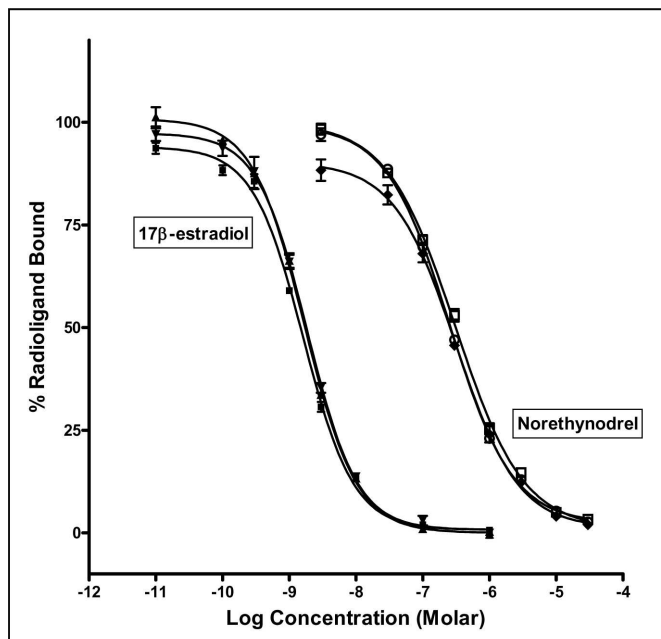
1. The hrER α competitive binding assay measures the binding of a single concentration of [3 H]-17 β -estradiol in the presence of increasing concentrations of a test chemical. The competitive binding curve is plotted as specific [3 H]-17 β -estradiol binding versus the concentration (log₁₀ units) of the competitor. The concentration of the test chemical that inhibits 50 % of the maximum specific [3 H]-17 β -estradiol binding is the IC₅₀.

Data Analysis for the Reference Estrogen and Weak Binder (1)

2. Data from the control runs are transformed (i.e. percent [3 H]-17 β -estradiol specific binding and the log concentration of the control chemical) for further analysis. Estimates of log(IC₅₀) values for the positive controls (e.g. reference estrogen and weak binder) should be determined using an appropriate nonlinear curve fitting software to fit a four parameter Hill equation (e.g. BioSoft; GraphPad Prism) (2). The top, bottom, slope, and log(IC₅₀) can typically be left unconstrained when fitting these curves. Robust regression should be used when determining the best fit unless justification is given. The method chosen for robust regression should be stated. Correction for ligand depletion was not needed for the FW or CERi hrER assays, but may be considered if needed. Following the initial analysis, each binding curve should be reviewed to ensure an appropriate fit to the model. The relative binding affinity (RBA) for the weak binder can be calculated as a percent of the log (IC₅₀) for the weak binder relative to the log (IC₅₀) for 17 β -estradiol. Results for the positive controls and the non-binder control should be evaluated using measures of assay performance and acceptability criteria as described in this test method (paragraph 20), Appendix 2 (FW Assay, paragraphs 41-51) and Appendix 3 (CERi Assay, paragraphs 41-51). Examples of 3 runs for the reference estrogen and weak binder are shown in Figure 1.

Figure 1

Examples of the competitive binding curves for the reference estrogen and the control weak binder

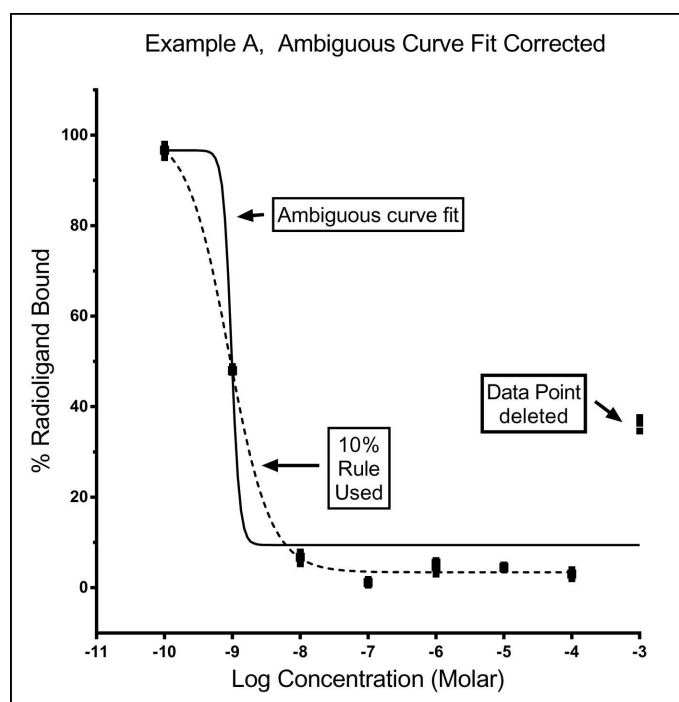


Data Analysis for Test Chemicals

- Data for all test chemicals should be analysed using a step-wise approach to ensure that data are appropriately analysed and that each competitive binding curve is properly classified. Each run for a test chemical should initially undergo a standardised data analysis that is identical to that used for the reference estrogen and weak binder controls. Once completed, a technical review of the curve fit parameters as well as a visual review of how well the data fit the generated competitive binding curve for each run should be conducted. During this technical review, the observations of a concentration dependent decrease in the percent [^3H]-17 β -estradiol specifically bound, low variability among the technical replicates at each chemical concentration, and consistency in fit parameters among the three runs are a good indication that the assay and data analyses were conducted appropriately. Professional judgment should be applied when reviewing results from each run for a test chemical, and the data used to classify each test chemical as a binder or non-binder should be scientifically defensible.
- Occasionally, there may be examples of data that require additional attention in order to appropriately analyse and interpret the hrER binding data. Previous studies had shown cases where the analysis and interpretation of competitive receptor binding data can be complicated by an upturn of the percent specific binding when testing chemicals at the highest concentrations (Figure 2). This is a well-known issue that has been encountered when using protocols for a number of competitive receptor binding assays (3). In these cases, a concentration dependent response is observed at lower concentrations, but as the concentration of the test chemical approaches the limit of solubility, the displacement of [^3H]17 β -estradiol no longer decreases. In these cases, data for the higher concentrations indicate that the biological limit of the assay has been reached. For example, this phenomenon is many times associated with chemical insolubility and precipitation at high concentrations, or may also be a reflection of exceeding the capacity of the dextran-coated charcoal to trap the unbound radiolabelled ligand during the separation procedure at the highest chemical concentrations. Leaving such data points in when fitting competitive binding data to a sigmoid curve can sometimes lead to a misclassification of the ER binding potential for a test chemical (Figure 2). To avoid this, the protocol for the FW and CER1 hrER binding assays includes an option to exclude from the analyses data points where the mean of the replicates for the percent [^3H]17 β -estradiol specific bound is 10 % or more above the mean observed at a lower concentration (i.e. This is commonly referred to as the 10 % rule). This rule can only be used once for a given curve, and there must be data remaining for at least 6 concentrations such that the curve can be correctly classified.

Figure 2

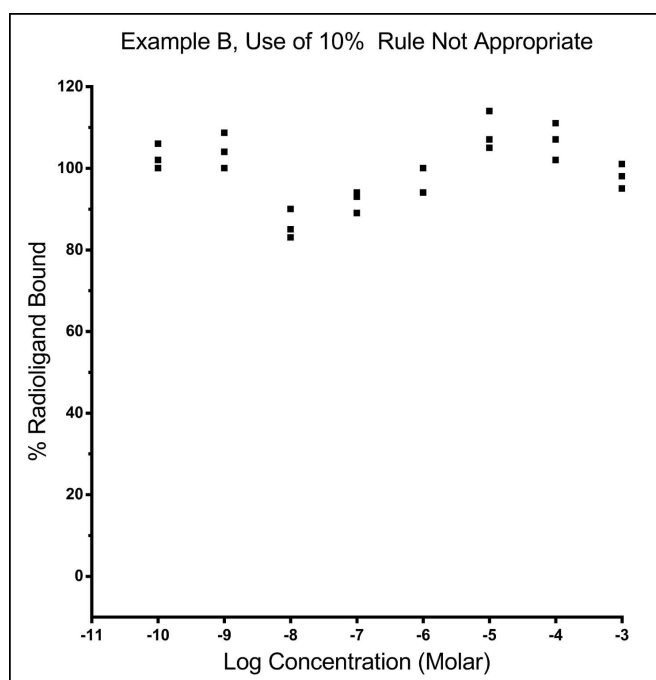
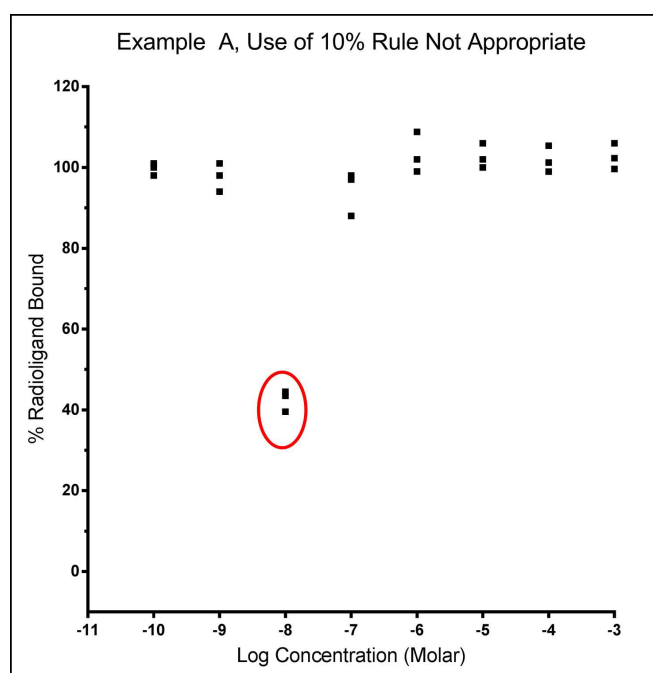
Examples, Competitive Binding Curves with and without Use of the 10 % Rule



5. The appropriate use of the 10 % rule to correct these curves should be carefully considered and reserved for those cases where there is a strong indication of a hrER binder. During the conduct of experiments for the validation study of the FW hrER Binding Assay, it was observed that the 10 % rule sometimes had an unintended and unforeseen consequence. Chemicals that did not interact with the receptor (i.e. true non-binders) often showed variability around 100 % radioligand binding that was greater than 10 % across the range of concentrations tested. If the lowest value happened to be at a low concentration, the data from all higher concentrations could potentially be deleted from the analysis by using the 10 % rule, even though those concentrations could be useful in establishing that the chemical is a non-binder. Figure 3 show examples where the use of the 10 % rule is not appropriate.

Figure 3

Examples, Competitive Binding Data Where Use of the 10 % Rule is Not Appropriate.



References

- (1) OECD (2015). *Integrated Summary Report: Validation of Two Binding Assays Using Human Recombinant Estrogen Receptor Alpha (hrERα)*, Health and Safety Publications, Series on Testing and Assessment (No 226), Organisation for Economic Cooperation and Development, Paris.
- (2) Motulsky H. and Christopoulos A. (2003). *The law of mass action*, In *Fitting Models to Biological Data Using Linear and Non-linear Regression*. GraphPad Software Inc., San Diego, CA, pp 187-191. www.graphpad.com/manuals/Prism4/RegressionBook.pdf
- (3) Laws SC, Yavanhaxay S, Cooper RL, Eldridge JC. (2006). *Nature of the Binding Interaction for 50 Structurally Diverse Chemicals with Rat Estrogen Receptors*. *Toxicological Sci.* 94(1):46-56.

B.71 IN VITRO SKIN SENSITISATION ASSAYS ADDRESSING THE KEY EVENT ON ACTIVATION OF DENDRITIC CELLS ON THE ADVERSE OUTCOME PATHWAY (AOP) FOR SKIN SENSITISATION

GENERAL INTRODUCTION

Activation of dendritic cells key event based test method

1. A skin sensitizer refers to a substance that will lead to an allergic response following skin contact as defined by the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) (1) and the European Union (EU) Regulation 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) (1). There is general agreement on the key biological events underlying skin sensitisation. The current knowledge of the chemical and biological mechanisms associated with skin sensitisation has been summarised as an Adverse Outcome Pathway (AOP) under the OECD AOP programme (2), starting with the molecular initiating event through intermediate events to the adverse effect, namely allergic contact dermatitis. In this instance, the molecular initiating event (i.e. the first key event) is the covalent binding of electrophilic substances to nucleophilic centres in skin proteins. The second key event in this AOP takes place in the keratinocytes and includes inflammatory responses as well as changes in gene expression associated with specific cell signalling pathways such as the antioxidant/electrophile response element (ARE)-dependent pathways. The third key event is the activation of dendritic cells (DC), typically assessed by expression of specific cell surface markers, chemokines and cytokines. The fourth key event is T-cell activation and proliferation, which is indirectly assessed in the murine Local Lymph Node Assay (LLNA) (3).
2. This test method (TM) is equivalent to OECD test guideline (TG) 442E (2017). It describes *in vitro* assays that address mechanisms described under the key event on activation of dendritic cells of the AOP for skin sensitisation (2). The TM comprises tests to be used for supporting the discrimination between skin sensitizers and non-sensitizers in accordance with the UN GHS and CLP.

The tests described in this TM are:

- Human Cell Line Activation Test (h-CLAT)
- U937 cell line activation Test (U-SENS™)
- Interleukin-8 Reporter Gene Assay (IL-8 Luc assay)

3. The tests included in this test method and the corresponding OECD TG may differ in relation to the procedure used to generate the data and the readouts measured but can be used indiscriminately to address countries' requirements for test results on the Key Event on activation of dendritic cells of the AOP for skin sensitisation while benefiting from the OECD Mutual Acceptance of Data.

(1) Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006, OJ L 353/1, 31.12.2008

Background and principles of the tests included in the key event based test method

4. The assessment of skin sensitisation has typically involved the use of laboratory animals. The classical methods that use guinea-pigs, the Guinea Pig Maximisation Test (GPMT) of Magnusson and Kligman, and the Buehler Test (TM B.6) (4), assess both the induction and elicitation phases of skin sensitisation. The murine tests, the LLNA (TM B.42) (3) and its two non-radioactive modifications, LLNA: DA (TM B.50) (5) and LLNA: BrdU-ELISA (TM B.51) (6), all assess the induction response exclusively, and have also gained acceptance, since they provide an advantage over the guinea pig tests in terms of animal welfare together with an objective measurement of the induction phase of skin sensitisation.
5. Recently mechanistically-based *in chemico* and *in vitro* test methods addressing the first key event (TM B.59; Direct Peptide Reactivity Assay (7)), and second key event (TM B.60; ARE-Nrf2 Luciferase Test Method (8)) of the skin sensitisation AOP have been adopted for contributing to the evaluation of the skin sensitisation hazard potential of chemicals.
6. Tests described in this test method either quantify the change in the expression of cell surface marker(s) associated with the process of activation of monocytes and DC following exposure to sensitisers (e.g. CD54, CD86) or the changes in IL-8 expression, a cytokine associated with the activation of DC. Skin sensitisers have been reported to induce the expression of cell membrane markers such as CD40, CD54, CD80, CD83, and CD86 in addition to induction of proinflammatory cytokines, such as IL-1 β and TNF- α , and several chemokines including IL-8 (CXCL8) and CCL3 (9) (10) (11) (12), associated with DC activation (2).
7. However, as DC activation represents only one key event of the skin sensitisation AOP (2) (13), information generated with tests measuring markers of DC activation alone may not be sufficient to conclude on the presence or absence of skin sensitisation potential of chemicals. Therefore data generated with the tests described in this test method are proposed to support the discrimination between skin sensitisers (i.e. UN GHS/CLP Category 1) and non-sensitisers when used within Integrated Approaches to Testing and Assessment (IATA), together with other relevant complementary information, e.g. derived from *in vitro* assays addressing other key events of the skin sensitisation AOP as well as non-testing methods, including read-across from chemical analogues (13). Examples of the use of data generated with these tests within Defined Approaches, i.e. approaches standardised both in relation to the set of information sources used and in the procedure applied to the data to derive predictions, have been published (13) and can be employed as useful elements within IATA.
8. The tests described in this test method cannot be used on their own, neither to sub-categorise skin sensitisers into subcategories 1A and 1B as defined by UN GHS/CLP, for authorities implementing these two optional subcategories, nor to predict potency for safety assessment decisions. However, depending on the regulatory framework, positive results generated with these methods may be used on their own to classify a chemical into UN GHS/CLP category 1.
9. The term "test chemical" is used in this test method to refer to what is being tested⁽¹⁾ and is not related to the applicability of the tests to the testing of mono-constituent substances, multi-constituent substances and/or mixtures. Limited information is currently available on the applicability of the tests to multi-constituent substances/mixtures (14) (15). The tests are nevertheless technically applicable to the testing of multi-constituent substances and mixtures. However, before use of this test method on a mixture for generating data for an intended regulatory purpose, it should be considered whether, and if so why, it may provide adequate results for that purpose⁽²⁾. Such considerations are not needed when there is a regulatory requirement for the testing of the mixture. Moreover, when testing multi-constituent substances or mixtures, consideration should be given to possible interference of cytotoxic constituents with the observed responses.

⁽¹⁾ In June 2013, the OECD Joint Meeting agreed that where possible, a more consistent use of the term "test chemical" describing what is being tested should be applied in new and updated OECD test guidelines.

⁽²⁾ This sentence was proposed and agreed at the April 2014 WNT meeting

LITERATURE

- (1) United Nations UN (2015). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Sixth revised edition. New York & Geneva: United Nations Publications. ISBN: 978-92-1-117006-1. Available at: https://www.unece.org/trans/danger/publi/ghs/ghs_rev06/06files_e.html
- (2) OECD (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. Series on Testing and Assessment No. 168. Available at: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO\(2012\)10/PART1&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2012)10/PART1&docLanguage=En)
- (3) Chapter B.42 of this Annex: The Local Lymph Node Assay.
- (4) Chapter B.6 of this Annex: Skin Sensitisation.
- (5) Chapter B.50 of this Annex: Skin Sensitisation: Local Lymph Node Assay: DA.
- (6) Chapter B.51 of this Annex: Skin sensitisation: Local Lymph Node Assay: BrdU-ELISA.
- (7) Chapter B.59 of this Annex: In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA).
- (8) Chapter B.60 of this Annex: *In Vitro* Skin Sensitisation: ARE-Nrf2 Luciferase Test Method.
- (9) Steinman RM. (1991). The dendritic cell system and its role in immunogenicity. *Annu Rev Immunol* 9:271-96.
- (10) Caux C, Vanbervliet B, Massacrier C, Azuma M, Okumura K, Lanier LL, and Banchereau J. (1994). B70/B7-2 is identical to CD86 and is the major functional ligand for CD28 expressed on human dendritic cells. *J Exp Med* 180:1841-7.
- (11) Aiba S, Terunuma A, Manome H, and Tagami H. (1997). Dendritic cells differently respond to haptens and irritants by their production of cytokines and expression of co-stimulatory molecules. *Eur J Immunol* 27:3031-8.

- (12) Aiba S, Manome H, Nakagawa S, Mollah ZU, Mizuashi M, Ohtani T, Yoshino Y, and Tagami. H. (2003). p38 mitogen-activated protein kinase and extracellular signal-regulated kinases play distinct roles in the activation of dendritic cells by two representative haptens, NiCl₂ and DNCB. *J Invest Dermatol* 120:390-8.
- (13) OECD (2016). Series on Testing & Assessment No 256: Guidance Document On The Reporting Of Defined Approaches And Individual Information Sources To Be Used Within Integrated Approaches To Testing And Assessment (IATA) For Skin Sensitisation, Annex 1 and Annex 2. ENV/JM/HA(2016)29. Organisation for Economic Cooperation and Development, Paris. Available at: <https://community.oecd.org/community/iatass>.
- (14) Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y, Miyazawa M, Ito Y, Nishiyama N, Itagaki H. (2010). A comparative evaluation of *in vitro* skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). *Altern. Lab. Anim.* 38, 275-284.
- (15) Piroird, C., Ovigne, J.M., Rousset, F., Martinozzi-Teissier, S., Gomes, C., Cotovio, J., Alépée, N. (2015). The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. *Toxicol. In Vitro* 29, 901-916.

Appendix 1

IN VITRO SKIN SENSITISATION: HUMAN CELL LINE ACTIVATION TEST (H-CLAT)?

INITIAL CONSIDERATIONS AND LIMITATIONS

1. The h-CLAT quantifies changes in the expression of cell surface markers associated with the process of activation of monocytes and dendritic cells (DC) (i.e. CD86 and CD54), in the human monocytic leukaemia cell line THP-1, following exposure to sensitisers (1)(2). The measured expression levels of CD86 and CD54 cell surface markers are then used for supporting the discrimination between skin sensitisers and non-sensitisers.
2. The h-CLAT has been evaluated in a European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)-coordinated validation study and subsequent independent peer review by the EURL ECVAM Scientific Advisory Committee (ESAC). Considering all available evidence and input from regulators and stakeholders, the h-CLAT was recommended by EURL ECVAM (3) to be used as part of an IATA to support the discrimination between sensitisers and non-sensitisers for the purpose of hazard classification and labelling. Examples of the use of h-CLAT data in combination with other information are reported in the literature (4)(5)(6)(7)(8)(9)(10)(11).
3. The h-CLAT proved to be transferable to laboratories experienced in cell culture techniques and flow cytometry analysis. The level of reproducibility in predictions that can be expected from the test is in the order of 80 % within and between laboratories (3)(12). Results generated in the validation study (13) and other published studies (14) overall indicate that, compared with LLNA results, the accuracy in distinguishing skin sensitisers (i.e. UN GHS/CLP Cat.1) from non-sensitisers is 85 % (N=142) with a sensitivity of 93 % (94/101) and a specificity of 66 % (27/41) (based on a re-analysis by EURL ECVAM (12) considering all existing data and not considering negative results for chemicals with a Log Kow greater than 3.5 as described in paragraph 4). False negative predictions with the h-CLAT are more likely to concern chemicals showing a low to moderate skin sensitisation potency (i.e. UN GHS/CLP subcategory 1B) than chemicals showing a high skin sensitisation potency (i.e. UN GHS/CLP subcategory 1A) (4)(13)(15). Taken together, this information indicates the usefulness of the h-CLAT method to contribute to the identification of skin sensitisation hazards. However, the accuracy values given here for the h-CLAT as a stand-alone test are only indicative, since the test should be considered in combination with other sources of information in the context of an IATA and in accordance with the provisions of paragraphs 7 and 8 in the General introduction. Furthermore, when evaluating non-animal methods for skin sensitisation, it should be kept in mind that the LLNA test as well as other animal tests may not fully reflect the situation in humans.
4. On the basis of the data currently available, the h-CLAT method was shown to be applicable to test chemicals covering a variety of organic functional groups, reaction mechanisms, skin sensitisation potency (as determined in *in vivo* studies) and physicochemical properties (3)(14)(15). The h-CLAT method is applicable to test chemicals soluble or that form a stable dispersion (i.e. a colloid or suspension in which the test chemical does not settle or separate from the solvent/vehicle into different phases) in an appropriate solvent/vehicle (see paragraph 14). Test chemicals with a Log Kow greater than 3.5 tend to produce false negative results (14). Therefore negative results with test chemicals with a Log Kow greater than 3.5 should not be considered. However, positive results obtained with test chemicals with a Log Kow greater than 3.5 could still be used to support the identification of the test chemical as a skin sensitiser. Furthermore, because of the limited metabolic capability of the cell line used (16) and because of the experimental conditions, pro-haptens (i.e. substances requiring enzymatic activation for example via P450 enzymes) and pre-haptens (i.e. substances activated by oxidation) in particular with a slow oxidation rate may also provide negative results in the h-CLAT (15). Fluorescent test chemicals can be assessed with the h-CLAT (17), nevertheless, strong fluorescent test chemicals emitting at the same wavelength as fluorescein isothiocyanate (FITC) or as propidium iodide (PI), will interfere with the flow cytometric detection and thus cannot be correctly evaluated using FITC-conjugated antibodies or PI. In such a case, other fluorochrome-tagged antibodies or other cytotoxicity markers, respectively, can be used as long as it can be shown they provide similar results as the FITC-tagged antibodies (see paragraph 24) or PI (see paragraph 18) e.g. by testing the proficiency substances in Appendix 1-2. In the light of the above, negative results should be interpreted in the context of the stated limitations and together with other information sources within the framework of IATA. In cases where there is evidence demonstrating the non-applicability of the h-CLAT method to other specific categories of test chemicals, it should not be used for those specific categories.

5. As described above, the h-CLAT method supports the discrimination between skin sensitisers from non-sensitisers. However, it may also potentially contribute to the assessment of sensitising potency (4)(5)(9) when used in integrated approaches such as IATA. Nevertheless, further work, preferably based on human data, is required to determine how h-CLAT results may possibly inform potency assessment.
6. Definitions are provided in Appendix 1.1.

PRINCIPLE OF THE TEST

7. The h-CLAT method is an *in vitro* assay that quantifies changes of cell surface marker expression (i.e. CD86 and CD54) on a human monocytic leukemia cell line, THP-1 cells, following 24 hours exposure to the test chemical. These surface molecules are typical markers of monocytic THP-1 activation and may mimic DC activation, which plays a critical role in T-cell priming. The changes of surface marker expression are measured by flow cytometry following cell staining with fluorochrome-tagged antibodies. Cytotoxicity measurement is also conducted concurrently to assess whether upregulation of surface marker expression occurs at sub-cytotoxic concentrations. The relative fluorescence intensity of surface markers compared to solvent/vehicle control are calculated and used in the prediction model (see paragraph 26), to support the discrimination between sensitisers and non-sensitisers

DEMONSTRATION OF PROFICIENCY

8. Prior to routine use of the test described in this Appendix to test method B.71, laboratories should demonstrate technical proficiency, using the 10 proficiency substances listed in Appendix 1.2. Moreover, test users should maintain an historical database of data generated with the reactivity checks (see paragraph 11) and with the positive and solvent/vehicle controls (see paragraphs 20-22), and use these data to confirm the reproducibility of the test in their laboratory is maintained over time.

PROCEDURE

9. This test is based on the h-CLAT DataBase service on ALternative Methods to animal experimentation (DB-ALM) protocol no 158 (18) which represents the protocol used for the EURL ECVAM-coordinated validation study. It is recommended that this protocol is used when implementing and using the h-CLAT method in the laboratory. The following is a description of the main components and procedures for the h-CLAT method, which comprises two steps: *dose finding assay* and *CD86/CD54 expression measurement*.

Preparation of cells

10. The human monocytic leukaemia cell line, THP-1, should be used for performing the h-CLAT method. It is recommended that cells (TIB-202™) are obtained from a well-qualified cell bank, such as the American Type Culture Collection.
11. THP-1 cells are cultured, at 37°C under 5 % CO₂ and humidified atmosphere, in RPMI-1640 medium supplemented with 10 % foetal bovine serum (FBS), 0,05 mM 2-mercaptoethanol, 100 units/ml penicillin and 100 µg/ml streptomycin. The use of penicillin and streptomycin in the culture medium can be avoided. However, in such a case users should verify that the absence of antibiotics in the culture medium has no impact on the results, for example by testing the proficiency substances listed in Appendix 1.2. In any case, in order to minimise the risk of contamination, good cell culture practices should be followed independently of the presence or not of antibiotics in the cell culture medium. THP-1 cells are routinely seeded every 2-3 days at the density of 0,1 to 0,2 × 10⁶ cells/ml. They should be maintained at densities from 0,1 to 1,0 × 10⁶ cells/ml. Prior to using them for testing, the cells should be qualified by conducting a reactivity check. The reactivity check of the cells should be performed using the positive controls, 2,4-dinitrochlorobenzene (DNCB) (CAS no 97-00-7, ≥ 99 % purity) and nickel sulfate (NiSO₄) (CAS no 10101-97-0, ≥ 99 % purity) and the negative control, lactic acid (LA) (CAS no 50-21-5, ≥ 85 % purity), two weeks after thawing. Both DNCB and NiSO₄ should produce a positive response of both CD86 and CD54 cell surface markers, and LA should produce a negative response of both CD86 and CD54 cell surface markers. Only the cells which passed the reactivity check are to be used for the assay. Cells can be propagated up to two months after thawing. Passage number should not exceed 30. The reactivity check should be performed according to the procedures described in paragraphs 20-24.

12. For testing, THP-1 cells are seeded at a density of either $0,1 \times 10^6$ cells/ml or $0,2 \times 10^6$ cells/ml, and pre-cultured in culture flasks for 72 hours or for 48 hours, respectively. It is important that the cell density in the culture flask just after the pre-culture period be as consistent as possible in each experiment (by using one of the two pre-culture conditions described above), because the cell density in the culture flask just after pre-culture could affect the CD86/CD54 expression induced by allergens (19). On the day of testing, cells harvested from culture flask are resuspended with fresh culture medium at 2×10^6 cells/ml. Then, cells are distributed into a 24 well flat-bottom plate (500 μ l, 1×10^6 cells/well) or a 96-well flat-bottom plate (80 μ l, $1,6 \times 10^5$ cells/well).

Dose finding assay

13. A *dose finding assay* is performed to determine the CV75, being the test chemical concentration that results in 75 % cell viability (CV) compared to the solvent/vehicle control. The CV75 value is used to determine the concentration of test chemicals for the CD86/CD54 expression measurement (see paragraphs 20-24).

Preparation of test chemicals and control substances

14. The test chemicals and control substances are prepared on the day of testing. For the h-CLAT method, test chemicals are dissolved or stably dispersed (see also paragraph 4) in saline or medium as first solvent/vehicle options or dimethyl sulfoxide (DMSO, 99 % purity) as a second solvent/vehicle option if the test chemical is not soluble or does not form a stable dispersion in the previous two solvents/vehicles, to final concentrations of 100 mg/ml (in saline or medium) or 500 mg/ml (in DMSO). Other solvents/vehicles than those described above may be used if sufficient scientific rationale is provided. Stability of the test chemical in the final solvent/vehicle should be taken into account.
15. Starting from the 100 mg/ml (in saline or medium) or 500 mg/ml (in DMSO) stock solutions of the test chemicals, the following dilution steps should be taken:
 - For saline or medium as solvent/vehicle: Eight stock solutions (eight concentrations) are prepared, by two-fold serial dilutions using the corresponding solvent/vehicle. These stock solutions are then further diluted 50-fold into culture medium (working solutions). If the top final concentration in the plate of 1 000 μ g/ml is non-toxic, the maximum concentration should be re-determined by performing a new cytotoxicity test. The final concentration in the plate should not exceed 5 000 μ g/ml for test chemicals dissolved or stably dispersed in saline or medium.
 - For DMSO as solvent/vehicle: Eight stock solutions (eight concentrations) are prepared, by two-fold serial dilutions using the corresponding solvent/vehicle. These stock solutions are then further diluted 250-fold into culture medium (working solutions). The final concentration in plate should not exceed 1 000 μ g/ml even if this concentration is non-toxic.

The working solutions are finally used for exposure by adding an equal volume of working solution to the volume of THP-1 cell suspension in the plate (see also paragraph 17) to achieve a further two-fold dilution (usually, the final range of concentrations in the plate is 7,81–1 000 μ g/ml).

16. The solvent/vehicle control used in the h-CLAT method is culture medium (for test chemicals solubilised or stably dispersed (see paragraph 4) either with medium or saline) or DMSO (for test chemicals solubilised or stably dispersed in DMSO) tested at a single final concentration in the plate of 0,2 %. It undergoes the same dilution as described for the working solutions in paragraph 15.

Application of test chemicals and control substances

17. The culture medium or working solutions described in paragraphs 15 and 16 are mixed 1:1 (v/v) with the cell suspensions prepared in the 24-well or 96-well flat-bottom plate (see paragraph 12). The treated plates are then incubated for $24 \pm 0,5$ hours at 37°C under 5 % CO_2 . Care should be taken to avoid evaporation of volatile test chemicals and cross-contamination between wells by test chemicals, e.g. by sealing the plate prior to the incubation with the test chemicals (20).

Propidium iodide (PI) staining

18. After 24±0.5 hours of exposure, cells are transferred into sample tubes and collected by centrifugation. The supernatants are discarded and the remaining cells are resuspended with 200 µl (in case of 96-well) or 600 µl (in case of 24-well) of a phosphate buffered saline containing 0,1 % bovine serum albumin (staining buffer). 200 µl of cell suspension is transferred into 96-well round-bottom plate (in case of 96-well) or micro tube (in case of 24-well) and washed twice with 200 µl (in case of 96-well) or 600 µl (in case of 24-well) of staining buffer. Finally, cells are resuspended in staining buffer (e.g. 400 µl) and PI solution (e.g. 20 µl) is added (for example, final concentration of PI is 0,625 µg/ml). Other cytotoxicity markers, such as 7-Aminoactinomycin D (7-AAD), Trypan blue or others may be used if the alternative stains can be shown to provide similar results as PI, for example by testing the proficiency substances in Appendix 1.2.

Cytotoxicity measurement by flow cytometry and estimation of CV75 value

19. The PI uptake is analysed using flow cytometry with the acquisition channel FL-3. A total of 10 000 living cells (PI negative) are acquired. The cell viability can be calculated using the following equation by the cytometer analysis program. When the cell viability is low, up to 30 000 cells including dead cells should be acquired. Alternatively, data can be acquired for one minute after the initiation of the analysis.

$$\text{Cell viability} = \frac{\text{Number of living cells}}{\text{Total Number of aquired cells}} \times 100$$

The CV75 value (see paragraph 13), i.e. a concentration showing 75 % of THP-1 cell survival (25 % cytotoxicity), is calculated by log-linear interpolation using the following equation:

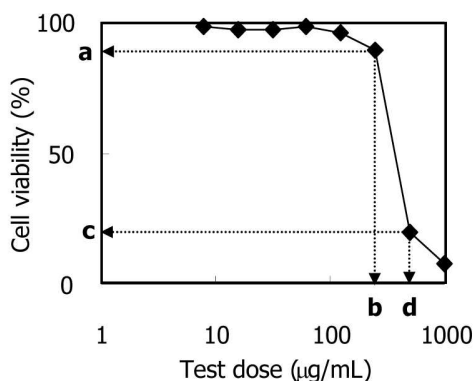
$$\text{Log CV75} = \frac{(75 - c) \times \text{Log}(b) - (75 - a) \times \text{Log}(d)}{a - c}$$

Where:

a is the minimum value of cell viability over 75 %

c is the maximum value of cell viability below 75 %

b and d are the concentrations showing the value of cell viability a and c respectively



Other approaches to derive the CV75 can be used as long as it is demonstrated that this has no impact on the results (e.g. by testing the proficiency substances).

CD86/CD54 expression measurement

Preparation of the test chemicals and control substances

20. The appropriate solvent/vehicle (saline, medium or DMSO; see paragraph 14) is used to dissolve or stably disperse the test chemicals. The test chemicals are first diluted to the concentration corresponding to 100-fold (for saline or medium) or 500-fold (for DMSO) of the $1.2 \times CV75$ determined in the *dose finding assay* (see paragraph 19). If the CV75 cannot be determined (i.e. if sufficient cytotoxicity is not observed in the *dose finding assay*), the highest soluble or stably dispersed concentration of test chemical prepared with each solvent/vehicle should be used as starting concentration. Please note that the final concentration in the plate should not exceed 5 000 µg/ml (in case of saline or medium) or 1 000 µg/ml (in case of DMSO). Then, 1.2-fold serial dilutions are made using the corresponding solvent/vehicle to obtain the stock solutions (eight concentrations ranging from $100 \times 1.2 \times CV75$ to $100 \times 0.335 \times CV75$ (for saline or medium) or from $500 \times 1.2 \times CV75$ to $500 \times 0.335 \times CV75$ (for DMSO)) to be tested in the h-CLAT method (see DB-ALM protocol N_O. 158 for an example of dosing scheme). The stock solutions are then further diluted 50-fold (for saline or medium) or 250-fold (for DMSO) into the culture medium (working solutions). These working solutions are finally used for exposure with a further final two-fold dilution factor in the plate. If the results do not meet the acceptance criteria described in the paragraphs 29 and 30 regarding cell viability, the *dose finding assay* may be repeated to determine a more precise CV75. Please note that only 24-well plates can be used for CD86/CD54 expression measurement.
21. The solvent/vehicle control is prepared as described in paragraph 16. The positive control used in the h-CLAT method is DNCB (see paragraph 11), for which stock solutions are prepared in DMSO and diluted as described for the stock solutions in paragraph 20. DNCB should be used as the positive control for CD86/CD54 expression measurement at a final single concentration in the plate (typically 4,0 µg/ml). To obtain a 4,0 µg/ml concentration of DNCB in the plate, a 2 mg/ml stock solution of DNCB in DMSO is prepared and further diluted 250-fold with culture medium to a 8 µg/ml working solution. Alternatively, the CV75 of DNCB, which is determined in each test facility, could be also used as the positive control concentration. Other suitable positive controls may be used if historical data are available to derive comparable run acceptance criteria. For positive controls, the final single concentration in the plate should not exceed 5 000 µg/ml (in case of saline or medium) or 1 000 µg/ml (in case of DMSO). The run acceptance criteria are the same as those described for the test chemical (see paragraph 29), except for the last acceptance criterion since the positive control is tested at a single concentration.

Application of test chemicals and control substances

22. For each test chemical and control substance, one experiment is needed to obtain a prediction. Each experiment consists of at least two independent runs for CD86/CD54 expression measurement (see paragraphs 26-28). Each independent run is performed on a different day or on the same day provided that for each run: a) independent fresh stock solutions and working solutions of the test chemical and antibody solutions are prepared and b) independently harvested cells are used (i.e. cells are collected from different culture flasks); however, cells may come from the same passage. Test chemicals and control substances prepared as working solutions (500 µl) are mixed with 500 µl of suspended cells (1×10^6 cells) at 1:1 ratio, and cells are incubated for 24 ± 0.5 hours as described in paragraphs 20 and 21. In each run, a single replicate for each concentration of the test chemical and control substance is sufficient because a prediction is obtained from at least two independent runs.

Cell staining and analysis

23. After 24 ± 0.5 hours of exposure, cells are transferred from 24 well plate into sample tubes, collected by centrifugation and then washed twice with 1ml of staining buffer (if necessary, additional washing steps may be done). After washing, cells are blocked with 600 µl of blocking solution (staining buffer containing 0,01 % (w/v) globulin (Cohn fraction II, III, human; SIGMA, #G2388-10G or equivalent)) and incubated at 4°C for 15 min. After blocking, cells are split in three aliquots of 180 µl into a 96-well round-bottom plate or micro tube.
24. After centrifugation, cells are stained with 50 µl of FITC-labelled anti-CD86, anti-CD54 or mouse IgG1 (isotype) antibodies at 4°C for 30 min. The antibodies described in the h-CLAT DB-ALM protocol no 158 (18) should be used by diluting 3:25 v/v (for CD86 (BD-PharMingen, #555657; Clone: Fun-1)) or 3:50 v/v (for CD54 (DAKO, #F7143; Clone: 6.5B5) and IgG1 (DAKO, #X0927)) with staining buffer. These antibody dilution factors were defined by the

test developers as those providing the best signal-to-noise ratio. Based on the experience of the test developers, the fluorescence intensity of the antibodies is usually consistent between different lots. However, users may consider titrating the antibodies in their own laboratory's conditions to define the best concentrations for use. Other fluorochrome-tagged anti-CD86 and/or anti-CD54 antibodies may be used if they can be shown to provide similar results as FITC-conjugated antibodies, for example by testing the proficiency substances in Appendix 1.2. It should be noted that changing the clone or supplier of the antibodies as described in the h-CLAT DB-ALM protocol no 158 (18) may affect the results. After washing twice or more with 150 µl of staining buffer, cells are resuspended in staining buffer (e.g. 400 µl), and the PI solution (e.g. 20 µl to obtain a final concentration of 0,625 µg/ml) or another cytotoxicity marker's solution (see paragraph 18) is added. The expression levels of CD86 and CD54, and cell viability are analysed using flow cytometry.

DATA AND REPORTING

Data evaluation

25. The expression of CD86 and CD54 is analysed with flow cytometry with the acquisition channel FL-1. Based on the geometric mean fluorescence intensity (MFI), the relative fluorescence intensity (RFI) of CD86 and CD54 for positive control (ctrl) cells and chemical-treated cells are calculated according to the following equation:

$$RFI = \frac{MFI \text{ of chemical-treated cell} - MFI \text{ of chemical-treated isotype control cells}}{MFI \text{ of solvent/vehicle-treated ctrl cells} - MFI \text{ of solvent/vehicle-treated isotype ctrl cells}} \times 100$$

The cell viability from the isotype control (ctrl) cells (which are stained with mouse IgG1 (isotype) antibodies) is also calculated according to the equation described in paragraph 19.

Prediction model

26. For CD86/CD54 expression measurement, each test chemical is tested in at least two independent runs to derive a single prediction (POSITIVE or NEGATIVE). An h-CLAT prediction is considered POSITIVE if at least one of the following conditions is met in 2 of 2 or in at least 2 of 3 independent runs, otherwise the h-CLAT prediction is considered NEGATIVE (Figure 1):

- The RFI of CD86 is equal to or greater than 150 % at any tested concentration (with cell viability \geq 50 %);
- The RFI of CD54 is equal to or greater than 200 % at any tested concentration (with cell viability \geq 50 %).

27. Based on the above, if the first two runs are both positive for CD86 and/or are both positive for CD54, the h-CLAT prediction is considered POSITIVE and a third run does not need to be conducted. Similarly, if the first two runs are negative for both markers, the h-CLAT prediction is considered NEGATIVE (with due consideration of the provisions of paragraph 30) without the need for a third run. If however, the first two runs are not concordant for at least one of the markers (CD54 or CD86), a third run is needed and the final prediction will be based on the majority result of the three individual runs (i.e. 2 out of 3). In this respect, it should be noted that if two independent runs are conducted and one is only positive for CD86 (hereinafter referred to as P₁) and the other is only positive for CD54 (hereinafter referred to as P₂), a third run is required. If this third run is negative for both markers (hereinafter referred to as N), the h-CLAT prediction is considered NEGATIVE. On the other hand, if the third run is positive for either marker (P₁ or P₂) or for both markers (hereinafter referred to as P₁₂), the h-CLAT prediction is considered POSITIVE.

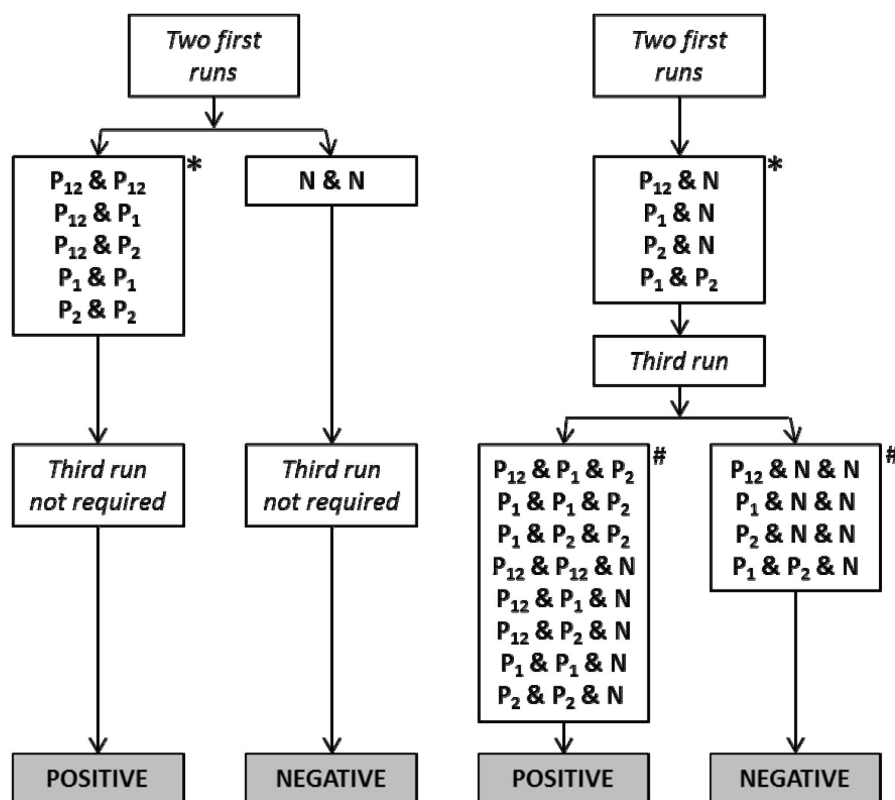


Figure 1: Prediction model used in the h-CLAT method. An h-CLAT prediction should be considered in the framework of an IATA and in accordance with the provision of paragraphs 7 and 8 in the General introduction.

P₁: run with only CD86 positive; P₂: run with only CD54 positive; P₁₂: run with both CD86 and CD54 positive; N: run with neither CD86 nor CD54 positive.

* The boxes show the relevant combinations of results from the first two runs, independently of the order in which they may be obtained.

The boxes show the relevant combinations of results from the three runs on the basis of the results obtained in the first two runs shown in the box above, but do not reflect the order in which they may be obtained.

28. For the test chemicals predicted as POSITIVE with the h-CLAT, optionally, two Effective Concentrations (EC) values, the EC150 for CD86 and EC200 for CD54, i.e. the concentration at which the test chemicals induced a RFI of 150 or 200, may be determined. These EC values potentially could contribute to the assessment of sensitising potency (9) when used in integrated approaches such as IATA (4) (5) (6) (7) (8). They can be calculated by the following equations:

$$EC\ 150\ (for\ CD86) = B_{conc} + [(150 - B_{RFI}) / (A_{RFI} - B_{RFI})] \times (A_{conc} - B_{conc})$$

$$EC\ 200\ (for\ CD86) = B_{conc} + [(200 - B_{RFI}) / (A_{RFI} - B_{RFI})] \times (A_{conc} - B_{conc})$$

where

A_{conc} is the lowest concentration in µg/ml with RFI > 150 (CD86) or 200 (CD54)

B_{conc} is the highest concentration in µg/ml with RFI < 150 (CD86) or 200 (CD54)

A_{RFI} is the RFI at the lowest concentration with RFI > 150 (CD86) or 200 (CD54)

B_{RFI} is the RFI at the highest concentration with RFI < 150 (CD86) or 200 (CD54)

For the purpose of more precisely deriving the EC150 and EC200 values, three independent runs for CD86/CD54 expression measurement may be required. The final EC150 and EC200 values are then determined as the median value of the ECs calculated from the three independent runs. When only two of three independent runs meet the criteria for positivity (see paragraphs 26-27), the higher EC150 or EC200 of the two calculated values is adopted.

Acceptance criteria

29. The following acceptance criteria should be met when using the h-CLAT method (22) (27).

- The cell viabilities of medium and solvent/vehicle controls should be higher than 90 %.
- In the solvent/vehicle control, RFI values of both CD86 and CD54 should not exceed the positive criteria (CD86 RFI 150 % and CD54 RFI 200 %). RFI values of the solvent/vehicle control are calculated by using the formula described in paragraph 25 ("MFI of chemical" should be replaced with "MFI of solvent/vehicle", and "MFI of solvent/vehicle" should be replaced with "MFI of (medium) control").
- For both medium and solvent/vehicle controls, the MFI ratio of both CD86 and CD54 to isotype control should be > 105 %.
- In the positive control (DNCB), RFI values of both CD86 and CD54 should meet the positive criteria (CD86 RFI 150 and CD54 RFI 200) and cell viability should be more than 50 %.
- For the test chemical, the cell viability should be more than 50 % in at least four tested concentrations in each run.

30. Negative results are acceptable only for test chemicals exhibiting a cell viability of less than 90 % at the highest concentration tested (i.e. $1,2 \times CV75$ according to the serial dilution scheme described in paragraph 20). If the cell viability at $1,2 \times CV75$ is equal or above 90 % the negative result should be discarded. In such a case it is recommended to try to refine the dose selection by repeating the CV75 determination. It should be noted that when 5 000 µg/ml in saline (or medium or other solvents/vehicles), 1 000 µg/ml in DMSO or the highest soluble concentration is used as the maximal test concentration of a test chemical, a negative result is acceptable even if the cell viability is above 90 %.

Test report

31. The test report should include the following information.

Test chemical

Mono-constituent substance

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Physical appearance, Log Kow, water solubility, DMSO solubility, molecular weight, and additional relevant physicochemical properties, to the extent available;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc.;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Concentration(s) tested;
- Storage conditions and stability to the extent available;
- Justification for choice of solvent/vehicle for each test chemical.

Multi-constituent substance, UVCB and mixture

- Characterisation as far as possible by e.g. chemical identity (see above), purity, quantitative occurrence and relevant physicochemical properties (see above) of the constituents, to the extent available;
- Physical appearance, water solubility, DMSO solubility and additional relevant physicochemical properties, to the extent available;
- Molecular weight or apparent molecular weight in case of mixtures/polymers of known compositions or other information relevant for the conduct of the study;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Concentration(s) tested;
- Storage conditions and stability to the extent available;
- Justification for choice of solvent/vehicle for each test chemical.

Controls

Positive control

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Physical appearance, Log K_{ow} , water solubility, DMSO solubility, molecular weight, and additional relevant physicochemical properties, to the extent available and where applicable;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc.;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Concentration(s) tested;
- Storage conditions and stability to the extent available;
- Reference to historical positive control results demonstrating suitable run acceptance criteria, if applicable.

Negative and solvent/vehicle control

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc.;
- Physical appearance, molecular weight, and additional relevant physicochemical properties in the case other control solvent/vehicle than those mentioned in the Test Guideline are used and to the extent available;
- Storage conditions and stability to the extent available;
- Justification for choice of solvent/vehicle for each test chemical.

Test conditions

- Name and address of the sponsor, test facility and study director;
- Description of test used;
- Cell line used, its storage conditions and source (e.g. the facility from which they were obtained);
- Flow cytometry used (e.g. model), including instrument settings, globulin, antibodies and cytotoxicity marker used;
- The procedure used to demonstrate proficiency of the laboratory in performing the test by testing of proficiency substances, and the procedure used to demonstrate reproducible performance of the test over time, e.g. historical control data and/or historical reactivity checks' data.

Test acceptance criteria

- Cell viability, MFI and RFI values obtained with the solvent/vehicle control in comparison to the acceptance ranges;
- Cell viability and RFI values obtained with the positive control in comparison to the acceptance ranges;
- Cell viability of all tested concentrations of the tested chemical.

Test procedure

- Number of runs used;
- Test chemical concentrations, application and exposure time used (if different than the one recommended)
- Description of evaluation and decision criteria used;
- Description of any modifications of the test procedure.

Results

- Tabulation of the data, including CV75 (if applicable), individual geometric MFI, RFI, cell viability values, EC150/EC200 values (if applicable) obtained for the test chemical and for the positive control in each run, and an indication of the rating of the test chemical according to the prediction model;
- Description of any other relevant observations, if applicable.

Discussion of the results

- Discussion of the results obtained with the h-CLAT method;
- Consideration of the test results within the context of an IATA, if other relevant information is available.

Conclusions

LITERATURE

- (1) Ashikaga T, Yoshida Y, Hirota M, Yoneyama K, Itagaki H, Sakaguchi H, Miyazawa M, Ito Y, Suzuki H, Toyoda H. (2006). Development of an *in vitro* skin sensitization test using human cell lines: The human Cell Line Activation Test (h-CLAT) I. Optimization of the h-CLAT protocol. *Toxicol. In Vitro* 20, 767–773.
- (2) Miyazawa M, Ito Y, Yoshida Y, Sakaguchi H, Suzuki H. (2007). Phenotypic alterations and cytokine production in THP-1 cells in response to allergens. *Toxicol. In Vitro* 21, 428-437.
- (3) EC EURL-ECVAM (2013). Recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitisation testing. Accessible at: <https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations>
- (4) Takenouchi O, Fukui S, Okamoto K, Kurotani S, Imai N, Fujishiro M, Kyotani D, Kato Y, Kasahara T, Fujita M, Toyoda A, Sekiya D, Watanabe S, Seto H, Hirota M, Ashikaga T, Miyazawa M. (2015). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *J Appl Toxicol.* 35, 1318-1332.
- (5) Hirota M, Fukui S, Okamoto K, Kurotani S, Imai N, Fujishiro M, Kyotani D, Kato Y, Kasahara T, Fujita M, Toyoda A, Sekiya D, Watanabe S, Seto H, Takenouchi O, Ashikaga T, Miyazawa M. (2015). Evaluation of combinations of *in vitro* sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. *J Appl Toxicol.* 35, 1333-1347.
- (6) Bauch C, Kolle SN, Ramirez T, Fabian E, Mehling A, Teubner W, van Ravenzwaay B, Landsiedel R. (2012). Putting the parts together: combining *in vitro* methods to test for skin sensitizing potentials. *Regul Toxicol Pharmacol.* 63, 489-504.
- (7) Van der Veen JW, Rorije E, Emter R, Natch A, van Loveren H, Ezendam J. (2014). Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. *Regul Toxicol Pharmacol.* 69, 371-379.
- (8) Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul Toxicol Pharmacol.* 71, 337-351.
- (9) Jaworska JS, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M. (2015). Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. *Arch Toxicol.* 89, 2355-2383.

- (10) Strickland J, Zang Q, Kleinstreuer N, Paris M, Lehmann DM, Choksi N, Matheson J, Jacobs A, Lowit A, Allen D, Casey W. (2016). Integrated decision strategies for skin sensitization hazard. *J Appl Toxicol*. DOI 10.1002/jat.3281.
- (11) Nukada Y, Ashikaga T, Miyazawa M, Hirota M, Sakaguchi H, Sasa H, Nishiyama N. (2012). Prediction of skin sensitization potency of chemicals by human Cell Line Activation Test (h-CLAT) and an attempt at classifying skin sensitization potency. *Toxicol. In Vitro* 26, 1150-60.
- (12) EC EURL ECVAM (2015). Re-analysis of the within and between laboratory reproducibility of the human Cell Line Activation Test (h-CLAT). Accessible at: <https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/eurl-ecvam-recommendation-on-the-human-cell-line-activation-test-h-clat-for-skin-sensitisation-testing>
- (13) EC EURL ECVAM (2012). human Cell Line Activation Test (h-CLAT) Validation Study Report Accessible at: <https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations>
- (14) Takenouchi O, Miyazawa M, Saito K, Ashikaga T, Sakaguchi H. (2013). Predictive performance of the human Cell Line Activation Test (h-CLAT) for lipophilic with high octanol-water partition coefficients. *J. Toxicol. Sci.* 38, 599-609.
- (15) Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y, Miyazawa M, Ito Y, Nishiyama N, Itagaki H. (2010). A comparative evaluation of *in vitro* skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). *Altern. Lab. Anim.* 38, 275-284.
- (16) Fabian E., Vogel D., Blatz V., Ramirez T., Kolle S., Eltze T., van Ravenzwaay B., Oesch F., Landsiedel R. (2013). Xenobiotic metabolizing enzyme activities in cells used for testing skin sensitization *in vitro*. *Arch Toxicol* 87, 1683-1969.
- (17) Okamoto K, Kato Y, Kosaka N, Mizuno M, Inaba H, Sono S, Ashikaga T, Nakamura T, Okamoto Y, Sakaguchi H, Kishi M, Kuwahara H, Ohno Y. (2010). The Japanese ring study of a human Cell Line Activation Test (h-CLAT) for predicting skin sensitization potential (6th report): A study for evaluating oxidative hair dye sensitization potential using h-CLAT. *AATEX* 15, 81-88.
- (18) DB-ALM (INVITTOX) (2014). Protocol 158: human Cell Line Activation Test (h-CLAT), 23pp. Accessible at: <http://ecvam-dbal.m.jrc.ec.europa.eu/>
- (19) Mizuno M, Yoshida M, Kodama T, Kosaka N, Okamoto K, Sono S, Yamada T, Hasegawa S, Ashikaga T, Kuwahara H, Sakaguchi H, Sato J, Ota N, Okamoto Y, Ohno Y. (2008). Effects of pre-culture conditions on the human Cell Line Activation Test (h-CLAT) results; Results of the 4th Japanese inter-laboratory study. *AATEX* 13, 70-82.

- (20) Sono S, Mizuno M, Kosaka N, Okamoto K, Kato Y, Inaba H, Nakamura T, Kishi M, Kuwahara H, Sakaguchi H, Okamoto Y, Ashikaga T, Ohno Y. (2010). The Japanese ring study of a human Cell Line Activation Test (h-CLAT) for predicting skin sensitization potential (7th report): Evaluation of volatile, poorly soluble fragrance materials. AATEX 15, 89-96.
- (21) OECD (2005). Guidance Document No 34 on The Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Series on Testing and Assessment. Organization for Economic Cooperation and Development, Paris, France, 2005, 96 pp.
- (22) OECD (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. Series on Testing and Assessment No 168. Available at: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO\(2012\)10/PART1&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2012)10/PART1&docLanguage=En)
- (23) United Nations UN (2013). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Fifth revised edition. New York & Geneva: United Nations Publications. ISBN: 978-92-1-117006-1. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html
- (24) ECETOC (2003). Contact sensitization: Classification according to potency. European Centre for Ecotoxicology & Toxicology of Chemicals (Technical Report No 87).
- (25) Ashikaga T, Sakaguchi H, Okamoto K, Mizuno M, Sato J, Yamada T, Yoshida M, Ota N, Hasegawa S, Kodama T, Okamoto Y, Kuwahara H, Kosaka N, Sono S, Ohno Y. (2008). Assessment of the human Cell Line Activation Test (h-CLAT) for Skin Sensitization; Results of the First Japanese Inter-laboratory Study. AATEX 13, 27-35.

Appendix 1.1

DEFINITIONS

Accuracy: The closeness of agreement between test results and accepted reference values. It is a measure of test performance and one aspect of relevance. The term is often used interchangeably with concordance to mean the proportion of correct outcomes of a test (21).

AOP (Adverse Outcome Pathway): sequence of events from the chemical structure of a target chemical or group of similar chemicals through the molecular initiating event to an *in vivo* outcome of interest (22).

Chemical: A substance or a mixture.

CV75: The estimated concentration showing 75 % cell viability.

EC150: the concentrations showing the RFI values of 150 in CD86 expression

EC200: the concentrations showing the RFI values of 200 in CD54 expression

Flow cytometry: a cytometric technique in which cells suspended in a fluid flow one at a time through a focus of exciting light, which is scattered in patterns characteristic to the cells and their components; cells are frequently labeled with fluorescent markers so that light is first absorbed and then emitted at altered frequencies.

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

IATA (Integrated Approach to Testing and Assessment): A structured approach used for hazard identification (potential), hazard characterisation (potency) and/or safety assessment (potential/potency and exposure) of a chemical or group of chemicals, which strategically integrates and weights all relevant data to inform regulatory decision regarding potential hazard and/or risk and/or the need for further targeted and therefore minimal testing.

Medium control: An untreated replicate containing all components of a test system. This sample is processed with test chemical-treated samples and other control samples to determine whether the solvent/vehicle interacts with the test system.

Mixture: A mixture or a solution composed of two or more substances.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration $\geq 10\%$ (w/w) and $< 80\%$ (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

Positive control: A replicate containing all components of a test system and treated with a substance known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Pre-haptens: chemicals which become sensitisers through abiotic transformation

Pro-haptens: chemicals requiring enzymatic activation to exert skin sensitisation potential

Relative fluorescence intensity (RFI): Relative values of geometric mean fluorescence intensity (MFI) in chemical-treated cells compared to MFI in solvent/vehicle-treated cells.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test (21).

Reliability: Measures of the extent that a test can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability (21).

Run: A run consists of one or more test chemicals tested concurrently with a solvent/vehicle control and with a positive control.

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test. It is a measure of accuracy for a test that produces categorical results, and is an important consideration in assessing the relevance of a test(21).

Staining buffer: A phosphate buffered saline containing 0,1 % bovine serum albumin.

Solvent/vehicle control: An untreated sample containing all components of a test system except of the test chemical, but including the solvent/vehicle that is used. It is used to establish the baseline response for the samples treated with the test chemical dissolved or stably dispersed in the same solvent/vehicle. When tested with a concurrent medium control, this sample also demonstrates whether the solvent/vehicle interacts with the test system.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test. It is a measure of accuracy for a test that produces categorical results and is an important consideration in assessing the relevance of a test (21).

Substance: A chemical element and its compounds in the natural state or obtained by any production process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Test chemical: Any substance or mixture tested using this method.

United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS): A system proposing the classification of chemicals (substances and mixtures) according to standardised types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (23).

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

Valid test: A test considered to have sufficient relevance and reliability for a specific purpose and which is based on scientifically sound principles. A test is never valid in an absolute sense, but only in relation to a defined purpose (21).

Appendix 1.2

PROFICIENCY SUBSTANCES

Prior to routine use of the test described in this Appendix to test method B.71, laboratories should demonstrate technical proficiency by correctly obtaining the expected h-CLAT prediction for the 10 substances recommended in Table 1 and by obtaining CV75, EC150 and EC200 values that fall within the respective reference range for at least 8 out of the 10 proficiency substances. Proficiency substances were selected to represent the range of responses for skin sensitisation hazards. Other selection criteria were that the substances are commercially available, and that high-quality *in vivo* reference data as well as high quality *in vitro* data generated with the h-CLAT method are available. Also, published reference data are available for the h-CLAT method (3) (14).

Table 1

Recommended substances for demonstrating technical proficiency with the h-CLAT method

Proficiency substances	CASRN	Physical state	<i>In vivo</i> prediction (1)	CV75 Reference Range in µg/ml (2)	h-CLAT results for CD86 (EC150 Reference Range in µg/ml) (2)	h-CLAT results for CD54 (EC200 Reference Range in µg/ml) (2)
2,4-Dinitrochlorobenzene	97-00-7	Solid	Sensitiser(extreme)	2-12	Positive(0.5-10)	Positive(0.5-15)
4-Phenylenediamine	106-50-3	Solid	Sensitiser(strong)	5-95	Positive(<40)	Negative(>1,5) (3)
Nickel sulfate	10101-97-0	Solid	Sensitiser(mod-erate)	30-500	Positive(<100)	Positive(10-100)
2-Mercaptbenzothiazole	149-30-4	Solid	Sensitiser(mod-erate)	30-400	Negative(>10) (3)	Positive(10-140)
R(+)-Limonene	5989-27-5	Liquid	Sensitiser(weak)	>20	Negative(>5) (3)	Positive(<250)
Imidazolidinyl urea	39236-46-9	Solid	Sensitiser(weak)	25-100	Positive(20-90)	Positive(20-75)
Isopropanol	67-63-0	Liquid	Non-sensitiser	>5000	Negative(>5000)	Negative(>5000)
Glycerol	56-81-5	Liquid	Non-sensitiser	>5000	Negative(>5000)	Negative(>5000)
Lactic acid	50-21-5	Liquid	Non-sensitiser	1500-5000	Negative(>5000)	Negative(>5000)
4-Aminobenzoic acid	150-13-0	Solid	Non-sensitiser	>1000	Negative(>1000)	Negative(>1000)

Abbreviations: CAS RN = Chemical Abstracts Service Registry Number

(1) The *in vivo* hazard (and potency) prediction is based on LLNA data (3) (14). The *in vivo* potency is derived using the criteria proposed by ECETOC (24).

(2) Based on historical observed values (13) (25).

(3) Historically, a majority of negative results have been obtained for this marker and therefore a negative result is mostly expected. The range provided was defined on the basis of the few historical positive results observed. In case a positive result is obtained, the EC value should be within the reported reference range.

Appendix 2

IN VITRO SKIN SENSITISATION: U937 CELL LINE ACTIVATION TEST (U-SENS™)

INITIAL CONSIDERATIONS AND LIMITATIONS

1. The U-SENS™ test quantifies the change in the expression of a cell surface marker associated with the process of activation of monocytes and dendritic cells (DC) (i.e. CD86), in the human histiocytic lymphoma cell line U937, following exposure to sensitisers (1). The measured expression levels of CD86 cell surface marker in the cell line U937 is then used for supporting the discrimination between skin sensitisers and non-sensitisers.
2. The U-SENS™ test has been evaluated in a validation study (2) coordinated by L'Oreal and subsequently independent peer reviewed by the European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) Scientific Advisory Committee (ESAC) (3). Considering all available evidence and input from regulators and stakeholders, the U-SENS™ was recommended by EURL ECVAM (4) to be used as part of an IATA to support the discrimination between sensitisers and non-sensitisers for the purpose of hazard classification and labelling. In its guidance document on the reporting of structured approaches to data integration and individual information sources used within IATA for skin sensitisation, the OECD currently discusses a number of case studies describing different testing strategies and prediction models. One of the different defined approaches is based on the U-SENS assay (5). Examples of the use of U-SENS™ data in combination with other information, including historical data and existing valid human data (6), are also reported elsewhere in the literature (4) (5) (7).
3. The U-SENS™ test proved to be transferable to laboratories experienced in cell culture techniques and flow cytometry analysis. The level of reproducibility in predictions that can be expected from the test is in the order of 90 % and 84 % within and between laboratories, respectively (8). Results generated in the validation study (8) and other published studies (1) overall indicate that, compared with LLNA results, the accuracy in distinguishing skin sensitisers (i.e. UN GHS/CLP Cat.1) from non-sensitisers is 86 % (N=166) with a sensitivity of 91 % (118/129) and a specificity of 65 % (24/37). Compared with human results, the accuracy in distinguishing skin sensitisers (i.e. UN GHS/CLP Cat.1) from non-sensitisers is 77 % (N=101) with a sensitivity of 100 % (58/58) and a specificity of 47 % (20/43). False negative predictions compared to LLNA with the U-SENS™ are more likely to concern chemicals showing a low to moderate skin sensitisation potency (i.e. UN GHS/CLP subcategory 1B) than chemicals showing a high skin sensitisation potency (i.e. UN GHS/CLP subcategory 1A) (1) (8) (9). Taken together, this information indicates the usefulness of the U-SENS™ test to contribute to the identification of skin sensitisation hazards. However, the accuracy values given here for the U-SENS™ as a stand-alone test are only indicative, since the test should be considered in combination with other sources of information in the context of an IATA and in accordance with the provisions of paragraphs 7 and 8 in the General introduction. Furthermore, when evaluating non-animal methods for skin sensitisation, it should be kept in mind that the LLNA test as well as other animal tests may not fully reflect the situation in humans.
4. On the basis of the data currently available, the U-SENS™ test was shown to be applicable to test chemicals (including cosmetics ingredients e.g. preservatives, surfactants, actives, dyes) covering a variety of organic functional groups, of physicochemical properties, skin sensitisation potency (as determined in *in vivo* studies) and the spectrum of reaction mechanisms known to be associated with skin sensitisation (i.e. Michael acceptor, Schiff base formation, acyl transfer agent, substitution nucleophilic bi-molecular [SN2], or nucleophilic aromatic substitution [SNAr]) (1) (8) (9) (10). The U-SENS™ test is applicable to test chemicals that are soluble or that form a stable dispersion (i.e. a colloid or suspension in which the test chemical does not settle or separate from the solvent/vehicle into different phases) in an appropriate solvent/vehicle (see paragraph 13). Chemicals in the dataset reported to be pre-haptens (i.e. substances activated by oxidation) or pro-haptens (i.e. substances requiring enzymatic activation for example via P450 enzymes) were correctly predicted by the U-SENS™ (1) (10). Membrane disrupting substances can lead to false

positive results due to a non-specific increase of CD86 expression, as 3 out of 7 false positives relative to the *in vivo* reference classification were surfactants (1). As such positive results with surfactants should be considered with caution whereas negative results with surfactants could still be used to support the identification of the test chemical as a non-sensitiser. Fluorescent test chemicals can be assessed with the U-SENS™ (1), nevertheless, strong fluorescent test chemicals emitting at the same wavelength as fluorescein isothiocyanate (FITC) or as propidium iodide (PI), will interfere with the flow cytometric detection and thus cannot be correctly evaluated using FITC-conjugated antibodies (potential false negative) or PI (viability not measurable). In such a case, other fluorochrome-tagged antibodies or other cytotoxicity markers, respectively, can be used as long as it can be shown they provide similar results as the FITC-tagged antibodies or PI (see paragraph 18) e.g. by testing the proficiency substances in Appendix 2.2. In the light of the above, positive results with surfactants and negative results with strong fluorescent test chemicals should be interpreted in the context of the stated limitations and together with other information sources within the framework of IATA. In cases where there is evidence demonstrating the non-applicability of the U-SENS™ test to other specific categories of test chemicals, it should not be used for those specific categories.

5. As described above, the U-SENS™ test supports the discrimination between skin sensitisers from non-sensitisers. However, it may also potentially contribute to the assessment of sensitising potency when used in integrated approaches such as IATA. Nevertheless, further work, preferably based on human data, is required to determine how U-SENS™ results may possibly inform potency assessment.
6. Definitions are provided in Appendix 2.1.

PRINCIPLE OF THE TEST

7. The U-SENS™ test is an *in vitro* assay that quantifies changes of CD86 cell surface marker expression on a human histiocytic lymphoma cell line, U937 cells, following 45±3 hours exposure to the test chemical. The CD86 surface marker is one typical marker of U937 activation. CD86 is known to be a co-stimulatory molecule that may mimic monocytic activation, which plays a critical role in T-cell priming. The changes of CD86 cell surface marker expression are measured by flow cytometry following cell staining typically with fluorescein isothiocyanate (FITC)-labelled antibodies. Cytotoxicity measurement is also conducted (e.g. by using PI) concurrently to assess whether upregulation of CD86 cell surface marker expression occurs at sub-cytotoxic concentrations. The stimulation index (S.I.) of CD86 cell surface marker compared to solvent/vehicle control is calculated and used in the prediction model (see paragraph 19), to support the discrimination between sensitisers and non-sensitisers.

DEMONSTRATION OF PROFICIENCY

8. Prior to routine use of the test described in this Appendix to test method B.71, laboratories should demonstrate technical proficiency, using the 10 Proficiency Substances listed in Appendix 2.2 in compliance with the Good *in vitro* Method Practices (11). Moreover, test users should maintain a historical database of data generated with the reactivity checks (see paragraph 11) and with the positive and solvent/vehicle controls (see paragraphs 15-16), and use these data to confirm the reproducibility of the test in their laboratory is maintained over time.

PROCEDURE

9. This test is based on the U-SENS™ DataBase service on ALternative Methods to animal experimentation (DB-ALM) protocol no 183 (12). The Standard Operating Procedures (SOP) should be employed when implementing and using the U-SENS™ test in the laboratory. An automated system to run the U-SENS™ can be used if it can be shown to provide similar results, for example by testing the proficiency substances in Appendix 2.2. The following is a description of the main components and procedures for the U-SENS™ test.

Preparation of cells

10. The human histiocytic lymphoma cell line, U937 (13) should be used for performing the U-SENS™ test. Cells (clone CRL1593.2) should be obtained from a well-qualified cell bank such as the American Type Culture Collection.

11. U937 cells are cultured, at 37 °C under 5 % CO₂ and humidified atmosphere, in RPMI-1 640 medium supplemented with 10 % foetal calf serum (FCS), 2 mM L-glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin (complete medium). U937 cells are routinely passaged every 2-3 days at the density of 1,5 or 3 × 10⁵ cells/ml, respectively. The cell density should not exceed 2 × 10⁶ cells/ml and the cell viability measured by trypan blue exclusion should be ≥ 90 % (not to be applied at the first passage after thawing). Prior to using them for testing, every batch of cells, FCS or antibodies should be qualified by conducting a reactivity check. The reactivity check of the cells should be performed using the positive control, picrylsulfonic acid (2,4,6-Trinitro-benzene-sulfonic acid: TNBS) (CASRN 2 508-19-2, ≥ 99 % purity) and the negative control lactic acid (LA) (CASRN 50-21-5, ≥ 85 % purity), at least one week after thawing. For the reactivity check, six final concentrations should be tested for each of the 2 controls (TNBS: 1, 12.5, 25, 50, 75, 100µg/ml and LA: 1, 10, 20, 50, 100, 200µg/ml). TNBS solubilised in complete medium should produce a positive and concentration-related response of CD86 (e.g. when a positive concentration, CD86 S.I. ≥ 150, is followed by a concentration with an increasing CD86 S.I.), and LA solubilised in complete medium should produce negative response of CD86 (see paragraph 21). Only the batch of cells which passed the reactivity check 2 times should be used for the assay. Cells can be propagated up to seven weeks after thawing. Passage number should not exceed 21. The reactivity check should be performed according to the procedures described in paragraphs 18-22.

12. For testing, U937 cells are seeded at a density of either 3 × 10⁵ cells/ml or 6 × 10⁵ cells/ml, and pre-cultured in culture flasks for 2 days or 1 day, respectively. Other pre-cultured conditions than those described above may be used if sufficient scientific rationale is provided and if it can be shown to provide similar results, for example by testing the proficiency substances in Appendix 2.2. In the day of testing, cells harvested from culture flask are resuspended with fresh culture medium at 5 × 10⁵ cells/ml. Then, cells are distributed into a 96-well flat-bottom plate with 100 µl (final cell density of 0,5 × 10⁵ cells/well).

Preparation of test chemicals and control substances

13. Assessment of solubility is conducted prior to testing. For this purpose, test chemicals are dissolved or stably dispersed at a concentration of 50 mg/ml in complete medium as first solvent option or dimethyl sulfoxide (DMSO, 99 % purity) as a second solvent/vehicle option if the test chemical is not soluble in the complete medium solvent/vehicle. For the testing, the test chemical is dissolved to a final concentration of 0,4 mg/ml in complete medium if the chemical is soluble in this solvent/vehicle. If the chemical is soluble only in DMSO, the chemical is dissolved at a concentration of 50 mg/ml. Other solvents/vehicles than those described above may be used if sufficient scientific rationale is provided. Stability of the test chemical in the final solvent/vehicle should be taken into account.

14. The test chemicals and control substances are prepared on the day of testing. Because a dose finding assay is not conducted, for the first run, 6 final concentrations should be tested (1, 10, 20, 50, 100 and 200 µg/ml) into the corresponding solvent/vehicle either in complete medium or in 0,4 % DMSO in medium. For the subsequent runs, starting from the 0,4 mg/ml in complete medium or 50 mg/ml in DMSO, solutions of the test chemicals, at least 4 working solutions (i.e. at least 4 concentrations), are prepared using the corresponding solvent/vehicle. The working solutions are finally used for treatment by adding an equal volume of U937 cell suspension (see paragraph 11 above) to the volume of working solution in the plate to achieve a further 2-fold dilution (12). The concentrations (at least 4 concentrations) for any further run are chosen based on the individual results of all previous runs (8). The usable final concentrations are 1, 2, 3, 4, 5, 7.5, 10, 12.5, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180 and 200 µg/ml. The maximum final concentration is 200 µg/ml. In the case of a CD86 positive value at 1 µg/ml is observed, then 0,1 µg/ml is evaluated in order to find the concentration of the test chemical that does not

induce CD86 above the positive threshold. For each run, the EC150 (concentration at which a chemical reaches the CD86 positive threshold of 150 %, see paragraph 19) is calculated if a CD86 positive concentration-response is observed. Where the test chemical induces a positive CD86 response not concentration related, the calculation of the EC150 might not be relevant as described in the U-SENS™ DB-ALM protocol no 183 (12). For each run, CV70 (concentration at which a chemical reaches the cytotoxicity threshold of 70 %, see paragraph 19) is calculated whenever possible (12). To investigate the concentration response effect of CD86 increase, any concentrations from the usable concentrations should be chosen evenly spread between the EC150 (or the highest CD86 negative non cytotoxic concentration) and the CV70 (or the highest concentration allowed i.e. 200 µg/ml). A minimum of 4 concentrations should be tested per run with at least 2 concentrations being common with the previous run(s), for comparison purposes.

15. The solvent/vehicle control used in the U-SENS™ test is complete medium (for test chemicals solubilised or stably dispersed in complete medium) (see paragraph 4) or 0,4 % DMSO in complete medium (for test chemicals solubilised or stably dispersed in DMSO).

16. The positive control used in the U-SENS™ test is TNBS (see paragraph 11), prepared in complete medium. TNBS should be used as the positive control for CD86 expression measurement at a final single concentration in plate (50 µg/ml) yielding > 70 % of cell viability. To obtain a 50 µg/ml concentration of TNBS in plate, a 1 M (i.e. 293 mg/ml) stock solution of TNBS in complete medium is prepared and further diluted 2 930-fold with complete medium to a 100 µg/ml working solution. Lactic acid (LA, CAS 50-21-5) should be used as the negative control at 200 µg/ml solubilised in complete medium (from a 0,4 mg/ml stock solution). In each plate of each run, three replicates of complete medium untreated control, solvent/vehicle control, negative and positive controls are prepared (12). Other suitable positive controls may be used if historical data are available to derive comparable run acceptance criteria. The run acceptance criteria are the same as described for the test chemical (see paragraph 12).

Application of test chemicals and control substances

17. The solvent/vehicle control or working solutions described in paragraphs 14-16 are mixed 1:1 (v/v) with the cell suspensions prepared in the 96-well flat-bottom plate (see paragraph 12). The treated plates are then incubated for 45±3 hours at 37 °C under 5 % CO₂. Prior to incubation, plates are sealed with semi permeable membrane, to avoid evaporation of volatile test chemicals and cross-contamination between cells treated with test chemicals (12).

Cell staining

18. After 45±3 hours of exposure, cells are transferred into V-shaped microtiter plate and collected by centrifugation. Solubility interference is defined as crystals or drops observed under the microscope at 45 ± 3 hours post treatment (before the cell staining). The supernatants are discarded and the remaining cells are washed once with 100 µl of an ice-cold phosphate buffered saline (PBS) containing 5 % foetal calf serum (staining buffer). After centrifugation, cells are re-suspended with 100 µl of staining buffer and stained with 5 µl (e.g. 0,25 µg) of FITC-labelled anti-CD86 or mouse IgG1 (isotype) antibodies at 4 °C for 30 min protected from light. The antibodies described in the U-SENS™ DB-ALM protocol no 183 (12) should be used (for CD86: BD-PharMingen #5556 57 Clone: Fun-1, or Caltag/Invitrogen # MHCD8601 Clone: BU63; and for IgG1: BD-PharMingen #5557 48, or Caltag/Invitrogen # GM4992). Based on the experience of the test developers, the fluorescence intensity of the antibodies is usually consistent between different lots. Other clones or supplier of the antibodies which passed the reactivity check may be used for the assay (see paragraph 11). However, users may consider titrating the antibodies in their own laboratory's

conditions to define the best concentration for use. Other detection system e.g. fluorochrome-tagged anti-CD86 antibodies may be used if they can be shown to provide similar results as FITC-conjugated antibodies, for example by testing the proficiency substances in Appendix 2.2. After washing with 100 µl of staining buffer two times and once with 100 µl of an ice-cold PBS, cells are resuspended in ice-cold PBS (e.g. 125 µl for samples being analysed manually tube by tube, or 50 µl using an auto-sampler plate) and PI solution is added (final concentration of 3 µg/ml). Other cytotoxicity markers, such as 7-Aminoactinomycin D (7-AAD) or Trypan blue may be used if the alternative stains can be shown to provide similar results as PI, for example by testing the proficiency substances in Appendix 2.2.

Flow cytometry analysis

19. Expression level of CD86 and cell viability are analysed using flow cytometry. Cells are displayed within a size (FSC) and granularity (SSC) dot plot set to log scale in order to clearly identify the population in a first gate R1 and eliminate the debris. A targeting total of 10 000 cells in gate R1 are acquired for each well. Cells from the same R1 gate are displayed within a FL3 or FL4 / SSC dot plot. Viable cells are delineated by placing a second gate R2 selecting the population of propidium iodide-negative cells (FL3 or FL4 channel). The cell viability can be calculated using the following equation by the cytometer analysis program. When the cell viability is low, up to 20 000 cells including dead cells could be acquired. Alternatively, data can be acquired for one minute after the initiation of the analysis.

$$\text{Cell viability} = \frac{\text{Number of living cells}}{\text{Total Number of aquired cells}} \times 100$$

Percentage of FL1-positive cells is then measured among these viable cells gated on R2 (within R1). Cell surface expression of CD86 is analysed in a FL1 / SSC dot plot gated on viable cells (R2).

For the complete medium / IgG1 wells, the analysis marker is set close to the main population so that the complete medium controls have IgG1 within the target zone of 0,6 to 0,9 %.

Colour interference is defined as a shift of the FITC-labelled IgG1 dot-plot (IgG1 FL1 Geo Mean S.I. \geq 150 %).

The stimulation index (S.I.) of CD86 for controls cells (untreated or in 0,4 % DMSO) and chemical-treated cells are calculated according to the following equation:

$$\text{S.I.} = \frac{\% \text{ of CD86}^+ \text{ treated cells} - \% \text{ of IgG1}^+ \text{ treated cells}}{\% \text{ of CD86}^+ \text{ control cells} - \% \text{ of IgG1}^+ \text{ control cells}} \times 100$$

% of IgG1⁺ untreated control cells: referred to as percentage of FL1-positive IgG1 cells defined with the analysis marker (accepted range of \geq 0,6 % and $<$ 1,5 %, see paragraph 22) among the viable untreated cells.

% of IgG1⁺/CD86⁺ control/treated cells: referred to as percentage of FL1-positive IgG1/CD86 cells measured without moving the analysis marker among the viable control/treated cells.

DATA AND REPORTING

Data evaluation

20. The following parameters are calculated in the U-SENS™ test: CV70 value, i.e. a concentration showing 70 % of U937 cell survival (30 % cytotoxicity) and the EC150 value, i.e. the concentration at which the test chemicals induced a CD86 stimulation index (S.I.) of 150 %.

CV70 is calculated by log-linear interpolation using the following equation:

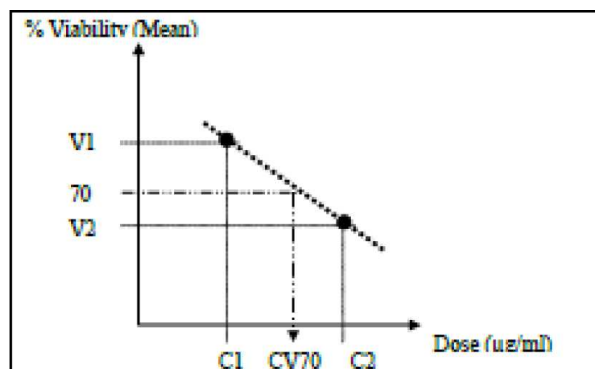
$$CV70 = C1 + [(V1 - 70) / (V1 - V2) * (C2 - C1)]$$

Where:

V1 is the minimum value of cell viability over 70 %

V2 is the maximum value of cell viability below 70 %

C1 and C2 are the concentrations showing the value of cell viability V1 and V2 respectively.



Other approaches to derive the CV70 can be used as long as it is demonstrated that this has no impact on the results (e.g. by testing the proficiency substances).

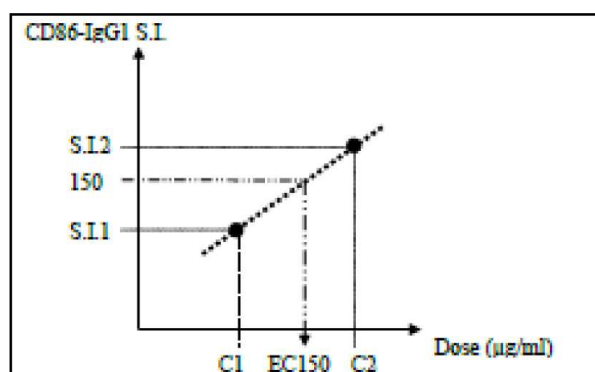
EC150 is calculated by log-linear interpolation using the following equation:

$$EC150 = C1 + [(150 - S.I.1) / (S.I.2 - S.I.1) * (C2 - C1)]$$

Where:

C1 is the highest concentration in µg/ml with a CD86 S.I. < 150 % (S.I. 1)

C2 is the lowest concentration in µg/ml with a CD86 S.I. ≥ 150 % (S.I. 2).



The EC150 and CV70 values are calculated

- for each run: the individual EC150 and CV70 values are used as tools to investigate the concentration response effect of CD86 increase (see paragraph 14),
- based on the average viabilities, the overall CV70 is determined (12),
- based on the average S.I. of CD86 values, the overall EC150 is determined for the test chemical predicted as POSITIVE with the U-SENS™ (see paragraph 21) (12).

Prediction model

21. For CD86 expression measurement, each test chemical is tested in at least four concentrations and in at least two independent runs (performed on a different day) to derive a single prediction (NEGATIVE or POSITIVE).
 - The individual conclusion of an U-SENS™ run is considered Negative (hereinafter referred to as N) if the S.I. of CD86 is less than 150 % at all non-cytotoxic concentrations (cell viability ≥ 70 %) and if no interference is observed (cytotoxicity, solubility: see paragraph 18 or colour: see paragraph 19 regardless of the non-cytotoxic concentrations at which the interference is detected). In all other cases: S.I. of CD86 higher or equal to 150 % and/or interferences observed, the individual conclusion of an U-SENS™ run is considered Positive (hereinafter referred to as P).
 - An U-SENS™ prediction is considered NEGATIVE if at least two independent runs are negative (N) (Figure 1). If the first two runs are both negative (N), the U-SENS™ prediction is considered NEGATIVE and a third run does not need to be conducted.
 - An U-SENS™ prediction is considered POSITIVE if at least two independent runs are positive (P) (Figure 1). If the first two runs are both positive (P), the U-SENS™ prediction is considered POSITIVE and a third run does not need to be conducted.
 - Because a dose finding assay is not conducted, there is an exception if, in the first run, the S.I. of CD86 is higher or equal to 150 % at the highest non-cytotoxic concentration only. The run is then considered to be NOT CONCLUSIVE (NC), and additional concentrations (between the highest non cytotoxicity concentration and the lowest cytotoxicity concentration - see paragraph 20) should be tested in additional runs. In case a run is identified as NC, at least 2 additional runs should be conducted, and a fourth run in case runs 2 and 3 are not concordant (N and/or P independently) (Figure 1). Follow up runs will be considered positive even if only one non cytotoxic concentration gives a CD86 equal or above 150 %, since the concentration setting has been adjusted for the specific test chemical. The final prediction will be based on the majority result of the three or four individual runs (i.e. 2 out of 3 or 2 out of 4) (Figure 1).

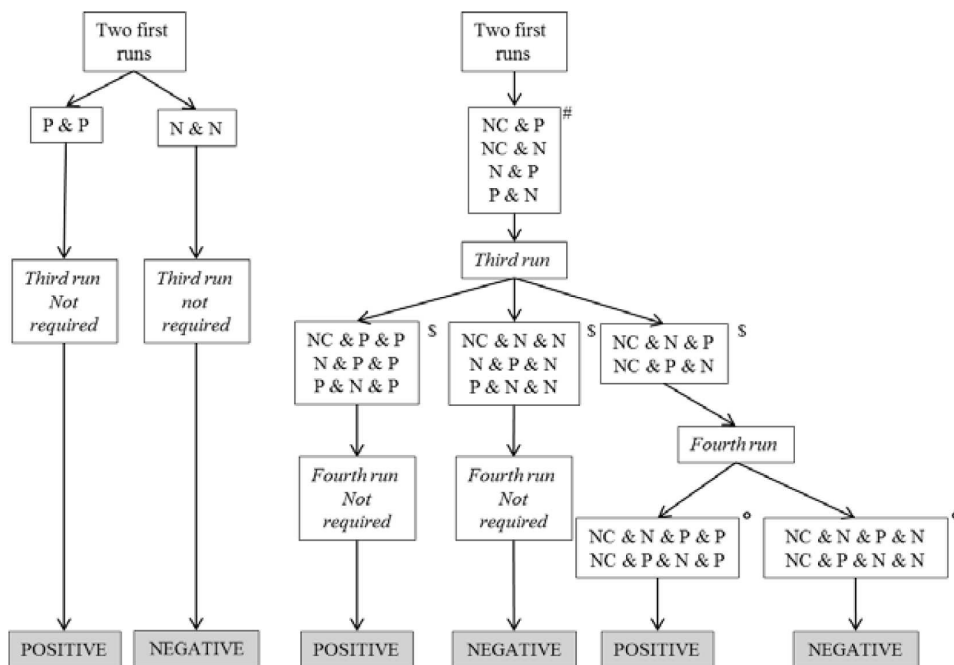


Figure 1: Prediction model used in the U-SENS™ test. An U-SENS™ prediction should be considered in the framework of an IATA and in accordance with the provision of paragraph 4 and of the General introduction paragraphs 7, 8 and 9.

N: Run with no CD86 positive or interference observed;

P: Run with CD86 positive and/or interference(s) observed;

NC: Not Conclusive. First run with No Conclusion when CD86 is positive at the highest non-cytotoxic concentration only;

#: A Not Conclusive (NC) individual conclusion attributed only to the first run conducts automatically to the need of a third run to reach a majority of Positive (P) or Negative (N) conclusions in at least 2 of 3 independent runs.

\$. The boxes show the relevant combinations of results from the three runs on the basis of the results obtained in the first two runs shown in the box above.

°. The boxes show the relevant combinations of results from the four runs on the basis of the results obtained in the first three runs shown in the box above.

Acceptance criteria

22. The following acceptance criteria should be met when using the U-SENS™ test (12).

- At the end of the 45 ± 3 hours exposure period, the mean viability of the triplicate untreated U937 cells had to be $> 90\%$ and no drift in CD86 expression is observed. The CD86 basal expression of untreated U937 cells had to be comprised within the range of $\geq 2\%$ and $\leq 25\%$.
- When DMSO is used as a solvent, the validity of the DMSO vehicle control is assessed by calculating a DMSO S.I. compared to untreated cells, and the mean viability of the triplicate cells had to be $> 90\%$. The DMSO vehicle control is valid if the mean value of its triplicate CD86 S.I. was smaller than 250 % of the mean of the triplicate CD86 S.I. of untreated U937 cells.
- The runs are considered valid if at least two out of three IgG1 values of untreated U937 cells fell within the range of $\geq 0,6\%$ and $< 1,5\%$.
- The concurrent tested negative control (lactic acid) is considered valid if at least two out of the three replicates were negative (CD86 S.I. $< 150\%$) and non-cytotoxic (cell viability $\geq 70\%$).
- The positive control (TNBS) was considered as valid if at least two out of the three replicates were positive (CD86 S.I. $\geq 150\%$) and non-cytotoxic (cell viability $\geq 70\%$).

Test report

23. The test report should include the following information.

Test Chemical

Mono-constituent substance

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;

- Physical appearance, complete medium solubility, DMSO solubility, molecular weight, and additional relevant physicochemical properties, to the extent available;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc.;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Concentration(s) tested;
- Storage conditions and stability to the extent available;
- Justification for choice of solvent/vehicle for each test chemical.

Multi-constituent substance, UVCB and mixture:

- Characterisation as far as possible by e.g. chemical identity (see above), purity, quantitative occurrence and relevant physicochemical properties (see above) of the constituents, to the extent available;
- Physical appearance, complete medium solubility, DMSO solubility and additional relevant physicochemical properties, to the extent available;
- Molecular weight or apparent molecular weight in case of mixtures/polymers of known compositions or other information relevant for the conduct of the study;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Concentration(s) tested;
- Storage conditions and stability to the extent available;
- Justification for choice of solvent/vehicle for each test chemical.

Controls

Positive control

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Physical appearance, DMSO solubility, molecular weight, and additional relevant physicochemical properties, to the extent available and where applicable;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc.;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Concentration(s) tested;
- Storage conditions and stability to the extent available;

- Reference to historical positive control results demonstrating suitable run acceptance criteria, if applicable.

Negative and solvent/vehicle control

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc.;
- Physical appearance, molecular weight, and additional relevant physicochemical properties in the case other control solvent/vehicle than those mentioned in the Test Guideline are used and to the extent available;
- Storage conditions and stability to the extent available;
- Justification for choice of solvent/vehicle for each test chemical.

Test Conditions

- Name and address of the sponsor, test facility and study director;
- Description of test used;
- Cell line used, its storage conditions and source (e.g. the facility from which they were obtained);
- Flow cytometry used (e.g. model), including instrument settings, antibodies and cytotoxicity marker used;
- The procedure used to demonstrate proficiency of the laboratory in performing the test by testing of proficiency substances, and the procedure used to demonstrate reproducible performance of the test over time, e.g. historical control data and/or historical reactivity checks' data.

Test Acceptance Criteria

- Cell viability and CD86 S.I values obtained with the solvent/vehicle control in comparison to the acceptance ranges;
- Cell viability and S.I. values obtained with the positive control in comparison to the acceptance ranges;
- Cell viability of all tested concentrations of the tested chemical.

Test procedure

- Number of runs used;
- Test chemical concentrations, application and exposure time used (if different than the one recommended)
- Duration of exposure;

- Description of evaluation and decision criteria used;
- Description of any modifications of the test procedure.

Results

- Tabulation of the data, including CV70 (if applicable), S.I., cell viability values, EC150 values (if applicable) obtained for the test chemical and for the positive control in each run, and an indication of the rating of the test chemical according to the prediction model;
- Description of any other relevant observations, if applicable.

Discussion of the Results

- Discussion of the results obtained with the U-SENS™ test;
- Consideration of the test results within the context of an IATA, if other relevant information is available.

Conclusions

LITERATURE

- (1) Piroird, C., Ovigne, J.M., Rousset, F., Martinozzi-Teissier, S., Gomes, C., Cotovio, J., Alépée, N. (2015). The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. *Toxicol. In Vitro* 29, 901-916.
- (2) EURL ECVAM (2017). The U-SENS™ test method Validation Study Report. Accessible at: http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations
- (3) EC EURL ECVAM (2016). ESAC Opinion No 2016-03 on the L'Oréal-coordinated study on the transferability and reliability of the U-SENS™ test method for skin sensitisation testing. EUR 28 178 EN; doi 10.2787/8157 37. Available at: [<http://publications.jrc.ec.europa.eu/repository/handle/JRC103705>].
- (4) EC EURL ECVAM (2017). EURL ECVAM Recommendation on the use of non-animal approaches for skin sensitisation testing. EUR 28 553 EN; doi 10.2760/5889 55. Available at: <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/eurl-ecvam-recommendation-use-non-animal-approaches-skin-sensitisation-testing>.
- (5) Steiling, W. (2016). Safety Evaluation of Cosmetic Ingredients Regarding their Skin Sensitization Potential. doi:10.3390/cosmetics3020014. *Cosmetics* 3, 14.
- (6) OECD (2016). Guidance Document on The Reporting of Defined Approaches and Individual Information Sources to be Used Within Integrated Approaches to Testing and Assessment (IATA) For Skin Sensitisation, Series on Testing & Assessment No 256, ENV/JM/MONO(2016)29. Organisation for Economic Cooperation and Development, Paris. Available at: [<http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm>].

- (7) Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., Kolle, S., Landsiedel, R., Jaworska, J., Kern, P.S., Gerberick, F., Natsch, A., Emter, R., Ashikaga, T., Miyazawa, M., Sakaguchi, H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul. Toxicol. Pharmacol.* 71, 337-351.
- (8) Alépée, N., Piroird, C., Aujoulat, M., Dreyfuss, S., Hoffmann, S., Hohenstein, A., Meloni, M., Nardelli, L., Gerbeix, C., Cotovio, J. (2015). Prospective multicentre study of the U-SENS test method for skin sensitization testing. *Toxicol In Vitro* 30, 373-382.
- (9) Reisinger, K., Hoffmann, S., Alépée, N., Ashikaga, T., Barroso, J., Elcombe, C., Gellatly, N., Galbiati, V., Gibbs, S., Groux, H., Hibatallah, J., Keller, D., Kern, P., Klaric, M., Kolle, S., Kuehnl, J., Lambrechts, N., Lindstedt, M., Millet, M., Martinozzi-Teissier, S., Natsch, A., Petersohn, D., Pike, I., Sakaguchi, H., Schepky, A., Tailhardat, M., Templier, M., van Vliet, E., Maxwell, G. (2014). Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. *Toxicol. In Vitro* 29, 259-270.
- (10) Fabian, E., Vogel, D., Blatz, V., Ramirez, T., Kolle, S., Eltze, T., van Ravenzwaay, B., Oesch, F., Landsiedel, R. (2013). Xenobiotic metabolizing enzyme activities in cells used for testing skin sensitization *in vitro*. *Arch. Toxicol.* 87, 1 683-1 696.
- (11) OECD. (2018). Draft Guidance document: Good *In Vitro* Method Practices (GIVIMP) for the Development and Implementation of *In Vitro* Methods for Regulatory Use in Human Safety Assessment. Organisation for Economic Cooperation and Development, Paris. Available at: [http://www.oecd.org/env/ehs/testing/OECD Final Draft GIVIMP.pdf](http://www.oecd.org/env/ehs/testing/OECD_Final_Draft_GIVIMP.pdf).
- (12) DB-ALM (2016). Protocol no 183: Myeloid U937 Skin Sensitization Test (U-SENS™), 33pp. Accessible at: [<http://ecvam-dbalm.jrc.ec.europa.eu/>].
- (13) Sundström, C., Nilsson, K. (1976). Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int. J. Cancer* 17, 565-577.
- (14) OECD (2005). Series on Testing and Assessment No. 34: Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Organisation for Economic Cooperation and Development, Paris. Available at: <http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm>.
- (15) United Nations UN (2015). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). ST/SG/AC.10/30/Rev.6, Sixth Revised Edition, New York & Geneva: United Nations Publications. Available at: http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev06/English/ST-SG-AC10-30-Rev6e.pdf.
- (16) OECD (2012). Series on Testing and Assessment No 168: The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. Organisation for Economic Cooperation and Development, Paris. Available at: <http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm>.
- (17) ECETOC (2003). Technical Report No 87: Contact sensitization: Classification according to potency. European Centre for Ecotoxicology & Toxicology of Chemicals, Brussels. Available at: https://ftp.cdc.gov/pub/Documents/OEL/06.%20Dotson/References/ECETOC_2003-TR87.pdf.

Appendix 2.1

DEFINITIONS

Accuracy: The closeness of agreement between test results and accepted reference values. It is a measure of test performance and one aspect of relevance. The term is often used interchangeably with concordance to mean the proportion of correct outcomes of a test (14).

AOP (Adverse Outcome Pathway): sequence of events from the chemical structure of a target chemical or group of similar chemicals through the molecular initiating event to an *in vivo* outcome of interest (15).

CD86 Concentration response: There is concentration-dependency (or concentration response) when a positive concentration (CD86 S.I. ≥ 150) is followed by a concentration with an increasing CD86 S.I.

Chemical: A substance or a mixture.

CV70: The estimated concentration showing 70 % cell viability.

Drift: A drift is defined by i) the corrected %CD86⁺ value of the untreated control replicate 3 is less than 50 % of the mean of the corrected %CD86⁺ value of untreated control replicates 1 and 2; and ii) the corrected %CD86⁺ value of the negative control replicate 3 is less than 50 % of mean of the corrected %CD86⁺ value of negative control replicates 1 and 2.

EC150: the estimated concentrations showing the 150 % S.I. of CD86 expression.

Flow cytometry: a cytometric technique in which cells suspended in a fluid flow one at a time through a focus of exciting light, which is scattered in patterns characteristic to the cells and their components; cells are frequently labeled with fluorescent markers so that light is first absorbed and then emitted at altered frequencies.

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

IATA (Integrated Approach to Testing and Assessment): A structured approach used for hazard identification (potential), hazard characterisation (potency) and/or safety assessment (potential/potency and exposure) of a chemical or group of chemicals, which strategically integrates and weights all relevant data to inform regulatory decision regarding potential hazard and/or risk and/or the need for further targeted and therefore minimal testing.

Mixture: A mixture or a solution composed of two or more substances.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration $\geq 10\%$ (w/w) and $< 80\%$ (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

Positive control: A replicate containing all components of a test system and treated with a substance known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Pre-haptens: chemicals which become sensitisers through abiotic transformation, e.g. through oxidation.

Pro-haptens: chemicals requiring enzymatic activation to exert skin sensitisation potential.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test (14).

Reliability: Measures of the extent that a test can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability (14).

Run: A run consists of one or more test chemicals tested concurrently with a solvent/vehicle control and with a positive control.

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test. It is a measure of accuracy for a test that produces categorical results, and is an important consideration in assessing the relevance of a test (14).

S.I.: Stimulation Index. Relative values of geometric mean fluorescence intensity in chemical-treated cells compared to solvent-treated cells.

Solvent/vehicle control: An untreated sample containing all components of a test system except of the test chemical, but including the solvent/vehicle that is used. It is used to establish the baseline response for the samples treated with the test chemical dissolved or stably dispersed in the same solvent/vehicle. When tested with a concurrent medium control, this sample also demonstrates whether the solvent/vehicle interacts with the test system.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test. It is a measure of accuracy for a test that produces categorical results and is an important consideration in assessing the relevance of a test (14).

Staining buffer: A phosphate buffered saline containing 5 % foetal calf serum.

Substance: A chemical element and its compounds in the natural state or obtained by any production process, including any additive necessary to preserve its stability and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Test chemical: Any substance or mixture tested using this test.

United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (16).

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

Valid test: A test considered to have sufficient relevance and reliability for a specific purpose and which is based on scientifically sound principles. A test is never valid in an absolute sense, but only in relation to a defined purpose (14).

Appendix 2.2

PROFICIENCY SUBSTANCES

Prior to routine use of the test described in this Appendix to test method B.71, laboratories should demonstrate technical proficiency by correctly obtaining the expected U-SENS™ prediction for the 10 substances recommended in Table 1 and by obtaining CV70 and EC150 values that fall within the respective reference range for at least 8 out of the 10 proficiency substances. Proficiency substances were selected to represent the range of responses for skin sensitisation hazards. Other selection criteria were that the substances are commercially available, and that high-quality *in vivo* reference data as well as high quality *in vitro* data generated with the U-SENS™ test are available. Also, published reference data are available for the U-SENS™ test (1) (8).

Table 1

Recommended substances for demonstrating technical proficiency with the U-SENS™ test

Proficiency substances	CASRN	Physical state	<i>In vivo</i> prediction (1)	U-SENS™ Solvent/Vehicle	U-SENS™ CV70 Reference Range in µg/ml (2)	U-SENS™ EC150 Reference Range in µg/ml (2)
4-Phenylenediamine	106-50-3	Solid	Sensitiser(strong)	Complete medium (3)	<30	Positive(≤10)
Picryl sulfonic acid	2508-19-2	Liquid	Sensitizer(strong)	Complete medium	>50	Positive(≤50)
Diethyl maleate	141-05-9	Liquid	Sensitiser(mod-erate)	DMSO	10-100	Positive(≤20)
Resorcinol	108-46-3	Solid	Sensitiser(mod-erate)	Complete medium	>100	Positive(≤50)
Cinnamic alcohol	104-54-1	Solid	Sensitiser(weak)	DMSO	>100	Positive(10-100)
4-Allylanisole	140-67-0	Liquid	Sensitiser(weak)	DMSO	>100	Positive(<200)
Saccharin	81-07-2	Solid	Non-sensitiser	DMSO	>200	Negative(>200)
Glycerol	56-81-5	Liquid	Non-sensitiser	Complete medium	>200	Negative(>200)
Lactic acid	50-21-5	Liquid	Non-sensitiser	Complete medium	>200	Negative(>200)
Salicylic acid	69-72-7	Solid	Non-sensitiser	DMSO	>200	Negative(>200)

Abbreviations: CAS RN = Chemical Abstracts Service Registry Number

(1) The *in vivo* hazard and (potency) prediction is based on LLNA data (1) (8). The *in vivo* potency is derived using the criteria proposed by ECETOC (17).

(2) Based on historical observed values (1) (8).

(3) Complete medium: RPMI-1640 medium supplemented with 10 % foetal calf serum, 2 mM L-glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin (8).

Appendix 3

IN VITRO SKIN SENSITISATION: IL-8 LUC ASSAY

INITIAL CONSIDERATIONS AND LIMITATIONS

1. In contrast to assays analysing the expression of cell surface markers, the IL8-Luc assay quantifies changes in IL-8 expression, a cytokine associated with the activation of dendritic cells (DC). In the THP-1-derived IL-8 reporter cell line (THP-G8, established from the human acute monocytic leukemia cell line THP-1), IL-8 expression is measured following exposure to sensitisers (1). The expression of luciferase is then used to aid discrimination between skin sensitisers and non-sensitisers.
2. The IL-8 Luc assay has been evaluated in a validation study (2) conducted by the Japanese Centre for the Validation of Alternatives Methods (JaCVAM), the **Ministry of Economy, Trade and Industry** (METI), and the Japanese Society **for Alternatives to Animal Experiments (JSAAE)** and subsequently subjected to independent peer review (3) under the auspices of JaCVAM and the Ministry of Health, Labour and Welfare (MHLW) with the support of the **International Cooperation on Alternative Test** Methods (ICATM). Considering all available evidence and input from regulators and stakeholders, the IL-8 Luc assay is considered useful as part of IATA to discriminate sensitisers from non-sensitisers for the purpose of hazard classification and labelling. Examples of the use of IL-8 Luc assay data in combination with other information are reported in the literature (4) (5) (6).
3. The IL-8 Luc assay proved to be transferable to laboratories experienced in cell culture and luciferase measurement. Within and between laboratory reproducibilities were 87,7 % and 87,5 %, respectively (2). Data generated in the validation study (2) and other published work (1) (6) show that versus the LLNA, the IL-8 Luc assay judged 118 out of 143 chemicals as positive or negative and judged 25 chemicals as inconclusive and the accuracy of the IL-8 Luc assay in distinguishing skin sensitisers (UN GHS/CLP Cat. 1) from non-sensitisers (UN GHS/CLP No Cat.) is 86 % (101/118) with a sensitivity of 96 % (92/96) and specificity of 41 % (9/22). Excluding substances outside the applicability domain described below (paragraph 5), the IL-8 Luc assay judged 113 out of 136 chemicals as positive or negative and judged 23 chemicals as inconclusive and the accuracy of the IL-8 Luc assay is 89 % (101/113) with sensitivity of 96 % (92/96) and specificity of 53 % (9/17). Using human data cited in Urbisch et al. (7), the IL-8 Luc assay judged 76 out of 90 chemicals as positive or negative and judged 14 chemicals as inconclusive and the accuracy is 80 % (61/76), sensitivity is 93 % (54/58) and specificity is 39 % (7/18). Excluding substances outside the applicability domain, the IL-8 Luc assay judged 71 out of 84 chemicals as positive or negative and judged 13 chemicals as inconclusive and the accuracy is 86 % (61/71) with sensitivity of 93 % (54/58) and specificity of 54 % (7/13). False negative predictions with the IL-8 Luc assay are more likely to occur with chemicals showing low/moderate skin sensitisation potency (UN GHS/CLP subcategory 1B) than those with high potency (UN GHS/CLP subcategory 1A) (6). Together, the information supports a role for the IL-8 Luc assay in the identification of skin sensitisation hazards. The accuracy given for the IL-8 Luc assay as a standalone test is only for guidance, as the test should be considered in combination with other sources of information in the context of an IATA and in accordance with the provisions of paragraphs 7 and 8 in the General introduction. Furthermore, when evaluating non-animal tests for skin sensitisation, it should be remembered that the LLNA and other animal tests may not fully reflect the situation in humans.
4. On the basis of the data currently available, the IL-8 Luc assay was shown to be applicable to test chemicals covering a variety of organic functional groups, reaction mechanisms, skin sensitisation potency (as determined in *in vivo* studies) and physicochemical properties (2) (6).

5. Although the IL-8 Luc assay uses X-VIVO™ 15 as a solvent, it correctly evaluated chemicals with a Log K_{ow} >3.5 and those with a water solubility of around 100 µg/ml as calculated by EPI Suite™ and its performance to detect sensitisers with poor water solubility is better than that of the IL-8 Luc assay using dimethyl sulfoxide (DMSO) as a solvent (2). However, negative results for test chemicals that are not dissolved at 20 mg/ml may produce false negative results due to their inability to dissolve in X-VIVO™ 15. Therefore, negative results for these chemicals should not be considered. A high false negative rate for anhydrides was seen in the validation study. Furthermore, because of the limited metabolic capability of the cell line (8) and the experimental conditions, pro-haptens (substances requiring metabolic activation) and pre-haptens (substances activated by air oxidation) might give negative results in the assay. However, although negative results for suspected pre/prohaptens should be interpreted with caution, the IL-8 Luc assay correctly judged 11 out of 11 pre-haptens, 6/6 pro-haptens, and 6/8 pre/pro-haptens in the IL-8 Luc assay data set (2). Based on the recent comprehensive review on three non-animal tests (the DPRA, the KeratinoSens™ and the h-CLAT) to detect pre and prohaptens (9), and based on the fact that THP-G8 cells used in the IL-8 Luc assay is a cell line derived from THP-1 that is used in the h-CLAT, the IL-8 Luc assay may also contribute to increase the sensitivity of non-animal tests to detect pre and pro-haptens in the combination of other tests. Surfactants tested so far gave (false) positive results irrespective of their type (e.g. cationic, anionic or non-ionic). Finally, chemicals that interfere with luciferase can confound its activity/measurement, causing apparent inhibition or increased luminescence (10). For example, phytoestrogen concentrations higher than 1µM were reported to interfere with luminescence signals in other luciferase-based reporter gene assays due to over-activation of the luciferase reporter gene. Consequently, luciferase expression obtained at high concentrations of phytoestrogens or compounds suspected of producing phytoestrogen-like activation of the luciferase reporter gene needs to be examined carefully (11). Based on the above, surfactants, anhydrides and chemicals interfering with luciferase are outside the applicability domain of this assay. In cases where there is evidence demonstrating the non-applicability of the IL-8 Luc assay to other specific categories of test chemicals, the test should not be used for those specific categories.

6. As described above, the IL-8 Luc assay supports discrimination of skin sensitisers from non-sensitisers. Further work, preferably based on human data, is required to determine whether IL-8 Luc results can contribute to potency assessment when considered in combination with other information sources.

7. Definitions are provided in Appendix 3.1.

PRINCIPLE OF THE TEST

8. The IL-8 Luc assay makes use of a human monocytic leukemia cell line THP-1 that was obtained from the American Type Culture Collection (Manassas, VA, USA). Using this cell line, the Dept. of Dermatology, Tohoku University School of Medicine, established a THP-1-derived IL-8 reporter cell line, THP-G8, that harbours the Stable Luciferase Orange (SLO) and Stable Luciferase Red (SLR) luciferase genes under the control of the IL-8 and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoters, respectively (1). This allows quantitative measurement of luciferase gene induction by detecting luminescence from well-established light producing luciferase substrates as an indicator of the activity of the IL-8 and GAPDH in cells following exposure to sensitising chemicals.

9. The dual-colour assay system comprises an orange-emitting luciferase (SLO; max = 580 nm) (12) for the gene expression of the IL-8 promoter as well as a red-emitting luciferase (SLR; max = 630 nm) (13) for the gene expression of the internal control promoter, GAPDH. The two luciferases emit different colours upon reacting with firefly d-luciferin and their luminescence is measured simultaneously in a one-step reaction by dividing the emission from the assay mixture using an optical filter (14) (Appendix 3.2).

10. THP-G8 cells are treated for 16 hours with the test chemical, after which SLO luciferase activity (SLO-LA) reflecting IL-8 promoter activity and SLR luciferase activity (SLR-LA) reflecting GAPDH promoter activity are measured. To make the abbreviations easy to understand, SLO-LA and SLR-LA are designated as IL8LA and GAPLA, respectively. Table 1 gives a description of the terms associated with luciferase activity in the IL-8 Luc assay. The measured values are used to calculate the normalised IL8LA (nIL8LA), which is the ratio of IL8LA to GAPLA; the induction of nIL8LA (Ind-IL8LA), which is the ratio of the arithmetic means of quadruple-measured values of the nIL8LA of THP-G8 cells treated with a test chemical and the values of the nIL8LA of untreated THP-G8 cells; and the inhibition of GAPLA (Inh-GAPLA), which is the ratio of the arithmetic means of quadruple-measured values of the GAPLA of THP-G8 cells treated with a test chemical and the values of the GAPLA of untreated THP-G8 cells, and used as an indicator for cytotoxicity.

Table 1

Description of terms associated with the luciferase activity in the IL-8 Luc assay

Abbreviations	Definition
GAPLA	SLR luciferase activity reflecting GAPDH promoter activity
IL8LA	SLO luciferase activity reflecting IL-8 promoter activity
nIL8LA	$IL8LA / GAPLA$
Ind-IL8LA	nIL8LA of THP-G8 cells treated with chemicals / nIL8LA of untreated cells
Inh-GAPLA	GAPLA of THP-G8 treated with chemicals / GAPLA of untreated cells
CV05	The lowest concentration of the chemical at which Inh-GAPLA becomes $< 0,05$.

11. Performance standards (PS) (15) are available to facilitate the validation of modified *in vitro* IL-8 luciferase tests similar to the IL-8 Luc assay and allow for timely amendment of OECD Test Guideline 442E for their inclusion. OECD Mutual Acceptance of Data (MAD) will only be guaranteed for tests validated according to the PS, if these tests have been reviewed and included in Test Guideline 442E by the OECD (16).

DEMONSTRATION OF PROFICIENCY

12. Prior to routine use of the test described in this Appendix to test method B.71, laboratories should demonstrate technical proficiency, using the 10 Proficiency Substances listed in Appendix 3.3 in compliance with the Good *in vitro* Method Practices (17). Moreover, test users should maintain a historical database of data generated with the reactivity checks (see paragraph 15) and with the positive and solvent/vehicle controls (see paragraphs 21-24), and use these data to confirm the reproducibility of the test in their laboratory is maintained over time.

PROCEDURE

13. The Standard Operating Procedure (SOP) for the IL-8 Luc assay is available and should be employed when performing the test (18). Laboratories willing to perform the test can obtain the recombinant THP-G8 cellline

from GPC Lab. Co. Ltd., Tottori, Japan, upon signing a Material Transfer Agreement (MTA) in line with the conditions of the OECD template. The following paragraphs provide a description of the main components and procedures of the assay.

Preparation of cells

14. The THP-G8 cell line from GPC Lab. Co. Ltd., Tottori, Japan, should be used for performing the IL-8 Luc assay (see paragraphs 8 and 13). On receipt, cells are propagated (2-4 passages) and stored frozen as a homogeneous stock. Cells from this stock can be propagated up to a maximum of 12 passages or a maximum of 6 weeks. The medium used for propagation is the RPMI-1 640 culture medium containing 10 % foetal bovine serum (FBS), antibiotic/antimycotic solution (100U/ml of penicillin G, 100µg/ml of streptomycin and 0,25µg/ml of amphotericin B in 0,85 % saline) (e.g. GIBCO Cat#15 240-062), 0,15µg/ml Puromycin (e.g. CAS:58-58-2) and 300µg/ml G418 (e.g. CAS:1083 21-42-2).
15. Prior to use for testing, the cells should be qualified by conducting a reactivity check. This check should be performed 1-2 weeks or 2-4 passages after thawing, using the positive control, 4-nitrobenzyl bromide (4-NBB) (CAS:100-11-8, ≥ 99 % purity) and the negative control, lactic acid (LA) (CAS:50-21-5, ≥ 85 % purity). 4-NBB should produce a positive response to Ind-IL8LA (≥ 1.4), while LA should produce a negative response to Ind-IL8LA (< 1.4). Only cells that pass the reactivity check are used for the assay. The check should be performed according to the procedures described in paragraphs 22-24.
16. For testing, THP-G8 cells are seeded at a density of 2 to 5×10^5 cells/ml, and pre-cultured in culture flasks for 48 to 96 hours. On the day of the test, cells harvested from the culture flask are washed with RPMI-1 640 containing 10 % FBS without any antibiotics, and then, resuspended with RPMI-1 640 containing 10 % FBS without any antibiotics at 1×10^6 cells/ml. Then, cells are distributed into a 96-well flat-bottom black plate (e.g. Costar Cat#3 603) with 50µl (5×10^4 cells/well).

Preparation of the test chemical and control substances

17. The test chemical and control substances are prepared on the day of testing. For the IL-8 Luc assay, test chemicals are dissolved in X-VIVO™ 15, a commercially available serum-free medium (Lonza, 04-418Q), to the final concentration of 20 mg/ml. X-VIVO™ 15 is added to 20 mg of test chemical (regardless of the chemical's solubility) in a microcentrifuge tube and brought to a volume of 1ml and then vortexed vigorously and shaken on a rotor at a maximum speed of 8 rpm for 30 min at an ambient temperature of about 20°C. Furthermore, if solid chemicals are still insoluble, the tube is sonicated until the chemical is dissolved completely or stably dispersed. For test chemicals soluble in X-VIVO™ 15, the solution is diluted by a factor of 5 with X-VIVO™ 15 and used as an X-VIVO™ 15 stock solution of the test chemical (4 mg/ml). For test chemicals not soluble in X-VIVO™ 15, the mixture is rotated again for at least 30 min, then centrifuged at 15,000 rpm ($\approx 20\ 000$ g) for 5 min; the resulting supernatant is used as an X-VIVO™ 15 stock solution of the test chemical. A scientific rationale should be provided for the use of other solvents, such as DMSO, water, or the culture medium. The detailed procedure for dissolving chemicals is shown in Appendix 3.5. The X-VIVO™ 15 solutions described in paragraphs 18-23 are mixed 1:1 (v/v) with the cell suspensions prepared in a 96-well flat-bottom black plate (see paragraph 16).
18. The first test run is aimed to determine the cytotoxic concentration and to examine the skin sensitising potential of chemicals. Using X-VIVO™ 15, serial dilutions of the X-VIVO™ 15 stock solutions of the test chemicals are made at a dilution factor of two (see Appendix 3.5) using a 96-well assay block (e.g. Costar Cat#EW-01 729-03). Next, 50 µl/well of diluted solution is added to 50 µl of the cell suspension in a 96-well flat-bottom black plate. Thus for test chemicals that are soluble in X-VIVO™ 15, the final concentrations of the test chemicals range from 0,002 to 2 mg/ml (Appendix 3.5). For test chemicals that are not soluble in X-VIVO™ 15 at 20 mg/ml, only dilution factors that range from 2 to 2^{10} , are determined, although the actual final concentrations of the test chemicals remain uncertain and are dependent on the saturated concentration of the test chemicals in the X-VIVO™ 15 stock solution.

19. In subsequent test runs (i.e. the second, third, and fourth replicates), the X-VIVO™ 15 stock solution is made at the concentration 4 times higher than the concentration of cell viability 05 (CV05; the lowest concentration at which the Inh-GAPLA becomes <0.05) in the first experiment. If Inh-GAPLA does not decrease below 0,05 at the highest concentration in the first run, the X-VIVO™ 15 stock solution is made at the first run highest concentration. The concentration of CV05 is calculated by dividing the concentration of the stock solution in the first run by dilution factor for CV05 (X) (dilution factor CV05 (X); the dilution factor required to dilute stock solution to CV05) (see Appendix 3.5). For test substances not soluble in X-VIVO at 20 mg/ml, CV05 is determined by the concentration of the stock solution x 1/X. For run 2 to 4, a second stock solution is prepared as 4 x CV05 (Appendix 3.5).
20. Serial dilutions of the X-VIVO™ 15 second stock solutions are made at a dilution factor of 1,5 using a 96-well assay block. Next, 50 µl/well of diluted solution is added to 50 µl of the cell suspension in the wells of a 96-well flat-bottom black plate. Each concentration of each test chemical should be tested in 4 wells. The samples are then mixed on a plate shaker and incubated for 16 hours at 37 °C and 5 % CO₂, after which the luciferase activity is measured as described below.
21. The solvent control is the mixture of 50 µl/well of X-VIVO™ 15 and 50 µl/well of cell suspension in RPMI-1 640 containing 10 % FBS.
22. The recommended positive control is 4-NBB. 20 mg of 4-NBB is prepared in a 1,5-ml microfuge tube, to which X-VIVO™ 15 is added up to 1 ml. The tube is vortexed vigorously and shaken on a rotor at a maximum speed of 8 rpm for at least 30 min. After centrifugation at 20 000g for 5 min, the supernatant is diluted by a factor of 4 with X-VIVO™ 15, and 500 µl of the diluted supernatant is transferred to a well in a 96-well assay block. The diluted supernatant is further diluted with X-VIVO™ 15 at factors of 2 and 4, and 50 µl of the solution is added to 50 µl of THP-G8 cell suspension in the wells of a 96-well flat-bottom black plate (Appendix 3.6). Each concentration of the positive control should be tested in 4 wells. The plate is agitated on a plate shaker, and incubated in a CO₂ incubator for 16 hours (37 °C, 5 % CO₂), after which the luciferase activity is measured as described in paragraph 29.
23. The recommended negative control is LA. 20 mg of LA prepared in a 1,5-ml microfuge tube, to which X-VIVO™ 15 is added up to 1 ml (20 mg/ml). Twenty mg/ml of LA solution is diluted by a factor of 5 with X-VIVO™ 15 (4 mg/ml); 500 µl of this 4 mg/ml LA solution is transferred to a well of a 96-well assay block. This solution is diluted by a factor of 2 with X-VIVO™ 15 and then diluted again by a factor of 2 to produce 2 mg/ml and 1 mg/ml solutions. 50 µl of these 3 solutions and vehicle control (X-VIVO™ 15) are added to 50 µl of THP-G8 cell suspension in the wells of a 96-well flat-bottom black plate. Each concentration of the negative control is tested in 4 wells. The plate is agitated on a plate shaker and incubated in a CO₂ incubator for 16 hours (37 °C, 5 % CO₂), after which the luciferase activity is measured as described in paragraph 29.
24. Other suitable positive or negative controls may be used if historical data are available to derive comparable run acceptance criteria.
25. Care should be taken to avoid evaporation of volatile test chemicals and cross-contamination between wells by test chemicals, e.g. by sealing the plate prior to the incubation with the test chemicals.
26. The test chemicals and solvent control require 2 to 4 runs to derive a positive or negative prediction (see Table 2). Each run is performed on a different day with fresh X-VIVO™ 15 stock solution of test chemicals and independently harvested cells. Cells may come from the same passage.

Luciferase activity measurements

27. Luminescence is measured using a 96-well microplate luminometer equipped with optical filters, e.g. Phelios (ATTO, Tokyo, Japan), Tristan 941 (Berthold, Bad Wildbad, Germany) and the ARVO series (PerkinElmer, Waltham, MA, USA). The luminometer must be calibrated for each test to ensure reproducibility (19). Recombinant orange and red emitting luciferases are available for this calibration.

28. 100µl of pre-warmed Tripluc® Luciferase assay reagent (Tripluc) is transferred to each well of the plate containing the cell suspension treated with or without chemical. The plate is shaken for 10 min at an ambient temperature of about 20 °C. The plate is placed in the luminometer to measure the luciferase activity. Bioluminescence is measured for 3 sec each in the absence (F0) and presence (F1) of the optical filter. Justification should be provided for the use of alternative settings, e.g. depending on the model of luminometer used.

29. Parameters for each concentration are calculated from the measured values, e.g. IL8LA, GAPLA, nIL8LA, Ind-IL8LA, Inh-GAPLA, the mean ±SD of IL8LA, the mean ±SD of GAPLA, the mean ±SD of nIL8LA, the mean ±SD of Ind-IL8LA, the mean ±SD of Inh-GAPLA, and the 95 % confidence interval of Ind-IL8LA. Definitions of the parameters used in this paragraph are provided in Appendices 3.1 and 3.4.

30. Prior to measurement, colour discrimination in multi-colour reporter assays is generally achieved using detectors (luminometer and plate reader) equipped with optical filters, such as sharp-cut (long-pass or short-pass) filters or band-pass filters. The transmission coefficients of the filters for each bioluminescence signal colour should be calibrated prior to testing, per Appendix 3.2.

DATA AND REPORTING

Data evaluation

31. Criteria for a positive/negative decision require that in each run:
 - an IL-8 Luc assay prediction is judged positive if a test chemical has a Ind-IL8LA 1.4 and the lower limit of the 95 % confidence interval of Ind-IL8LA 1.0

 - an IL-8 Luc assay prediction is judged negative if a test chemical has a Ind-IL8LA < 1.4 and/or the lower limit of the 95 % confidence interval of Ind-IL8LA < 1.0

Prediction model

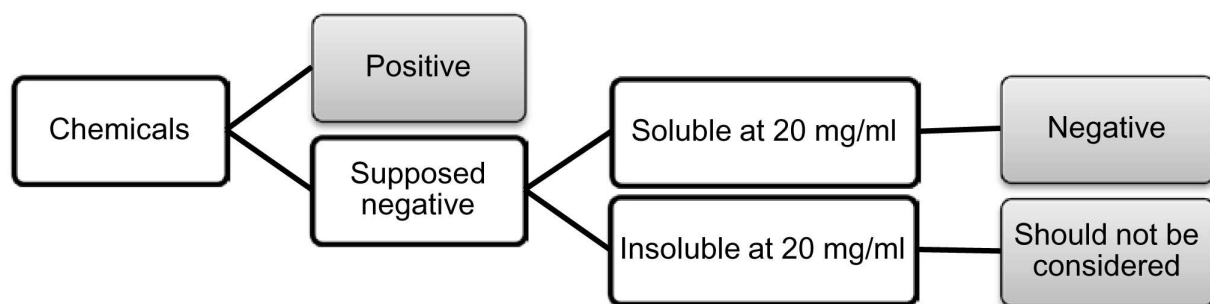
32. Test chemicals that provide two positive results from among the 1st, 2nd, 3rd or 4th runs are identified as positives whereas those that give three negative results from among the 1st, 2nd, 3rd or 4th runs are identified as supposed negative (Table 2). Among supposed negative chemicals, chemicals that are dissolved at 20 mg/ml of X-VIVO™ 15 are judged as negative, while chemicals that are not dissolved at 20 mg/ml of X-VIVO™ 15 should not be considered (Figure 1).

Table 2
Criteria for identifying positive and supposed negative

1st run	2nd run	3rd run	4th run	Final prediction
Positive	Positive	—	—	Positive
	Negative	Positive	—	Positive
		Negative	Positive	Positive
		Negative	Negative	Supposed negative
Negative	Positive	Positive	—	Positive
		Negative	Positive	Positive
		Negative	Negative	Supposed negative
	Negative	Positive	Positive	Positive
		Negative	Negative	Supposed negative
		Negative	—	Supposed negative

Figure 1

Prediction model for final judgment



Acceptance criteria

33. The following acceptance criteria should be met when using the IL-8 Luc assay:

- Ind-IL8LA should be more than 5.0 at least in one concentration of the positive control, 4-NBB, in each run.
- Ind-IL8LA should be less than 1.4 at any concentration of the negative control, lactic acid, in each run.

- Data from plates for which the GAPLA of control wells with cells and Tripluc but without chemicals is less than 5 times of that of well containing test medium only (50 µl/well of RPMI-1 640 containing 10 % FBS and 50 µl/well of X-VIVO™ 15) should be rejected.

- Data from plates for which the Inh-GAPLA of all concentrations of the test or control chemicals is less than 0,05 should be rejected. In this case, the first test should be repeated so the highest final concentration of the repeated test is the lowest final concentration of the previous test.

Test report

34. The test report should include the following information:

Test chemicals

Mono-constituent substance:

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;

- Physical appearance, water solubility, molecular weight, and additional relevant physicochemical properties, to the extent available;

- Purity, chemical identity of impurities as appropriate and practically feasible, etc.;

- Treatment prior to testing, if applicable (e.g. warming, grinding);

- Solubility in X-VIVO™ 15. For chemicals that are insoluble in X-VIVO™ 15, whether precipitation or flotation are observed after centrifugation;

- Concentration(s) tested;

- Storage conditions and stability to the extent available;

- Justification for choice of solvent/vehicle for each test chemical if X-VIVO™ 15 has not been used.

Multi-constituent substance, UVCB and mixture:

- Characterisation as far as possible by e.g. chemical identity (see above), purity, quantitative occurrence and relevant physicochemical properties (see above) of the constituents, to the extent available;

- Physical appearance, water solubility, and additional relevant physicochemical properties, to the extent available;
- Molecular weight or apparent molecular weight in case of mixtures/polymers of known compositions or other information relevant for the conduct of the study;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Solubility in X-VIVO™ 15. For chemicals that are insoluble in X-VIVO™ 15, whether precipitation or flotation are observed after centrifugation;
- Concentration(s) tested;
- Storage conditions and stability to the extent available.
- Justification for choice of solvent/vehicle for each test chemical, if X-VIVO™ 15 has not been used.

Controls

Positive control:

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
- Physical appearance, water solubility, molecular weight, and additional relevant physicochemical properties, to the extent available and where applicable;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc;
- Treatment prior to testing, if applicable (e.g. warming, grinding);
- Concentration(s) tested;
- Storage conditions and stability to the extent available;
- Reference to historical positive control results demonstrating suitable acceptance criteria, if applicable.

Negative control:

- Chemical identification, such as IUPAC or CAS name(s), CAS number(s), and/or other identifiers;
- Purity, chemical identity of impurities as appropriate and practically feasible, etc;

- Physical appearance, molecular weight, and additional relevant physicochemical properties in the case other negative controls than those mentioned in the Test Guideline are used and to the extent available;
- Storage conditions and stability to the extent available;
- Justification for choice of solvent for each test chemical.

Test conditions

- Name and address of the sponsor, test facility and study director;
- Description of test used;
- Cell line used, its storage conditions, and source (e.g. the facility from which it was obtained);
- Lot number and origin of FBS, supplier name, lot number of 96-well flat-bottom black plate, and lot number of Tripluc reagent;
- Passage number and cell density used for testing;
- Cell counting method used for seeding prior to testing and measures taken to ensure homogeneous cell number distribution;
- Luminometer used (e.g. model), including instrument settings, luciferase substrate used, and demonstration of appropriate luminescence measurements based on the control test described in Appendix 3.2;
- The procedure used to demonstrate proficiency of the laboratory in performing the test (e.g. by testing of proficiency substances) or to demonstrate reproducible performance of the test over time.

Test procedure

- Number of replicates and runs performed;
- Test chemical concentrations, application procedure and exposure time (if different from those recommended);
- Description of evaluation and decision criteria used;
- Description of study acceptance criteria used;
- Description of any modifications of the test procedure.

Results

- Measurements of IL8LA and GAPLA;
- Calculations for nIL8LA, Ind-IL8LA, and Inh-GAPLA;
- The 95 % confidence interval of Ind-IL8LA;
- A graph depicting dose-response curves for induction of luciferase activity and viability;
- Description of any other relevant observations, if applicable.

Discussion of the results

- Discussion of the results obtained with the IL-8 Luc assay;
- Consideration of the assay results in the context of an IATA, if other relevant information is available.

Conclusion

LITERATURE

- (1) Takahashi T, Kimura Y, Saito R, Nakajima Y, Ohmiya Y, Yamasaki K, and Aiba S. (2011). An *in vitro* test to screen skin sensitizers using a stable THP-1-derived IL-8 reporter cell line, THP-G8. *Toxicol Sci* 124:359-69.
- (2) OECD (2017). Validation report for the international validation study on the IL-8 Luc assay as a test evaluating the skin sensitizing potential of chemicals conducted by the IL-8 Luc Assay. Series on Testing and Assessment No 267, ENV/JM/MONO(2017)19. Organisation for Economic Cooperation and Development, Paris. Available at: <http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm>.
- (3) OECD (2017). Report of the Peer Review Panel for the IL-8 Luciferase (IL-8 Luc) Assay for *in vitro* skin sensitisation. Series on Testing and Assessment No 258, ENV/JM/MONO(2017)20. Organisation for Economic Cooperation and Development, Paris. Available at: <http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm>.
- (4) OECD (2016) Guidance Document On The Reporting Of Defined Approaches And Individual Information Sources To Be Used Within Integrated Approaches To Testing And Assessment (IATA) For Skin Sensitisation, Series on Testing & Assessment No 256, ENV/JM/MONO(2016)29. Organisation for Economic Cooperation and Development, Paris. Available at: <http://www.oecd.org/env/ehs/testing/series-testing-assessment-publications-number.htm>.

- (5) van der Veen JW, Rorije E, Emter R, Natsch A, van Loveren H, and Ezendam J. (2014). Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. *Regul Toxicol Pharmacol* 69:371-9.
- (6) Kimura Y, Fujimura C, Ito Y, Takahashi T, Nakajima Y, Ohmiya Y, and Aiba S. (2015). Optimization of the IL-8 Luc assay as an *in vitro* test for skin sensitization. *Toxicol In Vitro* 29:1 816-30.
- (7) Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, et al. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul Toxicol Pharmacol* 71:337-51.
- (8) Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y, Miyazawa M, Ito Y, Nishiyama N, and Itagaki H. (2010). A comparative evaluation of *in vitro* skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). *Alternatives to laboratory animals: ATLA* 38:275-84.
- (9) Patlewicz G, Casati S, Basketter DA, Asturiol D, Roberts DW, Lepoittevin J-P, Worth A and Aschberger K (2016) Can currently available non-animal methods detect pre and pro haptens relevant for skin sensitisation? *Regul Toxicol Pharmacol*, 82:147-155.
- (10) Thorne N, Inglese J, and Auld DS. (2010). Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology. *Chem Biol* 17:646-57.
- (11) OECD (2016). Test No 455: Performance-Based Test Guideline for Stably Transfected Transactivation *In Vitro* Assays to Detect Estrogen Receptor Agonists and Antagonists, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264265295-en>.
- (12) Viviani V, Uchida A, Suenaga N, Ryufuku M, and Ohmiya Y. (2001). Thr226 is a key residue for bioluminescence spectra determination in beetle luciferases. *Biochem Biophys Res Commun* 280:1 286-91.
- (13) Viviani VR, Bechara EJ, and Ohmiya Y. (1999). Cloning, sequence analysis, and expression of active Phrixothrix railroad-worms luciferases: relationship between bioluminescence spectra and primary structures. *Biochemistry* 38:8 271-9.
- (14) Nakajima Y, Kimura T, Sugata K, Enomoto T, Asakawa A, Kubota H, Ikeda M, and Ohmiya Y. (2005). Multicolor luciferase assay system: one-step monitoring of multiple gene expressions with a single substrate. *Biotechniques* 38:891-4.
- (15) OECD (2017). To be published - Performance Standards for the assessment of proposed similar or modified *in vitro* skin sensitisation IL-8 luc test methods. OECD Environment, Health and Safety Publications, Series on Testing and Assessment. OECD, Paris, France

-
- (16) OECD (2005). Guidance Document the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Environment, Health and Safety publications, OECD Series on Testing and Assessment No 34. OECD, Paris, France.
- (17) OECD (2018). Draft Guidance document: Good *In Vitro* Method Practices (GIVIMP) for the Development and Implementation of *In Vitro* Methods for Regulatory Use in Human Safety Assessment. Organisation for Economic Cooperation and Development, Paris. Available at: [http://www.oecd.org/env/ehs/testing/OECD Final Draft GIVIMP.pdf](http://www.oecd.org/env/ehs/testing/OECD_Final_Draft_GIVIMP.pdf).
- (18) JaCVAM (2016). IL-8 Luc assay protocol, Available at. http://www.jacvam.jp/en_effort/effort02.html.
- (19) Niwa K, Ichino Y, Kumata S, Nakajima Y, Hiraishi Y, Kato D, Viviani VR, and Ohmiya Y. (2010). Quantum yields and kinetics of the firefly bioluminescence reaction of beetle luciferases. *Photochem Photobiol* 86:1 046-9.
- (20) OECD (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins, Part 1: Scientific Evidence. OECD Environment, Health and Safety Publications, Series on Testing and Assessment No 168. OECD, Paris, France.
- (21) United Nations (2015). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Sixth revised edition. New York & Geneva: United Nations Publications. ISBN: 978-92-1-117006-1. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html.

Appendix 3.1

DEFINITIONS

Accuracy: The closeness of agreement between test results and accepted reference values. It is a measure of test performance and one aspect of relevance. The term is often used interchangeably with concordance to mean the proportion of correct outcomes of a test (16).

AOP (Adverse Outcome Pathway): Sequence of events from the chemical structure of a target chemical or group of similar chemicals through the molecular initiating event to an *in vivo* outcome of interest (20).

Chemical: A substance or a mixture.

CV05: Cell viability 05, i.e. minimum concentration at which chemicals show less than 0,05 of Inh-GAPLA.

FlnSLO-LA: Abbreviation used in the validation report and in previous publications regarding the IL-8 Luc assay to refer to Ind-IL8LA. See Ind-IL8LA for definition.

GAPLA: Luciferase Activity of Stable Luciferase Red (SLR) (max = 630 nm), regulated by GAPDH promoter and demonstrates cell viability and viable cell number.

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

IATA (Integrated Approach to Testing and Assessment): A structured approach used for hazard identification (potential), hazard characterisation (potency) and/or safety assessment (potential/potency and exposure) of a chemical or group of chemicals, which strategically integrates and weights all relevant data to inform regulatory decision regarding potential hazard and/or risk and/or the need for further targeted and therefore minimal testing.

II-SLR-LA: Abbreviation used in the validation report and in previous publications regarding the IL-8 Luc assay to refer to Inh-GAPLA. See Inh-GAPLA for definition

IL-8 (Interleukin-8): A cytokine derived from endothelial cells, fibroblasts, keratinocytes, macrophages, and monocytes that causes chemotaxis of neutrophils and T-cell lymphocytes.

IL8LA: Luciferase Activity of Stable Luciferase Orange (SLO) (max = 580 nm), regulated by IL-8 promoter.

Ind-IL8LA: Fold induction of nIL8LA. It is obtained by dividing the nIL8LA of THP-G8 cells treated with chemicals by that of non-stimulated THP-G8 cells and represents the induction of IL-8 promoter activity by chemicals.

Inh-GAPLA: Inhibition of GAPLA. It is obtained by dividing GAPLA of THP-G8 treated with chemicals with GAPLA of non-treated THP-G8 and represents cytotoxicity of chemicals.

Minimum induction threshold (MIT): the lowest concentration at which a chemical satisfies the positive criteria

Mixture: A mixture or a solution composed of two or more substances.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80 % (w/w).

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one of the main constituents is present in a concentration ≥ 10 % (w/w) and < 80 % (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

nIL8LA: The SLO luciferase activity reflecting IL-8 promoter activity (IL8LA) normalised by the SLR luciferase activity reflecting GAPDH promoter activity (GAPLA). It represents IL-8 promoter activity after considering cell viability or cell number.

nSLO-LA: Abbreviation used in the validation report and in previous publications regarding the IL-8 Luc assay to refer to nIL8LA. See nIL8LA for definition

Positive control: A replicate containing all components of a test system and treated with a substance known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Pre-haptens: Chemicals which become sensitisers through abiotic transformation.

Pro-haptens: Chemicals requiring enzymatic activation to exert skin sensitisation potential.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test (16).

Reliability: Measures of the extent that a test can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability (16).

Run: A run consists of one or more test chemicals tested concurrently with a solvent/vehicle control and with a positive control.

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test. It is a measure of accuracy for a test that produces categorical results, and is an important consideration in assessing the relevance of a test (16).

SLO-LA: Abbreviation used in the validation report and in previous publications regarding the IL-8 Luc assay to refer to IL8LA. See IL8LA for definition.

SLR-LA: Abbreviation used in the validation report and in previous publications regarding the IL-8 Luc assay to refer to GAPLA. See GAPLA for definition.

Solvent/vehicle control: An untreated sample containing all components of a test system except of the test chemical, but including the solvent/vehicle that is used. It is used to establish the baseline response for the samples treated with the test chemical dissolved or stably dispersed in the same solvent/vehicle. When tested with a concurrent medium control, this sample also demonstrates whether the solvent/vehicle interacts with the test system.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test. It is a measure of accuracy for a test that produces categorical results and is an important consideration in assessing the relevance of a test (16).

Substance: A chemical elements and its compounds in the natural state or obtained by any production manufacturing process, including any additive necessary to preserve the its stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Surfactant: Also called surface-active agent, this is a substance, such as a detergent, that can reduce the surface tension of a liquid and thus allow it to foam or penetrate solids; it is also known as a wetting agent. (TG437)

Test chemical: Any substance or mixture tested using this method.

THP-G8: An IL-8 reporter cell line used in IL-8 Luc assay. The human macrophage-like cell line THP-1 was transfected the SLO and SLR luciferase genes under the control of the IL-8 and GAPDH promoters, respectively.

United Nations Globally Harmonized System of Classification and Labeling of Chemicals (UN GHS): A system proposing the classification of chemicals (substances and mixtures) according to standardised types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (21).

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

Valid test method: A test considered to have sufficient relevance and reliability for a specific purpose and which is based on scientifically sound principles. A test is never valid in an absolute sense, but only in relation to a defined purpose.

Appendix 3.2

PRINCIPLE OF MEASUREMENT OF LUCIFERASE ACTIVITY AND DETERMINATION OF THE TRANSMISSION COEFFICIENTS OF OPTICAL FILTER FOR SLO AND SLR

MultiReporter Assay System -Tripluc- can be used with a microplate-type luminometer with a multi-colour detection system, which can equip an optical filter (e.g. Phelios AB-2350 (ATTO), ARVO (PerkinElmer), Tristar LB941 (Berthold)). The optical filter used in measurement is 600–620 nm long or short pass filter, or 600–700 nm band pass filter.

Measurement of two-colour luciferases with an optical filter

This is an example using Phelios AB-2350 (ATTO). This luminometer is equipped with a 600 nm long pass filter (R60 HOYA Co., 600 nm LP, Filter 1) for splitting SLO (max = 580 nm) and SLR (max = 630 nm) luminescence.

To determine transmission coefficients of the 600 nm LP, first, using purified SLO and SLR luciferase enzymes, measure i) the SLO and SLR bioluminescence intensity without filter (F0), ii) the SLO and SLR bioluminescence intensity that passed through 600 nm LP (Filter 1), and iii) calculate the transmission coefficients of 600 nm LP for SLO and SLR listed below.

Transmission coefficients		Abbreviation	Definition
SLO	Filter 1 Transmission coefficients	κO_{R60}	The filter's transmission coefficient for the SLO
SLR	Filter 1 Transmission coefficients	κR_{R60}	The filter's transmission coefficient for the SLR

When the intensity of SLO and SLR in test sample are defined as O and R, respectively, i) the intensity of light without filter (all optical) F0 and ii) the intensity of light that transmits through 600 nm LP (Filter 1) F1 are described as below.

$$F0=O+R$$

$$F1=\kappa O_{R60} \times O + \kappa R_{R60} \times R$$

These formulas can be rephrased as follows:

$$\begin{pmatrix} F0 \\ F1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \kappa O_{R60} & \kappa R_{R60} \end{pmatrix} \begin{pmatrix} O \\ R \end{pmatrix}$$

Then using calculated transmittance factors (κO_{R60} and κR_{R60}) and measured F0 and F1, you can calculate O and R-value as follows:

$$\begin{pmatrix} O \\ R \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \kappa O_{R60} & \kappa R_{R60} \end{pmatrix}^{-1} \begin{pmatrix} F0 \\ F1 \end{pmatrix}$$

Materials and methods for determining transmittance factor

(1) Reagents

Single purified luciferase enzymes:

Lyophilised purified SLO enzyme

Lyophilised purified SLR enzyme

(which for the validation work were obtained from GPC Lab. Co. Ltd., Tottori, Japan with THP-G8 cell line)

Assay reagent:

Tripluc® Luciferase assay reagent (for example from TOYOBO Cat#MRA-301)

Medium: for luciferase assay (30 ml, stored at 2 – 8°C)

Reagent	Conc.	Final conc. in medium	Required amount
RPMI-1640	—	—	27 ml
FBS	—	10 %	3 ml

(2) Preparation of enzyme solution

Dissolve lyophilised purified luciferase enzyme in tube by adding 200 µl of 10 ~ 100 mM Tris/HCl or HEPES/HCl (pH 7.5 ~ 8.0) supplemented with 10 % (w/v) glycerol, divide the enzyme solution into 10 µl aliquots in 1,5 ml disposable tubes and store them in a freezer at -80°C. The frozen enzyme solution can be used for up to 6 months. When used, add 1 ml of medium for luciferase assay (RPMI-1640 with 10 % FBS) to each tube containing the enzyme solutions (diluted enzyme solution) and keep them on ice to prevent deactivation.

(3) Bioluminescence measurement

Thaw Tripluc® Luciferase assay reagent (Tripluc) and keep it at room temperature either in a water bath or at ambient air temperature. Power on the luminometer 30 min before starting the measurement to allow the photomultiplier to stabilise. Transfer 100 µl of the diluted enzyme solution to a black 96 well plate (flat bottom) (the SLO reference sample to #B1, #B2, #B3, the SLR reference sample to #D1, #D2, #D3). Then, transfer 100 µl of pre-warmed Tripluc to each well of the plate containing the diluted enzyme solution using a pipetman. Shake the plate for 10 min at room temperature (about 25°C) using a plate shaker. Remove bubbles from the solutions in wells if they appear. Place the plate in the luminometer to measure the luciferase activity. Bioluminescence is measured for 3 sec each in the absence (F0) and presence (F1) of the optical filter.

Transmission coefficient of the optical filter was calculated as follows:

Transmission coefficient (SLO ($\kappa_{O_{R60}}$))= (#B1 of F1+ #B2 of F1+ #B3 of F1) / (#B1 of F0+ #B2 of F0+ #B3 of F0)

Transmission coefficient (SLR (κ_{R60}))= (#D1 of F1+ #D2 of F1+ #D3 of F1) / (#D1 of F0+ #D2 of F0+ #D3 of F0)

Calculated transmittance factors are used for all the measurements executed using the same luminometer.

Quality control of equipment

The procedures described in the IL-8 Luc protocol should be used (18).

Appendix 3.3

PROFICIENCY SUBSTANCES

Prior to routine use of the test described in this Appendix to test method B.71, laboratories should demonstrate technical proficiency by obtaining the expected IL-8 Luc assay prediction for the 9 substances recommended in Table 1 and by obtaining values that fall within the respective reference range for at least 8 out of the 9 proficiency substances (selected to represent the range of responses for skin sensitisation hazards). Other selection criteria were that the substances are commercially available, and that high-quality *in vivo* reference data as well as high quality *in vitro* data generated with the IL-8 Luc assay are available. Also, published reference data are available for the IL-8 Luc assay (6) (1).

Table 1

Recommended substances for demonstrating technical proficiency with the IL-8 Luc assay

Proficiency substances	CAS no.	State	Solubility in X-VIVO15 at 20 mg/ml	<i>In vivo</i> prediction (1)	IL-8 Luc prediction (2)	Reference range (µg/ml) (3)	
						CV05 (4)	IL-8 Luc MIT (5)
2,4-Dinitrochlorobenzene	97-00-7	Solid	Insoluble	Sensitiser(Extreme)	Positive	2.3-3.9	0.5-2.3
Formaldehyde	50-00-0	Liquid	Soluble	Sensitiser(Strong)	Positive	9-30	4-9
2-Mercaptobenzothiazole	149-30-4	Solid	Insoluble	Sensitiser(Moderate)	Positive	250-290	60-250
Ethylenediamine	107-15-3	Liquid	Soluble	Sensitiser(Moderate)	Positive	500-700	0.1-0.4
Ethyleneglycol dimethacrylate	97-90-5	Liquid	Insoluble	Sensitiser(Weak)	Positive	> 2000	0.04-0.1
4-Allylanisole (Estragol)	140-67-0	Liquid	Insoluble	Sensitiser(Weak)	Positive	> 2000	0.01-0.07
Streptomycin sulphate	3810-74-0	Solid	Soluble	Non-sensitiser	Negative	> 2000	> 2000
Glycerol	56-81-5	Liquid	Soluble	Non-sensitiser	Negative	> 2000	> 2000
Isopropanol	67-63-0	Liquid	Soluble	Non-sensitiser	Negative	> 2000	> 2000

Abbreviations: CAS no. = Chemical Abstracts Service Registry Number

(1) The *in vivo* potency is derived using the criteria proposed by ECETOC (19).

(2) Based on historical observed values (1) (6).

(3) CV05 and IL-8 Luc MIT were calculated using water solubility given by EPI Suite™.

(4) CV05: the minimum concentration at which chemicals show less than 0,05 of Inh-GAPLA.

(5) MIT: the lowest concentrations at which a chemical satisfies the positive criteria.

Appendix 3.4

INDEXES AND JUDGMENT CRITERIA

nIL8LA (nSLO-LA)

The j-th repetition ($j = 1-4$) of the i-th concentration ($i = 0-11$) is measured for IL8LA (SLO-LA) and GAPLA (SLR-LA) respectively. The normalised IL8LA, referred to as nIL8LA (nSLO-LA), and is defined as:

$$\text{nIL8LA}_{ij} = \text{IL8LA}_{ij} / \text{GAPLA}_{ij}$$

This is the basic unit of measurement in this assay.

Ind-IL8LA (FInSLO-LA)

The fold increase of the averaged nIL8LA (nSLO-LA) for the repetition on the i-th concentration compared with it at the 0 concentration, Ind-IL8LA, is the primary measure of this assay. This ratio is written by the following formula:

$$\text{Ind-IL8LA}_i = \left\{ (1/4) \times \sum_j \text{nIL8LA}_{ij} \right\} / \left\{ (1/4) \times \sum_j \text{nIL8LA}_{0j} \right\}$$

The lead laboratory has proposed that a value of 1.4 corresponds to a positive result for the tested chemical. This value is based on the investigation of the historical data of the lead laboratory. Data management team then used this value through all the phases of validation study. The primary outcome, Ind-IL8LA, is the ratio of 2 arithmetic means as shown in equation.

95 % confidence interval (95 % CI)

The 95 % confidence interval (95 % CI) based on the ratio can be estimated to show the precision of this primary outcome measure. The lower limit of the 95 % CI ≥ 1 indicates that the nIL8LA with the i-th concentration is significantly greater than that with solvent control. There are several ways to construct the 95 % CI. We used the method known as Fieller's theorem in this study. This 95 % confidence interval theorem is obtained from the following formula:

$$\left[\frac{-B - \sqrt{B^2 - 4AC}}{2A}, \frac{-B + \sqrt{B^2 - 4AC}}{2A} \right],$$

Where

$$A = \bar{x}_0^2 - t_{0.975(v)}^2 \times \frac{sd_0^2}{n_0},$$

$$B = -2 \times \bar{x} \times \bar{y},$$

$$C = \bar{y}_i^2 - t_{0.975(v)}^2 \times \frac{sd_{yi}^2}{n_{yi}}, \text{ and } n_0 = 4,$$

$$\bar{x}_0 = (1/n_0) \times \sum_j nIL8LA_{0j},$$

$$sd_0^2 = \{1/(n_0 - 1)\} \times \sum_j (nIL8LA_{0j} - \bar{x}_0)^2,$$

$$n_{yi} = 4,$$

$$\bar{y}_i = (1/n_{yi}) \times \sum_j (nIL8LA_{ij}),$$

$$sd_{yi}^2 = \{1/(n_{yi} - 1)\} \times \sum_j (nIL8LA_{ij} - \bar{y}_i)^2.$$

$t_{0.975(v)}$ is 97.5 percentile of the central t distribution with the v of the degree of freedom, where

$$v = \left(\frac{sd_0^2}{n_0} + \frac{sd_{yi}^2}{n_{yi}} \right) / \left\{ \left(\frac{sd_0^2}{n_0} \right)^2 / (n_0 - 1) + \left(\frac{sd_{yi}^2}{n_{yi}} \right) / (n_{yi} - 1) \right\}.$$

Inh-GAPLA (II-SLR-LA)

The Inh-GAPLA is a ratio of the averaged GAPLA (SLR-LA) for the repetition of the i-th concentration compared with that with solvent control, and this is written by

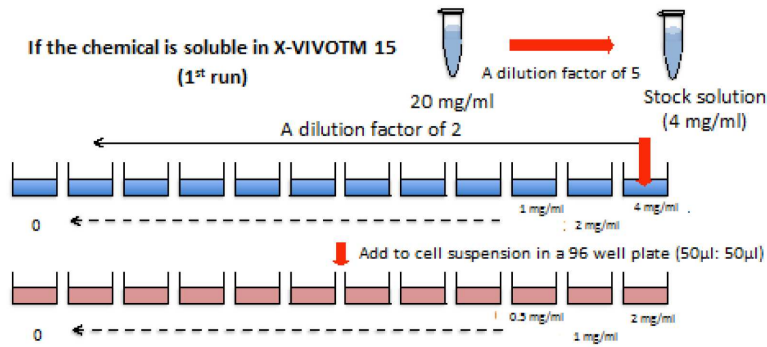
$$Inh - GAPLA_i = \left\{ (1/4) \times \sum_j GAPLA_{ij} \right\} / \left\{ (1/4) \times \sum_j GAPLA_{0j} \right\}.$$

Since the GAPLA is the denominator of the nIL8LA, an extremely small value causes large variation in the nIL8LA. Therefore, Ind-IL8LA values with an extremely small value of Inh-GAPLA (less than 0,05) might be considered poor precision.

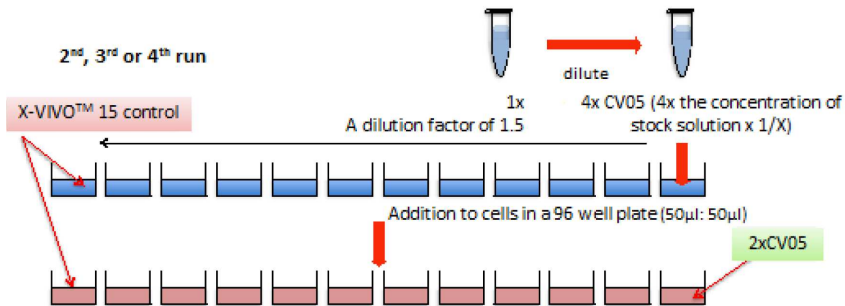
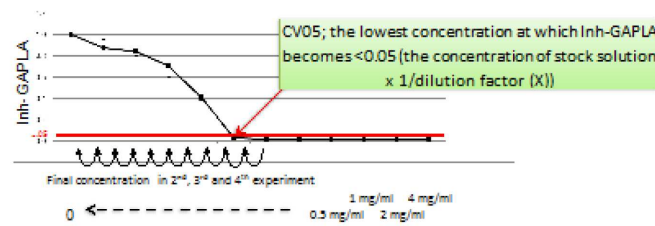
Appendix 3.5

THE SCHEME OF THE METHODS TO DISSOLVE CHEMICALS FOR THE IL-8 LUC ASSAY.

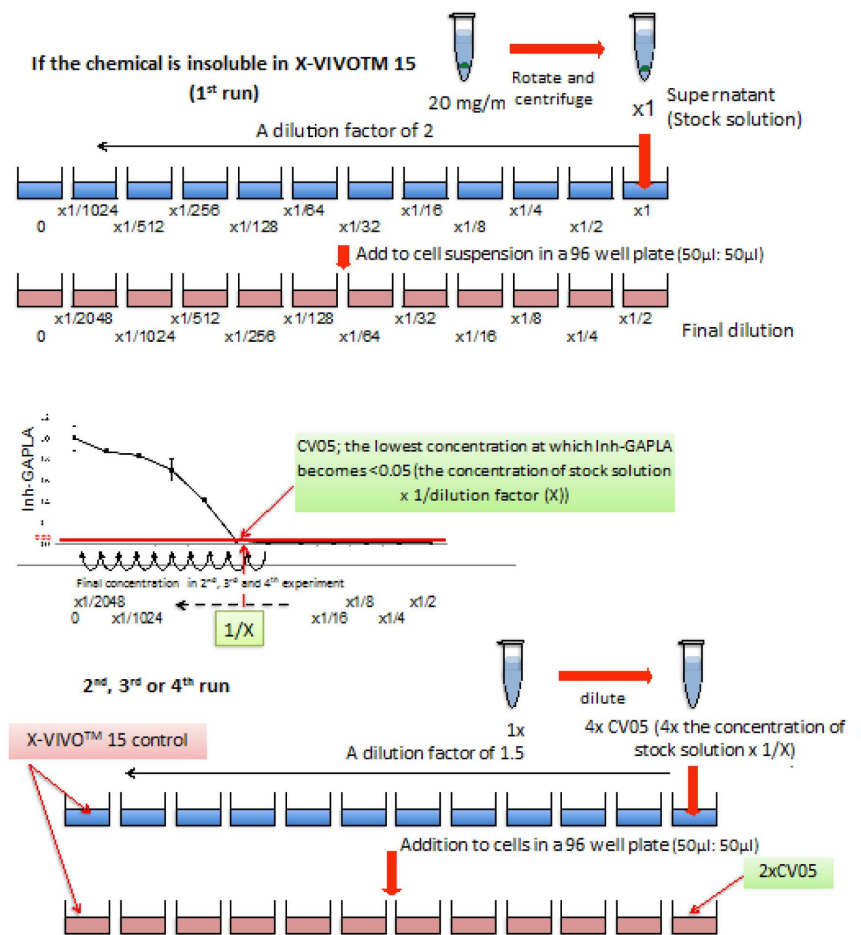
(a) For chemicals dissolved in X-VIVO™ 15 at 20 mg/ml



Determine the highest concentration of the following experiments



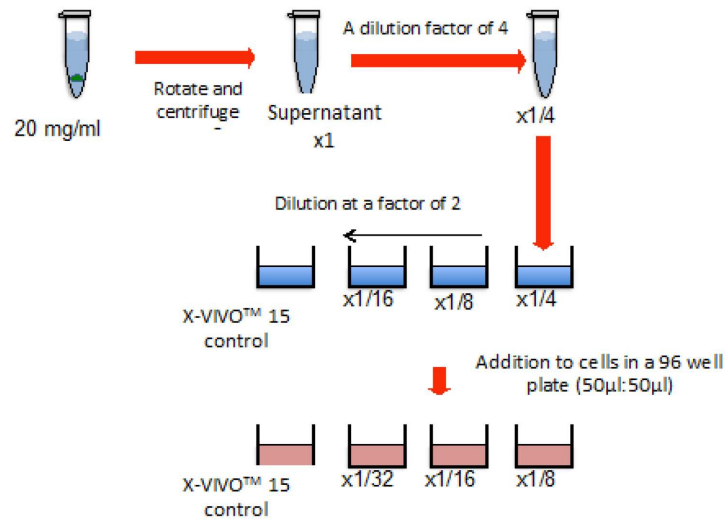
(b) For chemicals insoluble in X-VIVOTM 15 at 20 mg/ml



Appendix 3.6

THE SCHEME OF THE METHOD TO DISSOLVE 4-NBB FOR THE POSITIVE CONTROL OF THE IL-8 LUC ASSAY

The positive control : 4-NBB (insoluble in X-VIVO™ 15)



- (9) In Part C, the following Chapters are added:

"C.52 MEDAKA EXTENDED ONE GENERATION REPRODUCTION TEST (MEOGRT)

INTRODUCTION

1. This test method is equivalent to OECD test guideline (TG) 240 (2015). The Medaka Extended One Generation Test (MEOGRT) describes a comprehensive test method based on fish exposed over multiple generations to give data relevant to ecological hazard and risk assessment of chemicals, including suspected endocrine disrupting chemicals (EDCs). Exposure in the MEOGRT continues until hatching (until two weeks post fertilisation, wpf) in the second (F2) generation. Additional investigations would be needed to justify the utility of extending the F2 generation beyond hatching; at this time, there is insufficient information to provide relevant conditions or criteria for warranting the extension of the F2 generation. However, this test method may be updated as new information and data are considered. For example, guidance on extending the F2 generation through reproduction may be potentially useful under certain circumstances (e.g., chemicals with high bioconcentration potential or indications of trans-generational effects in other taxa). This test method can be used to evaluate the potential chronic effects of chemicals, including potential endocrine disrupting chemicals, on fish. The method gives primary emphasis to potential population relevant effects (namely, adverse impacts on survival, development, growth and reproduction) for the calculation of a No Observed Effect Concentration (NOEC) or an Effect Concentration (ECx), although it should be noted that ECx approaches are rarely suitable for large studies of this type where increasing the number of test concentrations to allow for determination of the desired ECx may be impractical which may also cause significant animal welfare concerns due to the large number of animals used. For chemicals not requiring assessment over "multi-generations" or chemicals that are not potential endocrine disrupting chemicals, other test methods may be more appropriate (1). The Japanese medaka is the appropriate species for use in this test method, given its short life-cycle and the possibility to determine its genetic sex (2), which is considered a critical component in this test method. The specific methods and observational endpoints detailed in this method are applicable to Japanese medaka alone. Other small fish species (e.g., zebrafish) may be adapted to a similar test protocol.
2. This test method measures several biological endpoints. Primary emphasis is given to potential adverse effects on population relevant parameters including survival, gross development, growth and reproduction. Secondly, in order to provide mechanistic information and provide linkage between results from other kinds of field and laboratory studies, where there is *a posteriori* evidence for a chemical having potential endocrine disrupter activity (e.g. androgenic or oestrogenic activity in other tests and assays) then other useful information is obtained by measuring *vitellogenin* (*vtg*) mRNA (or vitellogenin protein, VTG), phenotypic secondary sex characteristics (SSC) as related to genetic sex, and evaluating histopathology. It should be noted that if a test chemical or its metabolites are not suspected of being EDCs, it may not be necessary to measure these secondary endpoints and less resource and animal intensive studies may be more appropriate (1). Definitions used in this test method are given in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

3. Due to the limited number of chemicals tested and laboratories involved in the validation of this rather complex assay, it is anticipated that when a sufficient number of studies is available to ascertain the impact of this new study design, the test method will be reviewed and if necessary revised in light of experience gained. The data can be used at Level 5 of the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters (3). The test method begins by exposing adult fish (the F0 generation) to the test chemical during the reproduction phase. The exposure continues through development and reproduction in the F1 and hatch in the F2 generation; thus the assay allows evaluation of both structural and activational endocrine pathways. A weight of evidence approach may be undertaken when interpreting the endocrine related endpoints.
4. The test should include an adequate number of individuals to ensure sufficient power for the evaluation of reproduction-relevant endpoints (see Appendix 3) whilst ensuring that the number of animals used is the minimum required for animal welfare reasons. In view of the large numbers of test animals used, it is important to carefully consider the need for the test in relation to existing data which may already contain relevant information on many of the endpoints in the MEOGRT. Some assistance in this regard can be obtained from the OECD Fish Toxicity Testing Framework (1).

5. The test method has been designed primarily to distinguish the effects of a single substance. However, if a test on a mixture is required, then it should be considered whether it will provide acceptable results for the intended regulatory purpose.
6. Before beginning the test, it is important to have information about the physicochemical properties of the test chemical, particularly to allow the production of stable chemical solutions. It is also necessary to have an adequately sensitive analytical method for verifying test chemical concentrations.

PRINCIPLE OF THE TEST

7. The test is started by exposing sexually mature males and females (at least 12 wpf) in breeding pairs for 3 weeks, during which the test chemical is distributed in the organism of the parental generation (F0) according to its toxicokinetic behaviour. As near as possible to the first day of the fourth week, eggs are collected to start the F1 generation. During rearing of the F1 generation (a total of 15 weeks), hatchability and survival are assessed. In addition, fish are sampled at 9-10 wpf for developmental endpoints and spawning is assessed for three weeks from 12 through 14 wpf. An F2 generation is started after the third week of the reproduction assessment and reared until completion of hatching.

TEST VALIDITY CRITERIA

8. The following criteria for test validity apply:

- The dissolved oxygen concentration should be $\geq 60\%$ of air saturation value throughout the test;
- The mean water temperature over the entire duration of the study should be between 24 and 26°C. Brief excursions from the mean by individual aquaria should not be more than 2°C;
- The mean fecundity of controls in each of the generations (F0 and F1) should be greater than 20 eggs per pair per day. Fertility of all the eggs produced during the assessment should be greater than 80%. In addition, 16 of the recommended 24 control breeding pairs ($> 65\%$) should produce greater than 20 eggs per pair per day;
- Hatchability of eggs should be $\geq 80\%$ (average) in the controls (in each of the F1 and F2 generations);
- Survival after hatching until 3 wpf and from 3 wpf through termination for the generation F1 (i.e. 15 wpf) should be $\geq 80\%$ (average) and $\geq 90\%$ (average), respectively in the controls (F1);
- Evidence should be available to demonstrate that the concentrations of the test chemical in solution have been satisfactorily maintained within $\pm 20\%$ of the mean measured values.

Regarding water temperature, while not a validity criterion, replicates within a treatment should not be statistically different from each other, and treatment groups within the test should not be statistically different from each other (based on daily temperature measurements, and excluding brief excursions).

9. Although decreased reproduction may be observed in the higher exposure groups there should be sufficient reproduction in at least the third highest group and all lower groups of F0 to fill the hatching incubators. Furthermore, there should be adequate embryo survival in the third highest and lower exposure groups in F1 to allow endpoint evaluation at the sub-adult sampling (see paragraphs 36 and 38 and Appendix 9). Additionally, there should be at least minimal post-hatch survival ($\sim 20\%$) in the second highest exposure group of F1. These are not validity criteria, as such, but recommendations to permit robust NOECs to be calculated.

10. If a deviation from the test validity criteria is observed, the consequences should be considered in relation to the reliability of the test results and these deviations and considerations should be included in the test report.

DESCRIPTION OF THE METHOD

Apparatus

11. Normal laboratory equipment and especially the following:
 - (a) oxygen and pH meters;
 - (b) equipment for determination of water hardness and alkalinity;
 - (c) adequate apparatus for temperature control and preferably continuous monitoring;
 - (d) tanks made of chemically inert material and of a suitable capacity in relation to the recommended loading and stocking density (see Appendix 3);
 - (e) suitably accurate balance (i.e. accurate to ± 0.5 mg).

Water

12. Any water in which the test species shows suitable long-term survival and growth may be used as test water. It should be of constant quality during the period of the test. In order to ensure that the dilution water will not unduly influence the test result (for example by complexation of test chemical) or adversely affect the performance of the brood stock, samples should be taken at intervals for analysis. Measurements of heavy metals (e.g. Cu, Pb, Zn, Hg, Cd, Ni), major anions and cations (e.g. Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-}), pesticides, total organic carbon and suspended solids should be made, for example, every six months where a dilution water is known to be relatively constant in quality. Some chemical characteristics of acceptable dilution water are listed in Appendix 2. The pH of the water should be within the range 6.5 to 8.5, but during a given test it should be within a range of ± 0.5 pH units.

Exposure system

13. The design and materials used for the exposure system are not specified. Glass, stainless steel, or other chemically inert material should be used for construction of the test system that has not been contaminated during previous tests. For the purpose of this test, a well-suited exposure system may consist of a continuous flow-through system (4)(5)(6)(7)(8)(9)(10)(11)(12)(13).

Test solutions

14. Stock solution of the test chemical should be delivered into the exposure system by an appropriate pump. The flow rate of the stock solution should be calibrated in accordance with analytical confirmation of the test solutions before the initiation of exposure, and checked volumetrically periodically during the test. The test solution in each chamber is renewed adequately (e.g., minimum of 5 volume renewals/day to up to 16 volume renewals/day or up to 20 ml/min flow) depending on the test chemical stability and water quality.

15. Test solutions of the chosen concentrations are prepared by dilution of a stock solution. The stock solution should preferably be prepared by simply mixing or agitating the test chemical in dilution water by mechanical means (e.g. stirring and/or ultra-sonication). Saturation columns/systems or passive dosing methods (14) can be used for achieving a suitably concentrated stock solution. All efforts should be made to avoid solvents or carriers because: (1) certain solvents themselves may result in toxicity and/or undesirable or unexpected responses, (2) testing chemicals above their water solubility (as can frequently occur through the use of solvents) can result in inaccurate determinations of effective concentrations, (3) the use of solvents in longer-term tests can result in a significant degree of "bio-filming" associated with microbial activity which may impact environmental conditions as well as the ability to maintain exposure concentrations and (4) in the absence of historical data that demonstrates that the solvent does not influence the outcome of the study, use of solvents requires a solvent control treatment which has animal welfare implications as additional animals are required to conduct the test. For difficult to test chemicals, a solvent may be employed as a last resort, and the OECD Guidance Document 23 on Aquatic Toxicity Testing of Difficult Substances and Mixtures (15) should be consulted to determine the best method. The choice of solvent will be determined by the chemical properties of the test chemical and the availability of historical data on use of the solvent. If solvent carriers are used, appropriate solvent controls should be evaluated in addition to non-solvent (negative) controls (dilution water only). In the event that use of a solvent is unavoidable, and microbial activity (bio-filming) occurs, recommend recording/reporting of the bio-filming per tank (at least weekly) throughout the test. Ideally, the solvent concentration should be kept constant in the solvent control and all test treatments. If the concentration of solvent is not kept constant, the highest concentration of solvent in the test treatment should be used in the solvent control. In cases where solvent carrier is used, maximum solvent concentrations should not exceed 100 µl/l or 100 mg/l (15), and it is recommended to keep solvent concentration as low as possible (e.g. < 20 µl/l) to avoid potential effect of the solvent on endpoints measured (16).

Test animals

Selection and holding of fish

16. The test species is Japanese medaka *Oryzias latipes* because of its short life-cycle and the possibility to determine genetic sex. Although other small fish species may be adapted to a similar test protocol, the specific methods and observational endpoints detailed in this test method are applicable to Japanese medaka alone (see paragraph 1). The medaka is readily induced to breed in captivity; published methods exist for its culture (17) (18) (19), and data are available from short-term lethality, early life-stage and full life-cycle tests (5) (6) (8) (9) (20). All fish are maintained on a 16 h light:8 h dark photoperiod. The fish will be fed live brine shrimp, *Artemia* spp., nauplii which may be supplemented with a commercially available flake food if necessary. Commercially available flake food should be regularly analysed for contaminants.
17. As long as appropriate husbandry practices are followed, no specific culturing protocol is required. For example, medaka can be reared in 2 l tanks with 240 larval fish per tank until 4 wpf, then they can be reared in 2 l tanks with 10 fish per tank until 8 wpf, at which time, they transition to breeding pairs in 2 l tanks.

Acclimation and selection of fish

18. Test fish should be selected from a single laboratory stock which has been acclimated for at least two weeks prior to the test under conditions of water quality and illumination similar to those used in the test (Note: This acclimation period is not an *in situ* pre-exposure period). It is recommended that test fish be obtained from an in-house culture, as shipping of adult fish is stressful and may interfere with reliable spawning. Fish should be fed brine shrimp nauplii twice per day throughout the holding period and during the exposure phase, supplemented with a commercially available flake food if necessary. A minimum of 42 breeding pairs (54 breeding pairs if a solvent control is required due, in part, to lack of historical data to support the use of only the non-solvent control) are considered necessary to initiate this test to ensure adequate replication. In addition, each breeding pair of F0 should be verified to be XX-XY (i.e. normal complement of sex chromosomes in each sex) to avoid the possible inclusion of spontaneous XX males (see paragraph 39).
19. During the acclimation phase, mortalities in the culture fish should be recorded and the following criteria applied following a 48 h settling-down period:

- Mortalities of greater than 10 % of the culture population in seven days preceding transfer to the test system: reject the entire batch;

- Mortalities of between 5 % and 10 % of the population in the seven days preceding transfer to the test system: acclimation for seven additional days to the 2-week acclimation period; if more than 5 % mortality during the second seven days, reject the entire batch;
 - Mortalities of less than 5 % of the population in the seven days preceding transfer to the test system: accept the batch.
20. Fish should not receive treatment for disease in the two-week acclimation period preceding the test and during the exposure period, and disease treatment should be completely avoided if possible. Fish with clinical signs of disease should not be used in the study. A record of observations and any prophylactic and therapeutic disease treatments during the culture period preceding the test should be maintained.
21. The exposure phase should be started with sexually dimorphic, genetically sexed adult fish from a laboratory supply of reproductively mature animals cultured at 25 ± 2 °C. The fish should be identified as proven breeders (i.e. having produced viable offspring) during the week preceding exposure. For the whole group of fish used in the test, the range in individual weights by sex at the start of the test should be kept within ± 20 % of the arithmetic mean weight of the same sex. A subsample of fish should be weighed before the test to estimate the mean weight. The fish selected should be at least 12 wpf, being a weight ≥ 300 mg for females and ≥ 250 mg for males.

TEST DESIGN

Test concentrations

22. It is recommended to use five chemical concentrations plus control(s). All sources of information should be considered when selecting the range of test concentrations, including quantitative structure activity relationships (QSARs), read-across from analogues, results of fish tests such as acute toxicity assays (Chapter C.1 of this Annex), fish short-term reproduction assay (Chapter C.48 of this Annex) and other test methods e.g. Chapters C.15, C.37, C.41, C.47 or C.49 of this Annex (21) (22) (23) (24) (25) (26) if available, or if necessary, from a range-finding test possibly including a reproduction phase. If needed, the range-finding test may be conducted under conditions (water quality, test system, animal loading) similar to those used for the definitive test. If use of a solvent is necessary and no historical data are available, the range-finding test can be used to identify suitability of the solvent. The highest test concentration should not exceed the water solubility, 10 mg/l or $1/10^{\text{th}}$ of the 96h-LC50 (27). The lowest concentration should be a factor of 10- to 100-times lower than the highest concentration. The use of five concentrations in this test enables not only dose-response relationships to be measured, but also provides the Lowest Observed Effect Concentration (LOEC) and NOEC which are necessary for risk assessment in some regulatory programmes or jurisdictions. Generally, the spacing factor between nominal concentrations of the test chemical between adjacent treatment levels is ≤ 3.2 .

Replicates within treatment groups and controls

23. A minimum of six replicate test chambers per test concentration should be used (see Appendix 7). During the reproductive phase (except F0 generation), replication structure is doubled for fecundity assessment and each replicate has only one breeding pair (see paragraph 42).
24. A dilution water control and, if needed, a solvent control should be run in addition to the test concentrations. A doubled number of replicate chambers for the controls should be used to ensure adequate statistical power (i.e., at least twelve replicates should be used for controls). During the reproductive phase, the number of replicates in the controls are doubled (i.e. 24 replicates as a minimum and each replicate has only one mating pair). Following reproduction, control replicates should contain no more than 20 embryos (fish).

PROCEDURE

Initiation of test

25. The reproductively active adult fish used to start the F0 generation of the test are selected based on two criteria: age (typically more than 12 wpf but recommended not to exceed 16 wpf) and weight (should be ≥ 300 mg for females and ≥ 250 mg for males).

26. Female-male pairs that meet the above specifications are moved as individual pairs into each tank replicate, i.e. twelve replicates in controls and six replicates in chemical treatments at the initiation of the test. These tanks are randomly assigned a treatment (e.g., T1-T5 and control) and a replicate (e.g., A-L in controls and A-F in treatment), and then placed in the exposure system with the appropriate flow to each tank.

Conditions of exposure

27. A complete summary of test parameters and conditions can be found in Appendix 3. Adherence to these specifications should result in control fish with endpoint values similar to those listed in Appendix 4.
28. During the test, dissolved oxygen, pH, and temperature should be measured in at least one test vessel of each treatment group and the control. As a minimum, these measurements, except temperature, should be made once a week through the exposure period. The mean water temperature over the entire duration of the study should be between 24 and 26°C throughout the test. Temperature should be measured every day throughout the exposure period. The pH of the water should be within the range 6.5 to 8.5, but during a given test it should be within a range of ± 0.5 pH units. Replicates within a treatment should not be statistically different from each other, and treatment groups within the test should not be statistically different from each other (based on daily temperature measurements, and excluding brief excursions).

Duration of exposure

29. The test exposes sexually reproductive fish from F0 for three weeks. In week 4 on approximately test day 24, F1 is established and the F0 breeding pairs are humanely killed and weight and length are recorded (see Paragraph 34). This is followed by exposure of the F1 generation for 14 more weeks (total of 15 weeks for F1) and the F2 generation for two weeks until hatching. The total duration of the test is primarily 19 weeks (i.e., until F2 hatching). Timelines for the test are shown in Table 2 and further explained in detail in Appendix 9.

Feeding regime

30. Fish can be fed brine shrimp *Artemia* spp. (24-hours old nauplii) *ad libitum*, supplemented with a commercially available flake food if necessary. Commercially available flake food should be regularly analysed for contaminants such as organochlorine pesticides, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs). Food with an elevated level of endocrine active substances (i.e., phytoestrogens) that could compromise the response of the test should be avoided. Uneaten food and faecal material should be removed from the test vessels as required, e.g. by carefully cleaning the bottom of each tank using a siphon. The sides and bottom of each tank should also be cleaned once or twice per week (e.g., by scraping with a spatula). An example of a feeding schedule can be found in Appendix 5. Feeding rate is based upon number of fish per replicate. Therefore, feeding rate is reduced if there are mortalities in a replicate.

Analytical determination and measurements

31. Prior to initiation of the exposure period, proper function of the chemical delivery system should be ensured. All analytical methods needed should be established, including sufficient knowledge of the chemical's stability in the test system. During the test, the concentrations of the test chemical are determined at appropriate intervals, preferably at least every week in one replicate for each treatment group, rotating between replicates of the same treatment group every week.
32. During the test, the flow rates of diluent and stock solution should be checked at intervals accordingly (e.g. at minimum three times a week). It is recommended that results be based on measured concentrations. However, if concentration of the test chemical in solution has been satisfactorily maintained within $\pm 20\%$ of the mean measured values throughout the test, then the results can either be based on nominal or measured values. In case of chemicals that markedly accumulate in fish, the test concentrations may decrease as the fish grow. In such cases, it is recommended that the renewal rate of the test solution in each chamber be adapted to maintain test concentrations as constant as possible.

Observations and measured endpoints

33. Endpoints measured include fecundity, fertility, hatching, growth and survival for evaluation of possible population-level effects. Observations of behaviour should also be made daily, and any unusual behaviour noted. Other mechanistic endpoints include hepatic vtg mRNA or VTG protein levels by an immunoassay (28), sexual phenotypic markers such as characteristic male anal fin papillae, histological evaluation of gonadal sex, and histopathological evaluation of kidney, liver and gonad (see endpoint list in Table 1). All of these specific endpoints are evaluated in the context of a determination of the genetic sex of the individual, based on the presence or absence of the medaka male-sex determining gene *dmy* (see paragraph 41). Additionally, time to spawn is also evaluated. In addition, simple phenotypic sex ratios can be derived using the information from counts of anal fin papillae to define individual medaka as either phenotypically male or female. This test method would not be expected to detect modest deviations from the expected sex ratio because the relatively small numbers of fish per replicate will not provide sufficient statistical power. Also, during the course of the histopathological assessment, the gonad is evaluated and much more powerful analyses for assessing the gonad phenotype in the context of the genetic sex are conducted.
34. The primary purpose of this test method is to assess the potential population relevant effects of a test chemical. Mechanistic endpoints (VTG, SSCs and certain gonadal histopathology effects) can also assist in determining whether any effect is mediated via endocrine activity. However, these mechanistic endpoints can also be influenced by systemic and other toxicities. Consequently, liver and kidney histopathology may also be assessed in detail to help better understand any responses in mechanistic endpoints. However, if these detailed evaluations are not performed, gross abnormalities observed incidentally during the histopathological evaluation should still be noted and reported.

Humane killing of fish

35. At termination of F0 and F1 generation exposure, and when sub-adult fish are subsampled, the fish should be euthanised with appropriate amounts of anaesthetic solution (e.g. Tricaine methane sulfonate, MS-222 (CAS.886-86-2), 100-500 mg/l) buffered with 300 mg/l NaHCO₃ (sodium bicarbonate, CAS.144-55-8) to reduce mucous membrane irritation. If fish are showing signs of considerable suffering (very severe and death can be reliably predicted) and considered moribund, animals should be anaesthetised and euthanised and treated as mortality for data analysis. When a fish is euthanised due to morbidity, this should be noted and reported. Depending on when the fish is euthanised during the study, retaining the fish for histopathology analysis may be conducted (fixing the fish for possible histopathology).

Handling of eggs and larval fish

Collection of eggs from breeding pairs to propagate the next generation

36. Egg collection is done on the first day (or first two days, if needed) of Test Week 4 to go from F0 to F1 and Test Week 18 to go from F1 to F2. Test Week 18 corresponds to F1, 15 wpf (weeks post fertilisation) adult fish. It is important that all eggs are removed from each tank the day before the egg collection starts to ensure all eggs collected from a breeding pair are from a single spawn. Following spawning, female medaka sometimes carry their eggs near the vent until the eggs can be deposited onto a substrate. With no substrate present in the tank, the eggs can be found either attached to the female or at the bottom of the tank. Depending on their location, eggs are either carefully removed from the female or siphoned from the bottom in Test Week 4 of F0 and Test Week 18 of F1. All eggs collected within a treatment are pooled prior to distribution to incubation chambers.
37. Egg filaments, which hold spawned eggs together, should be removed. Fertilised eggs (up to 20) are collected from each breeding pair (1 pair per replicate), are pooled by treatment, and systematically distributed to suitable incubation chambers (Appendix 6, 7). Using a good quality dissecting microscope, one can see hallmarks of early fertilisation/development such as raising of the fertilisation membrane (chorion), ongoing cell division, or formation of the blastula. The incubator chambers may be placed in separate "incubator aquaria" set up for each treatment (in which case water quality parameters and test chemical concentrations need to be measured in these), or in the replicate aquarium in which hatched larvae (e.g., eleutheroembryo) will be contained. If a second day of collection (Test Day 23) is needed, all eggs from both days should be pooled and then systematically redistributed to each of the treatment replicates.

Rearing of eggs to hatching

38. Fertilised eggs are continually agitated *e.g.*, within the egg incubator by air bubbles or by vertically swinging the egg incubator. The mortalities of fertilised eggs (embryos) are checked and recorded daily. Dead eggs are removed from the incubators (Appendix 9). On the 7th day post fertilisation (dpf), the agitation is stopped or reduced so the fertilised eggs settle to the bottom of the incubator. This promotes hatching, typically over the next one or two days. For each treatment and control, hatchlings (young larvae; eleutheroembryo) are counted (pooled replicate basis). Fertilised eggs that have not hatched by twice the median day of hatch in the control (typically 16 or 18 dpf) are considered non-viable and discarded.
39. Twelve hatchlings are transferred into each replicate tank. The hatchlings from the incubation chambers are pooled and systematically distributed to replicate tanks (Appendix 7). This can be done by randomly selecting a hatchling from the treatment pool and sequentially adding a hatchling in an indiscriminate draw to a replicate aquarium. Each of the tanks should contain an equal number ($n=12$) of the hatched larvae (maximum 20 larvae each). If there are not enough hatchlings to fill all treatment replicates, then it is recommended to ensure as many replicates as possible have 12 hatchlings. Hatchlings can be handled safely with large-bore glass pipettes. Any additional hatchlings are humanely killed with anaesthetic. During the few weeks prior to the setup of breeding pairs, the day that the first spawning event is observed in each replicate should be recorded.

Setup of breeding pairs

Fin clipping and determination of genotypic sex

40. Determination of genotypic sex via fin clips is done at 9-10 wpf (*i.e.*, Test Week 12-13 for F1 generation). All fish within a tank are anaesthetised (using approved methods, *e.g.*, IACUC) and a small tissue sample is taken from either the dorsal or the ventral tip of the caudal fin of each fish to determine the genotypic sex of the individual (29). The fish from a replicate can be housed in small cages, if possible one per cage, in the replicate tank. Alternatively, two fish can be held in each cage if they are distinguishable from each other. One method is to differentially cut the caudal fin (*e.g.*, dorsal vs ventral tip) when taking the tissue sample.
41. The genotypic sex of medaka is determined by an identified and sequenced gene (*dmy*) which is located on the Y chromosome. The presence of *dmy* indicates a XY individual, regardless of phenotype, while the absence of *dmy* indicates a XX individual, regardless of phenotype (30); (31). Deoxyribose nucleic acid (DNA) from each fin clip is extracted and the presence or absence of *dmy* can be determined by polymerase chain reaction (PCR) methods (refer to Appendix 9 in Chapter C.41 of this Annex, or Appendix 3 and 4 in (29)).

Establishment of breeding pairs

42. The information on genotypic sex is used to establish XX-XY breeding pairs regardless of external phenotype which may be altered by exposure to a test chemical. On the day after the genotypic sex of each fish is determined, two XX fish and two XY fish from each replicate are randomly selected and two XX-XY breeding pairs are established. If a replicate does not have either two XX or two XY fish, appropriate fish should be obtained from other replicates within the treatment. The priority is to have the recommended number of replicate breeding pairs (12) in each treatment and in the controls (24). Fish with obvious abnormalities (swim bladder problems, spinal deformities, extreme size variations, etc.) would be precluded when establishing breeding pairs. During the reproductive phase for F1 each replicate tank should contain only one breeding pair.

Sampling of sub-adults and endpoint assessment

Sampling of non-breeding pair fish

43. After the setup of breeding pairs, the fish not selected for further breeding are humanely killed for measurement of sub-adult endpoints in Test Week 12-13 (F1). It is extremely important that the fish are handled in such a way so that the genotypic sex determined for breeding pair selection can still be traced to an individual fish. All the data collected are analysed in the context of the genotypic sex of the specific fish. Each fish is used for a variety of endpoint measurements including: determination of survival rates of juvenile/sub-adult fish (Test Weeks 7-12/13(F1),

growth in length (standard length may be measured if the caudal fin has been shortened due to sampling for genetic sex analysis. Total length can be measured if only a portion of the caudal fin, dorsal or ventral, is sampled for *dmy*) and body mass (i.e., wet weight, blotted dry), liver *vtg* mRNA (or VTG) and anal fin papillae (see Tables 1 and 2). Note that weights and lengths of the breeding pairs are also required for calculating mean growth in a treatment group.

Tissue sampling and vitellogenin measurement

44. The liver is dissected, and should be stored at ≤ -70 °C until the *vtg* mRNA (or VTG) measurements. The tail of the fish, including the anal fin, is preserved in an appropriate fixative (e.g. Davidson's) or photographed so that anal fin papillae can be counted at a later date. If desired, other tissues (i.e., gonad) may be sampled and preserved at this time). Liver VTG concentration should be quantified with a homologous ELISA technique (see the recommended procedures for medaka in Appendix 6 in Chapter C.48 of this Annex). Alternatively, the methods for *vtg* mRNA quantification, i.e., *vtg I* gene mRNA extraction from a liver sample and quantification of the number of copies of the *vtg I* gene (per ng of total mRNA) by quantitative PCR, have been established by the U.S EPA (29). Instead of determining the number of copies of the *vtg* gene in the control and treatment groups, a more resource friendly and less technically difficult method is to determine the relative (fold) change in *vtg I* expression from control and treatment groups.

Secondary sex characteristics

45. Under normal circumstances, only sexually mature male medaka have papillae, which develop on the joint plates of certain anal fin rays as a secondary sexual characteristic, providing a potential biomarker for endocrine disrupting effects. The method of counting anal fin papillae (the number of joint plates with papillae) is given in Appendix 8. Also the number of anal fin papillae per individual is used to categorise that individual as externally phenotypic male or female for the purpose of calculating a simple sex ratio per replicate. A medaka with any number greater than 0 is defined as a male; a medaka with 0 anal fin papillae is defined as a female.

Fecundity and fertility assessment

46. Fecundity and fertility are assessed in Test Weeks 1 through 3 in the F0 generation and Test Weeks 15 through 17 in the F1 generation. Eggs are collected daily from each breeding pair for 21 consecutive days. Eggs are gently removed from netted females and/or siphoned from the bottom of the aquarium each morning. Both fecundity and fertility are recorded daily for each replicate breeding pair. Fecundity is defined as the number of eggs spawned, and fertility is functionally defined as the number of fertilised and viable eggs at the time of counting. Counting should be done as soon as possible after egg collection.
47. Replicate fecundity is recorded daily as the number of eggs per breeding pair which is analysed by the recommended statistical procedures using the replicate means. Replicate fertility is the sum of the number of fertile eggs produced by a breeding pair divided by the sum of the number of eggs produced by that pair. Statistically fertility is analysed as a ratio per replicate. Replicate hatchability is the number of hatchlings divided by the number of embryos loaded (typically 20). Statistically hatchability is analysed as a ratio per replicate.

Sampling of adults and endpoint assessment

Sampling of breeding pair fish

48. Following Test Week 17 (i.e., after the F2 generation has successfully commenced), the F1 adults are humanely killed and various endpoints are assessed (see Tables 1 and 2). The anal fin is imaged for assessing anal fin papillae (see Appendix 8), and/or the tail, just posterior to the vent, is removed and fixed for counting papillae later. A portion of the caudal fin may be sampled and archived at this time for verification of genetic sex (*dmy*) if desired. If needed, a tissue sample can be taken to repeat the *dmy* analysis to verify genetic sex of specific fish. The body cavity is opened to allow perfusion with appropriate fixatives (e.g., Davidson's) prior to submersing the entire body in the fixative. However, if an appropriate permeabilisation step is performed prior to fixation, the body cavity does not need to be opened.

Histopathology

49. Each fish is evaluated histologically for pathology in the gonadal tissue (30); (29). As referenced in paragraph 33, other mechanistic endpoints evaluated in this assay (i.e., VTG, SSCs and certain gonadal histopathology effects) may be influenced by systemic or other toxicities. Consequently, liver and kidney histopathology may also be assessed in detail to help better understand any responses in mechanistic endpoints. However, if these detailed evaluations are not performed, gross abnormalities observed incidentally during the histopathological evaluation should still be noted and reported. 'Reading down' from the highest treatment group (compared to the control) to a treatment with no effect could be considered, however, it is recommended to consult the histopathology guidance (29). Typically all samples are processed/sectioned after which are read by the pathologist. If using a 'read-down' approach, it is noted that the Rao-Scott Cochran-Armitage by Slices (RSCABS) procedure uses the expectation that as dose levels increase the biological impact (the pathology) will increase as well. Therefore, one will lose power if only looking at a single high dose without any intermediate doses. If statistical analysis is not necessary to determine that the high dose has no effect, then this approach may be acceptable. The gonad phenotype is also derived from this evaluation

Other observations

50. The MEOGRT provides data that can be used (e.g., in a weight of evidence approach) to simultaneously evaluate at least two general types of adverse outcome pathways (AOPs) ending in reproductive impairment: (a) endocrine-mediated pathways involving disruption of the hypothalamus-pituitary-gonadal (HPG) endocrine axis; and, (b) pathways that cause reductions in survival, growth (length and weight), and reproduction through non-endocrine mediated toxicity. Endpoints typically measured in chronic toxicity tests such as the full life-cycle test and the early life-stage test are also included in this test and can be used to evaluate the hazards posed by both non-endocrine mediated toxic modes of action and endocrine-mediated toxicity pathways. During the test, observations of behaviour should be made daily, and any unusual behaviour should be noted. In addition, any mortality should be recorded and survival to the culling of fish (test week 6/7), survival after the culling to the sub-adult sampling (in 9-10 wpf), and survival from the pairing to the sampling of adult fish should be calculated.

Table 1

Endpoint overview of the MEOGRT (*)

Life-stage	Endpoint	Generation
Embryo(2 wpf)	Hatch (% and time to hatch)	F1, F2
Juvenile(4 wpf)	Survival	F1
Subadult(9 or10 wpf)	Survival	F1
	Growth(length and weight)	
	Vitellogenin(mRNA or protein)	
	Secondary sex characteristics(anal fin papillae)	
	External sex ratio	
	Time to 1 st spawn	
Adult(12-14 wpf)	Reproduction(fecundity and fertility)	F0, F1
Adult(15 wpf)	Survival	F1
	Growth(length and weight)	
	Secondary sex characteristics(anal fin papillae)	
	Histopathology(gonad, liver, kidney)	

(*) These endpoints are to be statistically analysed

TIMELINE

51. A timeline for the MEOGRT illustrated in Table 2 shows the test. The MEOGRT includes 4 weeks of exposure to F0 adults and 15 weeks of exposure to the F1 generation, and exposure period for the second generation (F2), until hatching (2 wpf). Activity through the course of the MEOGRT is summarised in Appendix 9.

Table 2

Exposure and measurement endpoint timelines for the MEOGRT

MEOGRT Exposure and Endpoint Timeline																			
F0	1	2	3	4															
F1				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
F2																	1	2	
Test Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Lifestage Key					Embryo			Larvae			Juvenile			Subadult		Adult			
Endpoints																			
Fecundity	F ₀																	F ₁	
Fertility	F ₀																	F ₁	
Hatch					F ₁														F ₂
Survival						F ₁						F ₁							F ₁
Growth				F ₀								F ₁							F ₁
Vitellogenin												F ₁							
Secondary sex												F ₁							F ₁
Histopathology																			F ₁
Test Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
<ul style="list-style-type: none"> • Experimental design has 7 groups of replicates <ul style="list-style-type: none"> ○ 5 for test chemical treatments ○ 2 for control treatments (4 if solvent is used) • Within-group design <ul style="list-style-type: none"> ○ 12 replicates for reproduction, adult pathology and SSC (Wks 10 through to 18) ○ 6 replicates for hatch, survival, Vtg; and - subadult SSC and growth (Wks 1 through to 9) SSC: secondary sex characters; Wks: weeks; Vtg: vitellogenin																			

DATA AND REPORTING

Statistical analysis

52. Since genotypic sex is determined for all test fish, the data should be analysed for each genotypic sex separately (i.e., XY males and XX females). Failure to do this will greatly reduce the statistical power of any analysis. Statistical analyses of the data should preferably follow procedures described in the OECD document on Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (32). Appendix 10 provides further guidance to the Statistical Analysis.
53. The test design and selection of statistical tests should permit adequate power to detect changes of biological importance in endpoints where a NOEC is to be reported (32). Reporting of relevant effect concentrations and parameters may depend upon the regulatory framework. The percent change in each endpoint that it is important to detect or estimate should be identified. The experimental design should be tailored to allow that. It is not likely that the same percent change applies to all endpoints, nor is it likely that a feasible experiment can be designed that will meet these criteria for all endpoints, so it is important to focus on the endpoints which are important for the respective experiment in designing the experiment appropriately. A statistical flow diagram and guidance is available in Appendix 10 to help with the treatment of data and in the choice of the most appropriate statistical test or model to use. Other statistical approaches may be used, provided they are scientifically justified.

54. It will be necessary for variations to be analysed within each set of replicates using analysis of variance or contingency table procedures and sufficient appropriate statistical analysis methods used based on this analysis. In order to make a multiple comparison between the results at the individual concentrations and those for the controls, the step-down procedure (e.g., Jonckheere-Terpstra test) is recommended for continuous responses. Where the data are not consistent with a monotone concentration-response, Dunnett's test or Dunn's test should be used (after an adequate data transform, if necessary).
55. For fecundity, egg counts taken daily, but may be analysed as total egg counts or as a repeated measure. Appendix 10 provides the details of how this endpoint is analysed. For histopathology data which are in the form of severity scores, a new statistical test, Rao-Scott Cochran-Armitage by Slices (RSCABS), has been developed (33).
56. Any endpoints observed in chemical treatments that are significantly different from appropriate controls should be reported.

Data analysis considerations

Use of compromised treatment levels

57. Several factors are considered when determining whether a replicate or entire treatment demonstrates overt toxicity and should be removed from analysis. Overt toxicity is defined as >4 mortalities in any replicate between 3 wpf and 9 wpf that cannot be explained by technical error. Other signs of overt toxicity include haemorrhage, abnormal behaviours, abnormal swimming patterns, anorexia, and any other clinical signs of disease. For sub-lethal signs of toxicity, qualitative evaluations may be necessary, and should always be made in reference to the dilution water control group (clean water only). If overt toxicity is evident in the highest treatment(s), it is recommended that those treatments be censored from the analysis.

Solvent controls

58. The use of a solvent should only be considered as a last resort, when all other chemical delivery options have been considered. If a solvent is used, then a dilution water control should be run in concert. At the termination of the test, an evaluation of the potential effects of the solvent should be performed. This is done through a statistical comparison of the solvent control group and the dilution water control group. The most relevant endpoints for consideration in this analysis are growth determinants (weight), as these can be affected through generalised toxicities. If statistically significant differences are detected in these endpoints between the dilution water control and solvent control groups, best professional judgment should be used to determine if the validity of the test is compromised. If the two controls differ, the treatments exposed to the chemical should be compared to the solvent control unless it is known that comparison to the dilution water control is preferred. If there is no statistically significant difference between the two control groups it is recommended that the treatments exposed to the test chemical are compared with the pooled (solvent and dilution-water control groups), unless it is known that comparison to either the dilution-water or solvent control group only is preferred.

Test report

59. The test report should include the following:

Test chemical: physical nature and, where relevant, physicochemical properties;

— Chemical identification data.

Mono-constituent substance:

— physical appearance, water solubility, and additional relevant physicochemical properties;

— chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc. (including the organic carbon content, if appropriate).

Multi-constituent substance, UVCBs and mixtures:

- characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physico-chemical properties of the constituents.

Test species:

- Scientific name, strain if available, source and method of harvesting of the fertilised eggs and subsequent handling.

Test conditions:

- Photoperiod(s);
- Test design (e.g. chamber size, material and water volume, number of test chambers and replicates, number of hatchlings per replicates);
- Method of preparation of stock solutions and frequency of renewal (the solubilising agent and its concentration should be given, when used);
- Method of dosing the test chemical (e.g. pumps, diluting systems);
- The recovery efficiency of the method and the nominal test concentrations, the limit of quantification, the means of the measured values and their standard deviations in the test vessels and the method by which these were attained and evidence that the measurements refer to the concentrations of the test chemical in true solution;
- Dilution water characteristics: pH, hardness, temperature, dissolved oxygen concentration, residual chlorine levels (if measured), total organic carbon (if measured), suspended solids (if measured), salinity of the test medium (if measured) and any other measurements made;
- The nominal test concentrations, the means of the measured values and their standard deviations;
- Water quality within test vessels, pH, temperature (daily) and dissolved oxygen concentration;
- Detailed information on feeding (e.g. type of foods, source, amount given and frequency).

Results:

- Evidence that controls met the overall validation criteria;
- Data for the control (plus solvent control when used) and the treatment groups as follows, hatching (hatchability and time to hatch) for F1 and F2, post hatch survival for F1, growth (length and body weight) for F1, genotypic sex and sexual differentiation (e.g. secondary sex characteristics based on anal fin papillae and gonadal histology) for F1, phenotypic sex for F1, secondary sex characteristics (anal fin papillae) for F1 *vtg* mRNA (or VTG protein) for F1, histopathology assessment (gonad, liver and kidney) for F1 and reproduction (fecundity and fertility) for F0, F1; (see Tables 1 and 2).
- Approach for the statistical analysis (regression analysis or analysis of the variance) and treatment of data (statistical tests and models used);
- No observed effect concentration (NOEC) for each response assessed;

- Lowest observed effect concentration (LOEC) for each response assessed (at $p = 0,05$); EC_x for each response assessed, if applicable, and confidence intervals (e.g. 90 % or 95 %) and a graph of the fitted model used for its calculation, the slope of the concentration-response curve, the formula of the regression model, the estimated model parameters and their standard errors.
- Any deviation from this test method and deviations from the acceptance criteria, and considerations of potential consequences on the outcome of the test.

60. For the results of endpoint measurements, mean values and their standard deviations (on both replicate and concentration basis, if possible) should be presented.

LITERATURE

- (1) OECD (2012a). Fish Toxicity Testing Framework, Environment, Health and Safety Publications, Series on Testing and Assessment (No. 171), Organisation for Economic Cooperation and Development, Paris.
- (2) Padilla S, Cowden J, Hinton DE, Yuen B, Law S, Kullman SW, Johnson R, Hardman RC, Flynn K and Au DWT. (2009). Use of Medaka in Toxicity Testing. *Current Protocols in Toxicology* 39: 1-36.
- (3) OECD (2012b). Guidance Document on Standardised Test Guidelines for Evaluating Endocrine Disrupters. Environment, Health and Safety Publications, Series on Testing and Assessment (No. 150), Organisation for Economic Cooperation and Development, Paris.
- (4) Benoit DA, Mattson VR, Olson DL. (1982). A Continuous-Flow Mini-Diluter System for Toxicity Testing. *Water Research* 16: 457-464.
- (5) Yokota H, Tsuruda Y, Maeda M, Oshima Y, Tadokoro H, Nakazono A, Honjo T and Kobayashi K. (2000). Effect of Bisphenol A on the Early Life Stage in Japanese Medaka (*Oryzias Latipes*). *Environmental Toxicology and Chemistry* 19: 1925-1930.
- (6) Yokota H, Seki M, Maeda M, Oshima Y, Tadokoro H, Honjo T and Kobayashi K. (2001). Life-Cycle Toxicity of 4-Nonylphenol to Medaka (*Oryzias Latipes*). *Environmental Toxicology and Chemistry* 20: 2552-2560.
- (7) Kang JJ, Yokota H, Oshima Y, Tsuruda Y, Yamaguchi T, Maeda M, Imada N, Tadokoro H and Honjo T. (2002). Effects of 17-Estradiol on the Reproduction of Japanese Medaka (*Oryzias Latipes*). *Chemosphere* 47: 71-80.
- (8) Seki M, Yokota H, Matsubara H, Tsuruda Y, Maeda M, Tadokoro H and Kobayashi K. (2002). Effect of Ethinylestradiol on the Reproduction and Induction of Vitellogenin and Testis-Ova in Medaka (*Oryzias Latipes*). *Environmental Toxicology and Chemistry* 21: 1692-1698.
- (9) Seki M, Yokota H, Matsubara H, Maeda M, Tadokoro H and Kobayashi K. (2003). Fish Full Life-Cycle Testing for the Weak Estrogen 4-Tert-Pentylphenol on Medaka (*Oryzias Latipes*). *Environmental Toxicology and Chemistry* 22: 1487-1496.
- (10) Hirai N, Nanba A, Koshio M, Kondo T, Morita M and Tatarazako N. (2006a). Feminization of Japanese Medaka (*Oryzias latipes*) Exposed to 17 β -Estradiol: Effect of Exposure Period on Spawning Performance in Sex-Transformed Females. *Aquatic Toxicology* 79: 288-295.
- (11) Hirai N, Nanba A, Koshio M, Kondo T, Morita M and Tatarazako N. (2006b). Feminization of Japanese Medaka (*Oryzias latipes*) Exposed to 17 β -Estradiol: Formation of Testis-Ova and Sex-Transformation During Early-Ontogeny. *Aquatic Toxicology* 77: 78-86.

- (12) Nakamaura A, Tamura I, Takanobu H, Yamamuro M, Iguchi T and Tatarazako N. (2015). Fish Multigeneration Test with Preliminary Short-Term Reproduction Assay for Estrone Using Japanese Medaka (*Oryzias Latipes*). *Journal of Applied Toxicology* 35:11-23.
- (13) U.S. Environmental Protection Agency (2013). Validation of the Medaka Multigeneration Test: Integrated Summary Report. Available at: <http://www.epa.gov/scipoly/sap/meetings/2013/062513meeting.html>.
- (14) Adolfsson-Erici M, Åkerman G, Jahnke A, Mayer P and McLachlan M. (2012). A Flow-Through Passive Dosing System for Continuously Supplying Aqueous Solutions of Hydrophobic Chemicals to Bioconcentration and Aquatic Toxicity Tests. *Chemosphere* 86: 593-599.
- (15) OECD (2000). Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. OECD Environment, Health and Safety Publications, Series on Testing and Assessment (No. 23.), Organisation for Economic Cooperation and Development, Paris.
- (16) Hutchinson TH., Shillabeer N., Winter MJ and Pickford DB. (2006). Acute and Chronic Effects of Carrier Solvents in Aquatic Organisms: A Critical Review. *Review. Aquatic Toxicology* 76: 69-92.
- (17) Denny JS, Spehar RL, Mead KE and Yousuff SC. (1991). Guidelines for Culturing the Japanese Medaka, *Oryzias latipes*. US EPA/600/3-91/064.
- (18) Koger CS, Teh SJ and Hinton DE. (1999). Variations of Light and Temperature Regimes and Resulting Effects on Reproductive Parameters in Medaka (*Oryzias Latipes*). *Biology of Reproduction* 61: 1287-1293.
- (19) Kinoshita M, Murata K, Naruse K and Tanaka M. (2009). *Medaka: Biology, Management, and Experimental Protocols*, Wiley- Blackwell.
- (20) Gormley K and Teather K. (2003). Developmental, Behavioral, and Reproductive Effects Experienced by Japanese Medaka in Response to Short-Term Exposure to Endosulfan. *Ecotoxicology and Environmental Safety* 54: 330-338.
- (21) Chapter C.15 of this Annex, Fish, Short-term Toxicity Test on Embryo and Sac-fry Stages.
- (22) Chapter C.37 of this Annex, 21-day Fish Assay: A Short-Term Screening for Oestrogenic and Androgenic Activity, and Aromatase Inhibition.
- (23) Chapter C.41 of this Annex, Fish Sexual Development Test.
- (24) Chapter C.48 of this Annex, Fish Short Term Reproduction Assay.
- (25) Chapter C.47 of this Annex, Fish, Early-life Stage Toxicity Test.

-
- (26) Chapter C.49 of this Annex, Fish Embryo Acute Toxicity (FET) Test.
- (27) Wheeler JR, Panter GH, Weltje L and Thorpe KL. (2013). Test Concentration Setting for Fish *In Vivo* Endocrine Screening Assays. *Chemosphere* 92: 1067-1076.
- (28) Tatarazako N, Koshio M, Hori H, Morita M and Iguchi T. (2004). Validation of an Enzyme-Linked Immunosorbent Assay Method for Vitellogenin in the Medaka. *Journal of Health Science* 50: 301-308.
- (29) OECD (2015). Guidance Document on Medaka Histopathology Techniques and Evaluation. Environment, Health and Safety Publications, Series on Testing and Assessment (No. 227). Organisation for Economic Cooperation and Development, Paris.
- (30) Nanda I, Hornung U, Kondo M, Schmid M and Schartl M. (2003). Common Spontaneous Sex-Reversed XX Males of the Medaka *Oryzias Latipes*. *Genetics* 163: 245-251.
- (31) Shinomiya, A, Otake H, Togashi K, Hamaguchi S and Sakaizumi M. (2004). Field Survey of Sex-Reversals in the Medaka, *Oryzias Latipes*: Genotypic Sexing of Wild Populations, *Zoological Science* 21: 613-619.
- (32) OECD (2014). Current Approaches in the Statistical Analysis of Ecotoxicity Data: A guidance to application (annexes to this publication exist as a separate document), OECD Publishing, Paris.
- (33) Green JW, Springer TA, Saulnier AN and Swintek J. (2014). Statistical Analysis of Histopathology Endpoints. *Environmental Toxicology and Chemistry* 33: 1108-1116.

Appendix 1

DEFINITIONS

Chemical: A substance or a mixture.

ELISA: Enzyme-Linked Immunosorbent Assay

Fecundity = number of eggs;

Fertility = number of viable eggs/fecundity;

Fork length (FL) refers to the length from the tip of the snout to the end of the middle caudal fin rays and is used in fishes in which it is difficult to tell where the vertebral column ends www.fishbase.org

Hatchability = hatchlings/number of embryos loaded into an incubator

IACUC: Institutional Animal Care and Use Committee

Standard length (SL) refers to the length of a fish measured from the tip of the snout to the posterior end of the last vertebra or to the posterior end of the midlateral portion of the hypural plate. Simply put, this measurement excludes the length of the caudal fin. (www.fishbase.org)

Total length (TL) refers to the length from the tip of the snout to the tip of the longer lobe of the caudal fin, usually measured with the lobes compressed along the midline. It is a straight-line measure, not measured over the curve of the body (www.fishbase.org)

Figure 1

Description of the different lengths, used

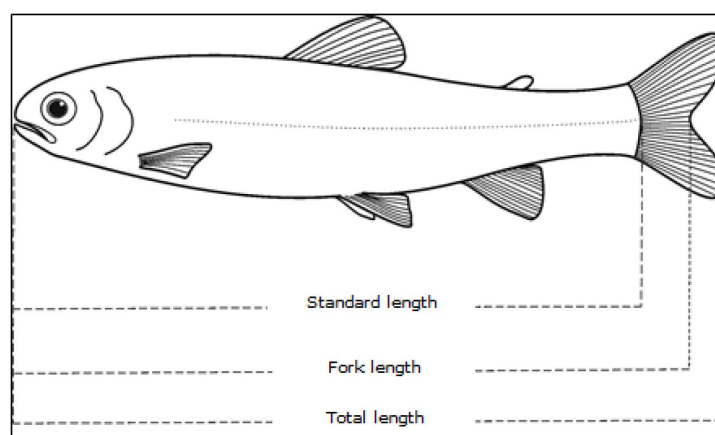


Figure 1: Description of the different lengths, used

EC_x: (Effect concentration for x % effect) is the concentration that causes an x % of an effect on test organisms within a given exposure period when compared with a control. For example, an EC₅₀ is a concentration estimated to cause an effect on a test end point in 50 % of an exposed population over a defined exposure period.

Flow-through test is a test with continued flow of test solutions through the test system during the duration of exposure.

HPG axis: hypothalamic-pituitary-gonadal axis.

IUPAC: International Union of Pure and Applied Chemistry.

Loading rate: The wet weight of fish per volume of water.

Lowest observed effect concentration (LOEC) is the lowest tested concentration of a test chemical at which the chemical is observed to have a statistically significant effect (at $p < 0,05$) when compared with the control. However, all test concentrations above the LOEC should have a harmful effect equal to or greater than those observed at the LOEC. When these two conditions cannot be satisfied, a full explanation should be given for how the LOEC (and hence the NOEC) has been selected. Appendix 5 and 6 provide guidance.

Median Lethal Concentration (LC50): is the concentration of a test chemical that is estimated to be lethal to 50 % of the test organisms within the test duration.

No observed effect concentration (NOEC) is the test concentration immediately below the LOEC, which when compared with the control, has no statistically significant effect ($p < 0,05$), within a stated exposure period.

SMILES: Simplified Molecular Input Line Entry Specification.

Stocking density: The number of fish per volume of water.

Test chemical: Any substance or mixture tested using this test method.

UVCB: Substances of unknown or variable composition, complex reaction products or biological materials.

VTG: Vitellogenin is a phospholipoglycoprotein precursor to egg yolk protein that normally occurs in sexually active females of all oviparous species.

WPF: Weeks post fertilisation

Appendix 2

SOME CHEMICAL CHARACTERISTICS OF AN ACCEPTABLE DILUTION WATER

Substance	Limit concentration
Particulate matter	5 mg/l
Total organic carbon	2 mg/l
Un-ionised ammonia	1 µg/l
Residual chlorine	10 µg/l
Total organophosphorous pesticides	50 ng/l
Total organochlorine pesticides plus polychlorinated biphenyls	50 ng/l
Total organic chlorine	25 ng/l
Aluminium	1 µg/l
Arsenic	1 µg/l
Chromium	1 µg/l
Cobalt	1 µg/l
Copper	1 µg/l
Iron	1 µg/l
Lead	1 µg/l
Nickel	1 µg/l
Zinc	1 µg/l
Cadmium	100 ng/l
Mercury	100 ng/l
Silver	100 ng/l

Appendix 3

TEST CONDITIONS FOR THE MEOGRT

1. Recommended species	Japanese medaka (<i>Oryzias latipes</i>)
2. Test type	Continuous flow-through
3. Water temperature	The nominal test temperature is 25°C. The mean temperature throughout the test in each tank is 24-26°C.
4. Illumination quality	Fluorescent bulbs (wide spectrum and ~150 lumens/m ²)(~150 lux).
5. Photoperiod	16 h light:8 h dark
6. Loading rate	F0: 2 adults/replicate; F1: initiated with maximum 20 eggs (embryos)/replicate, reduced to 12 embryos/replicate at hatch then 2 adults (XX-XY breeding pair) at 9-10 wpf for reproductive phase
7. Minimum test chamber usable volume	1,8 l (e.g., test chamber size: 18x9x15 cm)
8. Volume exchanges of test solutions	Minimum of 5 volume renewal/day to up to 16 volume renewal/day (or 20 ml/min flow)
9. Age of test organisms at initiation	F0: > 12 wpf but recommended not to exceed 16 wpf
10. Number of organisms per replicate	F0: 2 fish (male and female pair); F1: maximum 20 fish (eggs)/replicate (produced from F0 and F1 breeding pairs).
11. Number of treatments	5 test chemical treatments plus appropriate control(s)
12. Number of replicates per treatment	Minimum 6 replicates per treatment for test chemical and minimum 12 replicates for control, and for solvent control, if used (the number of replicates are doubled within reproduction phase in F1)
13. Number of organisms per test	Minimum of 84 fish in F0 and 504 in F1. (If solvent control is used, then 108 fish in F0 and 648 fish in F1). The unit counted is the post-eleutheroembryo.
14. Feeding regime	Fish are fed brine shrimp, <i>Artemia</i> spp., (24-hour old nauplii) <i>ad libitum</i> , supplemented with a commercially available flake food if needed (An example feeding schedule to ensure adequate growth and development to support robust reproduction can be found in Appendix 5).
15. Aeration	None unless dissolved oxygen approaches <60 % of air saturation value
16. Dilution water	Clean surface, well or reconstituted water or dechlorinated tap water.

17. Exposure period
Primarily 19 weeks (from F0 to F2 hatching)
18. Biological endpoints(primary)
Hatchability (F1 and F2); survival (F1, from hatch to 4 wpf (end of larval/beginning of juvenile), from 4 to 9 (or 10) wpf (beginning of juvenile to subadult) and from 9 to 15 wpf (subadult to adult termination)); growth (F1, length and weight at 9 and 15 wpf); secondary sex characteristics (F1, anal fin papillae at 9 and 15 wpf); vitellogenin (F1, vtg mRNA or VTG protein at 15wpf); phenotypic sex (F1, via gonad histology at 15 wpf); reproduction (F0 and F1, fecundity and fertility for 21 days); time to spawn (F1); and histopathology (F1, gonad, liver and kidney at 15 wpf)
19. Test validity criteria
Dissolved oxygen of ≥ 60 % air saturation value; mean water temperature of 24-26°C throughout the test; successful reproduction of ≥ 65 % females in control(s); mean daily fecundity of ≥ 20 eggs in control(s); hatchability of ≥ 80 % (average) in the controls (in each of the F1 and F2); survival after hatching until 3 wpf of ≥ 80 % (average) and from 3 wpf through termination for the generation of ≥ 90 % (average) in the controls (F1), concentrations of the test chemical in solution should be satisfactorily maintained within ± 20 % of the mean measured values.

Appendix 4

GUIDANCE ON TYPICAL CONTROL VALUES

It should be noted that these control values are based on a limited number of validation studies, and may be subject to amendment in the light of further experience.

Growth

Weight and length measurements are taken for all fish sampled at 9 (or 10) and 15 weeks post fertilisation (wpf). Following this protocol will yield expected wet weights at 9 wpf of 85-145 mg for males and 95-150 mg for females. The expected weights at 15 wpf are 250-330 mg for males and 280-350 mg for females. While there may be substantial deviations from these ranges for individual fish, control mean weights substantially outside of these ranges, especially lower, would suggest problems with feeding, temperature control, water quality, disease or any combination of these factors.

Hatch

Hatching success in controls is typically around 90 %, however, values as low as 80 % are not uncommon. Hatch success less than 75 % may indicate insufficient agitation of the developing eggs or inadequate care in handling the eggs such as lack of timely removal of dead eggs leading to fungal infestation.

Survival

Survival rates until 3 wpf from hatch and after 3 wpf are usually 90 % or greater for controls but survival rates in early life stages as low as 80 % for controls are not alarming. Survival rates in controls of less than 80 % would be cause for concern and may indicate insufficient cleaning of the aquaria leading to loss of larval fish through disease or from suffocation due to low dissolved oxygen levels. Mortality may also occur as a result of injury during tank cleaning and by the loss of larval fish to the drain system of the tank.

Vitellogenin gene

While absolute levels of *vitellogenin (vtg)* gene, expressed as copies/ng of total mRNA, may vary greatly between laboratories due to the procedures or instrumentation used, the ratio of *vtg* should be around 200 times greater in control females versus control males. It is not uncommon for this ratio to be as high as from 1 000 to 2000, however, ratios less than 200 are suspect and may indicate problems with sample contamination or problems with the procedure and/or reagents used.

Secondary sex characteristics

For males, the normal range of Secondary Sex Characteristics, defined as the total number of segments in the fin-rays of the anal fin papillae, is 40-80 segments at 9-10 wpf. By 15 wpf, the range for control males should be about 80-120 and 0 for control females. For unexplained reasons, in rare instances some males have no papillae present by 9 wpf but since all control males develop papillae by 15 wpf, this is most likely caused by delayed development. The presence of papillae in control females indicates the presence of XX males in the population.

XX-males

The normal background incidence of XX males in culture appears to be about 4 % or less at 25 °C with the incidence increasing with increased temperature. Steps should be taken to minimise the proportion of XX males in the population. Since the incidence of XX males appears to have a genetic component and is therefore heritable, monitoring the culture stock and ensuring that XX males are not used to propagate the culture stock is an effective means to reduce the incidence of XX males in the population.

Spawning activity

Spawning activity in the control replicates should be monitored daily prior to conducting the fecundity assessment. The control pairs can be qualitatively assessed visually for evidence of spawning activity. By 12-14 wpf most control pairs should be spawning. Low numbers of spawning pairs by this time indicates potential problems with the health, maturity or well-being of the fish.

Fecundity

Healthy, well fed 12-14 wpf medaka generally spawn daily, producing in the range of 15 to 50 eggs per day. Egg production for 16 of the recommended 24 control breeding pairs (> 65 %) should produce greater than 20 eggs per pair per day and may reach as high as about 40 eggs per day. Less than this amount may indicate immature, malnourished or unhealthy spawning pairs.

Fertility

The percentage of fertile eggs for control spawning pairs is typically in the 90 % range with values in the mid-to-upper 90s not uncommon. Fertility rates of less than 80 % for control eggs are suspect and may indicate either unhealthy individuals or less than ideal culture conditions.

Appendix 5

AN EXAMPLE OF A FEEDING SCHEDULE

An example of a feeding schedule to ensure adequate growth and development to support robust reproduction is shown in Table 1. Deviations from this feeding schedule may be acceptable, but it is recommended that they are tested to verify that acceptable growth and reproduction be observed. In order to follow the suggested feeding schedule, the dry weight of brine shrimp per volume of brine shrimp slurry needs to be determined prior to starting the test. This can be done by weighing a defined volume of brine shrimp slurry that has been dried for 24 hours at 60 °C on pre-weighed pans. To account for the weight of the salts in the slurry, an identical volume of the same salt solution used in the slurry should also be dried, weighed, and subtracted from the dried brine shrimp slurry weight. Alternatively, the brine shrimp can be filtered and rinsed with distilled water before drying, thereby eliminating the need to measure the weight of a “salt blank”. This information is used to convert the information in the Table from dry weight of brine shrimp to volume of brine shrimp slurry to be fed per fish. In addition, it is recommended that aliquots of the brine shrimp slurry are weighed weekly to verify the correct dry weight of brine shrimp being fed.

Table 1

Example of a feeding schedule

Time (post-hatch)	Brine Shrimp (mg dry weight/fish/day)
Day 1	0,5
Day 2	0,5
Day 3	0,6
Day 4	0,7
Day 5	0,8
Day 6	1,0
Day 7	1,3
Day 8	1,7
Day 9	2,2
Day 10	2,8
Day 11	3,5
Day 12	4,2
Day 13	4,5

Time (post-hatch)	Brine Shrimp (mg dry weight/fish/day)
Day 14	4,8
Day 15	5,2
Day 16-21	5,6
Week 4	7,7
Week 5	9,0
Week 6	11,0
Week 7	13,5
Week 8-sacrifice	22,5

Appendix 6

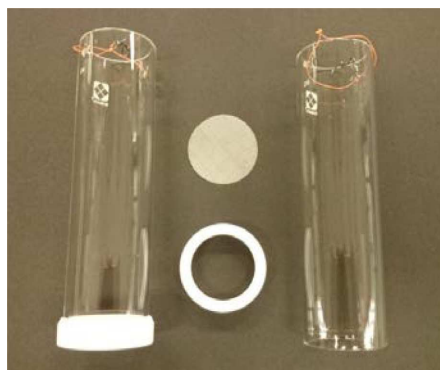
EXAMPLES OF AN EGG INCUBATION CHAMBER

Example A



This incubator consists of a transected glass centrifuge tube, connected by a stainless steel sleeve and held in place by the centrifuge screw top cap. A small glass or stainless steel tube projects through the cap and is positioned near the rounded bottom, gently bubbling air to suspend the eggs and reducing between-egg transmission of saprophytic fungal infections while also facilitating chemical exchange between the incubator and the holding tank.

Example B





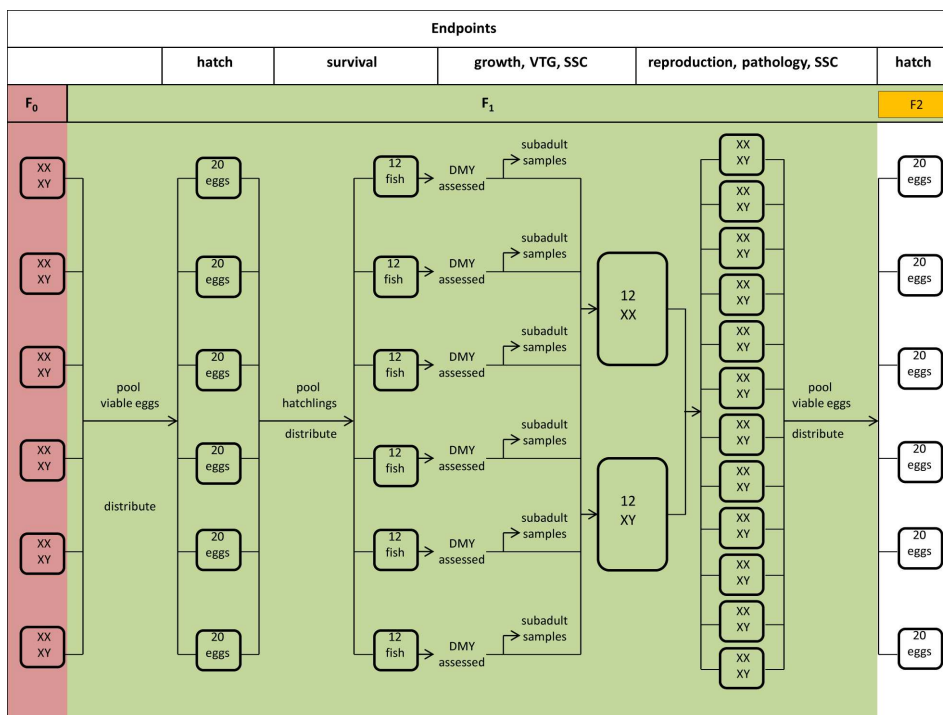
This incubator consists of a glass cylinder body (5 cm diameter and 10 cm height) and stainless wire mesh (0.25 ϕ and 32 mesh) which is attached to the bottom of the body with a PTFE ring. The incubators are suspended from the lifting bar to tanks, and shaken vertically (approximately 5 cm amplitude) in an appropriate cycle (approximately once every 4 seconds) for medaka eggs.

Appendix 7

SCHEMATIC DIAGRAM FOR POOLING AND POPULATING REPLICATES THROUGHOUT THE MEOGRT TEST METHOD

Figure 1

Pooling and repopulating replicates throughout the MEOGRT. The figure represents one treatment or 1/2 of a control. Due to pooling, replicate identity is not continuous throughout the test. Note that the term 'eggs' refers to viable, fertilised eggs (equivalent to embryos).



Treatments and Replication.

The test method recommends five test chemical treatments using technical grade material and a negative control. The number of replicates per treatment does not remain constant throughout the MEOGRT, and the number of replicates in the control treatment is double of any single test chemical treatment. In F₀, each test chemical treatment has six replicates while the negative control treatment has 12 replicates. Solvents are highly discouraged, and if used, a justification for both the use of a solvent and the choice of solvent should be included in the MEOGRT report. Also, if a solvent is used, two types of controls are necessary: a) a solvent control, and b) a negative control. These two control groups should each consist of a full complement of replicates at all points within the MEOGRT timeline. Throughout test organism development in the F₁ generation (and F₂, until hatch), this replicate structure remains the same. However, in the adult stage when F₁ breeding pairs are setup, the number of reproducing pair replicates per treatment is optimally doubled; therefore there are up to 12 replicate pairs in each test chemical treatment and 24 replicate pairs in the control group (and another 24 replicate pairs in the solvent control, if needed). The determination of hatch from embryos spawned by the F₁ pairs is done on the same replicate structure as was done for the embryos spawned by the F₀ pairs, meaning initially six replicates per test chemical treatment and 12 replicates in the control group(s).

Appendix 8

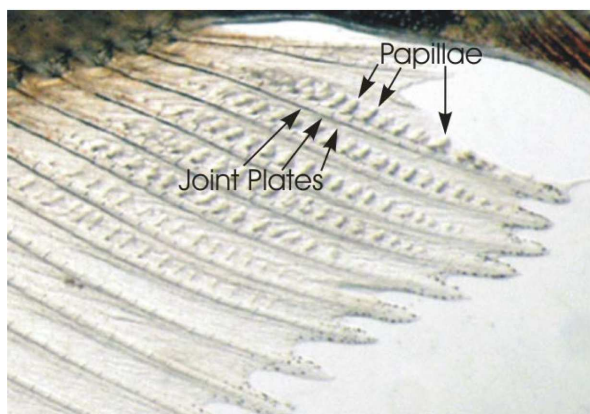
COUNTING ANAL FIN PAPILLAE

Major Materials and Reagents

- Dissecting microscope (with optional camera attached)
- Fixative (e.g., Davidson's (Bouin's is not recommended)), if not counting from image

Procedures

After necropsy, the anal fin should be imaged to allow for convenient counting of anal fin papillae. While imaging is the recommended method, the anal fin can be fixed with Davidson's fixative or other appropriate fixative for approximately 1 minute. It is important to keep the anal fin flat during fixation to allow for easier counting of papillae. The carcass with the anal fin can be stored in Davidson's fixative or other appropriate fixative until analysed. Count the number of joint plates (see **Figure 1**) with papillae which protrude from the posterior margin of the joint plate.

*Figure 1***Anal fin papillae**

Appendix 9

DETAILED TIMELINE OF MEOGRT

Test Weeks 1-3 (F0)

The F0 generation spawning fish that have met the selection criteria (see para. 16-20) are exposed for three weeks to allow the developing gametes and gonadal tissues to be exposed to the test chemical. Each replicate tank houses a single breeding fish pair (XX female-XY male breeding pair). Spawning eggs are collected, counted and assessed for fertility for 21 consecutive days, starting at Test Day 1.

Test Week 4 (F0 and F1)

It is preferable that the fertilised and viable eggs (embryos) are collected on a single day; however, if there are not enough embryos, the embryos may be collected over two days. If collected over two days, all embryos within the treatments that were collected on the first day are pooled with those collected on the second day. Then the total pooled embryos for each treatment are randomly distributed to each of the replicate incubators at 20 embryos per incubator. The mortalities of fertilised eggs (embryos) are checked and recorded daily. Dead eggs are removed from the incubators (death in fertilised eggs may be denoted by, particularly in the early stages, a marked loss of translucency and change in colouration, caused by coagulation and/or precipitation of protein, leading to a white opaque appearance; OECD 2010).

Note: If a single treatment requires a second day of collection, all treatments (including controls) need to follow this procedure. If after the second day of collection there are inadequate numbers of embryos within a treatment to load 20 embryos per incubator, then reduce the number of embryos loaded within that specific treatment to 15 embryos per incubator. If there are not enough embryos to load 15 per incubator, then reduce the number of replicate incubators until there are enough embryos for 15 per incubator. Additionally, more breeding pairs per treatment and controls could be added in F0 to produce more eggs to reach the recommended 20 per replicate.

On Test Day 24, the F0 breeding pairs are humanely killed and weight and length are recorded. If necessary F0 breeding pairs may be kept for an additional 1-2 days in order to restart F1.

Test Weeks 5-6 (F1)

One to two days before the anticipated start of hatching, stop or reduce the agitation of the incubating eggs to expedite hatching. As embryos hatch on each day, hatchlings are pooled by treatment and systematically distributed to each replicate larval tank within a specific treatment with no more than 12 hatchlings. This is done by randomly selecting hatchlings and placing a single hatchling in successive replicates in an indiscriminate draw, moving in order through the specific treatment replicates until all replicates within the treatment have 12 hatchlings. If there are not enough hatchlings to fill all replicates then ensure as many replicates as possible have 12 hatchlings to start the F1 phase.

The eggs that have not hatched by twice the median control day of hatch are considered non-viable and discarded. The number of hatchlings is recorded and hatching success (hatchability) is calculated in each replicate.

Test Weeks 7-11 (F1)

The survival of larval fish is checked and recorded daily in all replicates. On Test Day 43, the number of surviving fish in each replicate is recorded as well as the initial number of hatchlings placed in the replicate (nominally twelve). This allows for the calculation of the percent survival from hatch to the subadult stage.

Test Week 12 (F1)

On Test Day 78-85, a small sample is taken from the caudal fin of each fish to determine the genotypic sex of the individual (i.e., fin clipping) for all fish. This information is used to establish breeding pairs.

Within three days after the genotypic sex of each fish is determined, 12 breeding pairs per treatment and 24 pairs per control are randomly established. Two XX and XY fish from each replicate are randomly selected and then pooled by sex, and then randomly selected to establish a breeding pair (i.e., XX-XY pair). A minimum 12 replicates per chemical treatment and minimum 24 replicates for the control are established with one breeding pair per replicate. If a replicate does not have either two XX or two XY fish available for pooling, then fish with the appropriate gender genotype should be obtained from other replicates within the treatment.

The remaining fish (maximum 8 fish per replicate) are humanely killed and sampled for the various subadult endpoints. The *dmy* data (XX or XY) for all the subadult samples are retained to ensure that all endpoint data can be related to the genetic sex of each individual fish.

Test Weeks 13-14 (F1)

The exposure continues as the subadult breeding pairs develop into adults. On Test Day 98 (i.e. the day before egg collection is started), eggs are removed from both the aquaria and the females.

Test Weeks 15-17 (F1)

Spawned eggs are collected daily for 21 consecutive days in each replicate, and assessed for fecundity and fertility.

Test Week 18 (repeat of Test Week 4) (F1 and F2)

On Test Day 120, eggs collection is done in each replicate tank in the morning. The collected eggs are assessed and fertilised eggs (filaments removed) from each of the breeding pairs are pooled by treatment, and systematically distributed to egg incubation chambers with 20 fertilised eggs per incubator. The incubators may be placed in separate "incubator tanks" set up for each treatment or in the replicate tank that upon hatch will contain the hatched larvae. It is preferable that the embryos are collected on a single day; however, if there are not enough embryos, the embryos may be collected over two days. If collected over two days, all embryos within the treatments that were collected on the first day are pooled with those collected on the second day. Then the total pooled embryos for each treatment are randomly distributed to each of the replicate incubators at 20 embryos per incubator. Note: If a single treatment requires a second day of collection, all treatments (including controls) need to follow this procedure. If after the second day of collection there is inadequate numbers of embryos within a treatment to load 20 embryos per incubator, reduce the number of embryos loaded within that specific treatment to 15 embryos per incubator. If there are not enough embryos to load 15 per incubator, reduce the number of replicate incubators until there are enough embryos for 15 per incubator.

On Test Day 121 (or Test Day 122, to ensure the F2 has started well), the F1 breeding pairs are humanely killed and analysed for the adult endpoints. If necessary F1 breeding pairs maybe kept for an additional 1-2 days in order to restart F2.

Test Weeks 19-20 (F2)

One to two days before the anticipated start of hatching, stop or reduce the agitation of the incubating eggs to expedite hatching. If the test is terminated by the completion of the F2 hatching, each day the hatchlings are tallied and discarded. (Embryos that have not hatched after a prolonged incubation time, defined as twice the median control day of hatch, are considered non-viable).

Appendix 10

STATISTICAL ANALYSIS

The types of biological data generated in the MEOGRT are not unique to it and except for pathology data, many appropriate statistical methodologies have been developed to properly analyse similar data depending on the characteristics of the data including normality, variance homogeneity, whether the study design lends itself to hypothesis testing or regression analysis, parametric versus non-parametric tests, etc. In general principle, the suggested statistical analyses follow the recommendations of the OECD for ecotoxicity data (OECD 2006) and a decision flowchart for MEOGRT data analysis can be seen in Figure 2.

It is assumed that most often the datasets will display monotonic responses. Additionally, the issue of using a one-tailed statistical test versus a two-tailed statistical test should be considered. Unless there is a biological reasoning that would make a one-tailed test inappropriate, it is suggested that one-tailed tests be used. While the following section recommends certain statistical tests, if more appropriate and/or powerful statistical methods are developed for application to the specific data generated in the MEOGRT, those statistical tests would be used to leverage those advantages.

The MEOGRT data should be analysed separately for each genotypic sex. There are two strategies to analysing the data from sex reversed fish (either XX males or XY females). 1) Censor all data from sex reversed fish across the entire test except the prevalence of sex reversal in each replicate. 2) Leave the data from all sex reversed fish in the dataset and analyse based upon genotype.

Histopathology data

Histopathology data are reported as severity scores which are evaluated using a newly developed statistical procedure, the Rao-Scott Cochran-Armitage by Slices (RSCABS), (Green *et al.*, 2014). The Rao-Scott adjustment retains test-replication information; the *by Slices* procedure incorporates the biological expectation that severity scores tend to increase with increasing treatment concentrations. For each diagnosis, the RSCABS output specifies which treatments have higher prevalence of pathology than controls and the associated severity level.

Fecundity data

Analyses for fecundity data consist of a step-down Jonckheere-Terpstra or Williams' test to determine treatment effects, provided the data are consistent with a monotone concentration-response. With a step-down test, all comparisons are done at the 0,05 significance level and no adjustment for the number of comparisons made. The data are expected to be consistent with a monotone concentration response, but this can be verified either by visual inspection of the data or by constructing linear and quadratic contrasts of treatment means after a rank-order transform of the data. Unless the quadratic contrast is significant and the linear contrast is not significant, the trend test is done. Otherwise, Dunnett's test is used to determine treatment effects if the data are normally distributed with homogeneous variances. If those requirements are not met, then Dunn's test with a Bonferonni-Holm adjustment is used. All indicated tests are done independently of any overall F- or Kruskal-Wallis test. Further details are provided in OECD 2006.

Alternative methods can be used, such as a generalised linear model with Poisson errors for egg counts (with no transform), if justified statistically (Cameron and Trividi, 2013). Statistical advice is recommended if an alternative approach is used.

Daily Egg Count within a Single Generation

The ANOVA model is given by $Y = \text{Time} * \text{Time} + \text{Treatment} + * \text{Treatment} + \text{Time} * \text{Treatment} + * \text{Time} * \text{Treatment}$, with random effects of Replicate(Generation**Treatment*), and Time*Replicate(*Treatment*), allowing for unequal variance components of both types across generations. Here Time refers to the frequency of egg counts (e.g., Day or Week). This is a repeated measures analysis, with the correlations between observations on the same replicates accounting for the repeated measures nature of the data.

Main effects of treatment are tested using the Dunnett (or Dunnett-Hsu) test, which adjusts for the number of comparisons. Adjustments for the main effect of generation or time are needed, for with these two factors, there is no "control" level and every pair of levels is a comparison of possible interest. For these two main effects, if the F-test for the main effect is significant at the 0,05 level, then the pairwise comparisons across levels of that factor can then be tested at the 0,05 level without further adjustment.

The model includes two- and three-factor interactions, so that a main effect for, say, time, may not be significant even though time has a significant impact on results. Thus, if a two- or three-factor interaction involving time is significant at the 0,05 level, then one can accept the comparisons of levels of time at the 0,05 significance level without further adjustment.

Next are F-tests for significance of treatment within time, the so-called slices in the ANOVA table. If, for example, the slice for treatment within F1 and time 12, is significant at the 0,05 level, then the pairwise comparisons for treatment within F1 and time 12 can be accepted at the 0,05 level without further adjustment. Similar statements apply to tests for time within F1 and treatment and for generation within a time and treatment.

Finally, for comparisons not falling under any of the above categories, comparisons should be adjusted using the Bonferroni-Holm adjustment to p-values. Further information on analyses of such models can be found in Hocking (1985) and Hochberg and Tamhane (1987).

Alternatively, the raw data are recorded and presented in the study report as the fecundity (number of eggs) per replicate for each day. The replicate mean of the raw data should be calculated then a square root transformation applied. A one-way ANOVA on the transformed replicate means should be calculated followed by Dunnett contrasts. It may also be helpful to visually inspect the fecundity data of each treatment and/or replicate with a scatterplot that displays the data through time. This will allow an informal assessment of potential effects through time.

All other biological data

The statistical analyses are based on the underlying assumption that with proper dose selection the data will be monotonic. Thus, data are assumed to be monotonic and they are formally evaluated for monotonicity by using linear and quadratic contrasts. If the data are monotonic, a Jonckheere-Terpstra on replicate medians trend test (as advised in OECD 2006) is recommended. If the quadratic contrast is significant and the linear contrast is not, the data are considered non-monotonic.

If the data are non-monotonic, in particular because of the reduced response of the highest one or two treatments, consideration should be given to censoring the dataset so that the analysis is done without those treatments. This decision will need to be made with professional judgment and all available data, especially data that indicates overt toxicity at those treatment levels.

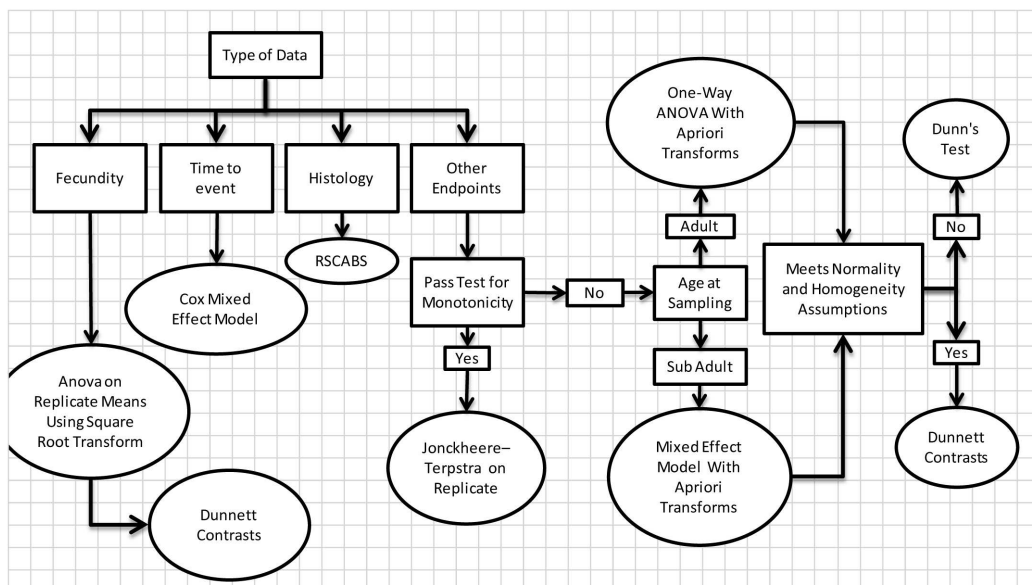
For weight and length, no transforms are recommended although they may occasionally be necessary. However, a log transformation is recommended for the vitellogenin data; a square root transformation is recommended for the SSC data (anal fin papillae); an arcsine-square root transformation is recommended for the data on proportion hatching, percent survival, sex ratio, and percent fertile eggs. Time to hatch and time to first spawn should be treated as time to event data, with individual embryos not hatching in the defined period or replicates never spawning treated as right-censored data. Time to hatch should be calculated from the median day of hatch of each replicate. These endpoints should be analysed using a mixed-effects Cox proportional hazard model.

The biological data from adult samples has one measurement per replicate, that is, there are one XX fish and one XY fish per replicate aquarium. Therefore, it is recommended that a one-way ANOVA be done on the replicate means. If the assumptions of the ANOVA (normality and variance homogeneity as assessed on the residuals of the ANOVA by Shapiro-Wilks test and Levene's test, respectively) are met, Dunnett contrasts should be used to determine treatments that were different from the control. On the other hand, if the assumptions of the ANOVA are not met, then a Dunn's test should be done to determine which treatments were different from control. A similar procedure is recommended for data that are in the form of percentages (fertility, hatch, and survival).

The biological data from subadult samples has from 1 to 8 measurements per replicate, that is, there can be variable numbers of individuals that contribute to the replicate mean for each genotypic sex. Therefore, it is recommended that a mixed effects ANOVA model be used followed by Dunnett contrasts, if the normality and variance homogeneity assumptions were met (on the residuals of the mixed effects ANOVA). If they were not met, then a Dunn's test should be done to determine which treatments were different from control.

Figure 2

Flow chart for the recommended statistical procedures for MEOGRT data analysis



LITERATURE

- (1) OECD (2014). Current Approaches in the Statistical Analysis of Ecotoxicity Data: A guidance to application (annexes to this publication exist as a separate document), OECD Publishing, Paris.
- (2) Cameron AC and Trivedi PK (2013). Regression Analysis of Count Data, 2nd edition, Econometric Society Monograph No 53, Cambridge University Press.
- (3) Hocking RR (1985). The Analysis of Linear Models, Monterey, CA: Brooks/Cole.
- (4) Hochberg Y and Tamhane AC (1987). Multiple Comparison Procedures. John Wiley and Sons, New York.

C.53 THE LARVAL AMPHIBIAN GROWTH AND DEVELOPMENT ASSAY (LAGDA)

INTRODUCTION

1. This test method is equivalent to OECD test guideline 241 (2015). The need to develop and validate an assay capable of identifying and characterising the adverse consequences of exposure to toxic chemicals in amphibians, originates from concerns that environmental levels of chemicals may cause adverse effects in both humans and wildlife. The OECD test guideline of the Larval Amphibian Growth and Development Assay (LAGDA) describes a toxicity test with an amphibian species that considers growth and development from fertilisation through the early juvenile period. It is an assay (typically 16 weeks) that assesses early development, metamorphosis, survival, growth, and partial reproductive maturation. It also enables measurement of a suite of other endpoints that allows for diagnostic evaluation of suspected endocrine disrupting chemicals (EDCs) or other types of developmental and reproductive toxicants. The method described in this test method is derived from validation work on African clawed frog (*Xenopus laevis*) by the U.S. Environmental Protection Agency (U.S. EPA) with supporting work in Japan (1). Although other amphibian species may be adapted to a growth and developmental test protocol with ability to determine genetic sex being an important component, the specific methods and observational endpoints detailed in this test method are applicable to *Xenopus laevis* alone.
2. The LAGDA serves as a higher tier test with an amphibian for collecting more comprehensive concentration-response information on adverse effects suitable for use in hazard identification and characterisation, and in ecological risk assessment. The assay fits at Level 4 of the OECD Conceptual Framework for Testing and Assessment of Endocrine Disruptors, where *in vivo* assays also provide data on adverse effects on endocrine relevant endpoints (2). The general experimental design entails exposing *X. laevis* embryos at Nieuwkoop and Faber (NF) stage 8-10 (3) to a minimum of four different concentrations of test chemical (generally spaced at not less than half-logarithmic intervals) and control(s) until 10 weeks after the median time to NF stage 62 in the control, with one interim sub-sample at NF stage 62 (≤ 45 post fertilisation; usually around 45 days (dpf)). There are four replicates in each test concentration with eight replicates for the control. Endpoints evaluated during the course of the exposure (at the interim sub-sample and final sample at completion of the test) include those indicative of generalised toxicity: mortality, abnormal behaviour, and growth determinations (length and weight), as well as endpoints designed to characterise specific endocrine toxicity modes of action targeting oestrogen, androgen or thyroid-mediated physiological processes. The method gives primary emphasis to potential population relevant effects (namely, adverse impacts on survival, development, growth and reproductive development) for the calculation of a No Observed Effect Concentration (NOEC) or an Effect Concentration causing x % change (ECx) in the endpoint measured. Although it should be noted that ECx approaches are rarely suitable for large studies of this type where increasing the number of test concentrations to allow for determination of the desired ECx may be impractical. It should also be noted that the method does not cover the reproductive phase itself. Definitions used in this test method are given in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

3. Due to the limited number of chemicals tested and laboratories involved in the validation of this rather complex assay, especially inter-laboratory reproducibility is not documented with experimental data so far, it is anticipated that when a sufficient number of studies is available to ascertain the impact of this new study design, OECD test guideline 241 will be reviewed and if necessary revised in light of experience gained. The LAGDA is an important assay to address potential contributors to amphibian population declines by evaluating the effects from exposure to chemicals during the sensitive larval stage, where effects on survival and development, including normal development of reproductive organs, may adversely affect populations.
4. The test is designed to detect an apical effect(s) resulting from both endocrine and non-endocrine mechanisms, and includes diagnostic endpoints which are partly specific to key endocrine modalities. It should be noted that until the LAGDA was developed, no validated assay existed that served this function for amphibians.
5. Before beginning the assay, it is important to have information about the physicochemical properties of the test chemical, particularly to allow the production of stable chemical solutions. It is also necessary to have an adequately sensitive analytical method for verifying test chemical concentrations. Over a duration of approximate 16 weeks, the assay requires a total number of 480 animals, i.e., *X. laevis* embryos, (or 640 embryos, if a solvent control is used) to ensure sufficient power of the test for the evaluation of population-relevant endpoints such as growth, development and reproductive maturation.
6. Before use of the test method for regulatory testing of a mixture, it should be considered whether it will provide acceptable results for the intended regulatory purpose. Furthermore, this assay does not evaluate fecundity directly, so it may not be applicable for use at a more advanced stage than Level 4 of the OECD Conceptual Framework for Testing and Assessment of Endocrine Disruptors.

SCIENTIFIC BASIS FOR THE TEST METHOD

7. Much of our current understanding of amphibian biology has been obtained using the laboratory model species *X. laevis*. This species can be routinely cultured in the laboratory, ovulation can be induced using human chorionic gonadotropin (hCG) and animal stocks are readily available from commercial breeders.
8. Like all vertebrates, reproduction in amphibians is under the control of the hypothalamic pituitary gonadal (HPG) axis (4). Oestrogens and androgens are mediators of this endocrine system, directing the development and physiology of sexually-dimorphic tissues. There are three distinct phases in the life cycle of amphibians when this axis is especially active: (1) gonadal differentiation during larval development, (2) development of secondary sex characteristics and gonadal maturation during the juvenile phase and (3) functional reproduction of adults. Each of these three developmental windows are likely susceptible to endocrine perturbation by certain chemicals such as estrogens and androgens, ultimately leading to a loss of reproductive fitness by the organisms.
9. The gonads begin development at NF stage 43, when the bipotential genital ridge first develops. Differentiation of the gonads begins at NF stage 52 when primordial germ cells either migrate to medullary tissue (males) or remain in the cortical region (females) of the developing gonads (3). This process of sexual differentiation of the gonads was first reported to be susceptible to chemical alteration in *Xenopus* in the 1950's (5) (6). Exposure of tadpoles to estradiol during this period of gonad differentiation results in sex reversal of males that when raised to adulthood are fully functional females (7) (8). Functional sex reversal of females into males is also possible and has been reported following implantation of testis tissue in tadpoles (9). However, although exposure to an aromatase inhibitor also causes functional sex reversal in *X. tropicalis* (10), this has not been shown to occur in *X. laevis*. Historically, toxicant effects on gonadal differentiation have been assessed by histological examination of the gonads at metamorphosis

and sex reversal could only be determined by analysis of sex ratios. Until recently, there had been no means to directly determine the genetic sex of *Xenopus*. However, recent establishment of sex linked markers in *X. laevis* make it possible to determine genetic sex and allows for the direct identification of sex reversed animals (11).

10. In males, juvenile development proceeds as blood levels of testosterone increase corresponding with the development of secondary sex characteristics as well as testis development. In females, estradiol is produced by the ovaries resulting in the appearance of vitellogenin (VTG) in the plasma, vitellogenic oocytes in the ovary and the development of oviducts (12). Oviducts are female secondary sex characteristics that function in oocyte maturation during reproduction. Jelly coats are applied to the outside of oocytes as they pass through the oviduct and collect in the ovisac, ready for fertilisation. Oviduct development appears to be regulated by oestrogens as development correlates with blood estradiol levels in *X. laevis* (13) and *X. tropicalis* (12). The development of oviducts in males following exposure to polychlorinated biphenyl compounds (14) and 4-*tert*-octylphenol (15) has been reported.

PRINCIPLE OF THE TEST

11. The test design entails exposing *X. laevis* embryos at NF stage 8-10 via the water route to four different concentrations of test chemical as well as control(s) until 10 weeks after the median time to NF stage 62 in the control with one interim sub-sample at NF stage 62. While it may also be possible to dose highly hydrophobic chemicals via the feed, there has been little experience using this exposure route in this assay to date. There are four replicates in each test concentration with eight replicates for each control used. Endpoints evaluated during the course of the exposure include those indicative of generalised toxicity (i.e., mortality, abnormal behaviour and growth determinations (length and weight)), as well as endpoints designed to characterise specific endocrine toxicity modes of action targeting oestrogen-, androgen-, or thyroid-mediated physiological processes (i.e. thyroid histopathology, gonad and gonad duct histopathology, abnormal development, plasma vitellogenin (optional), and genotypic/phenotypic sex ratios).

TEST VALIDITY CRITERIA

12. The following criteria for test validity apply:

- The dissolved oxygen concentration should be ≥ 40 % of air saturation value throughout the test;
- The water temperature should be in the range of 21 ± 1 °C and the inter-replicate and the inter-treatment differentials should not exceed 1,0 °C;
- pH of the test solution should be maintained between 6.5 and 8.5, and the inter-replicate and the inter-treatment differentials should not exceed 0,5;
- Evidence should be available to demonstrate that the concentrations of the test chemical in solution have been satisfactorily maintained within ± 20 % of the mean measured values;
- Mortality over the exposure period should be ≤ 20 % in each replicate in the controls;

- ≥ 70 % viability in the spawn chosen to start the study;
 - The median time to NF stage 62 of the controls should be ≤ 45 days.
 - The mean weight of test organisms at NF stage 62 and at the termination of the assay in controls and solvent controls (if used) should reach $1,0 \pm 0,2$ and $11,5 \pm 3$ g, respectively.
13. While not a validity criterion, it is recommended that at least three treatment levels with three uncompromised replicates be available for analysis. Excessive mortality, which compromises a treatment, is defined as > 4 mortalities (> 20 %) in 2 or more replicates that cannot be explained by technical error. At least three treatment levels without obvious overt toxicity should be available for analysis. Signs of overt toxicity may include, but are not limited to, floating on the surface, lying on the bottom of the tank, inverted or irregular swimming, lack of surfacing activity, and being nonresponsive to stimuli, morphological abnormalities (e.g., limb deformities), hemorrhagic lesions, and abdominal oedema.
14. In case a deviation from the test validity criteria is observed, the consequences should be considered in relation to the reliability of the test results, and these deviations and considerations should be included in the test report.

DESCRIPTION OF THE METHODS

Apparatus

15. Normal laboratory equipment and especially the following:
- (a) temperature controlling apparatus (e.g., heaters or coolers adjustable to 21 ± 1 °C);
 - (b) thermometer;
 - (c) binocular dissection microscope and dissection tools;
 - (d) digital camera with at least 4 megapixel resolution and micro function (if needed);
 - (e) analytical balance capable of measuring to 0,001 mg or 1 µg;
 - (f) dissolved oxygen meter and pH meter;
 - (g) light intensity meter capable of measuring in lux units.

Water

Source and quality

16. Any dilution water that is locally available (e.g. spring water or charcoal-filtered tap water) and permits normal growth and development of *X. laevis* can be used, and evidence of normal growth in this water should be available. Because local water quality can differ substantially from one area to another, analysis of water quality should be undertaken, particularly if historical data on the utility of the water for raising amphibian larvae is not available. Measurements of heavy metals (e.g. Cu, Pb, Zn, Hg, Cd, Ni), major anions and cations (e.g. Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, SO₄²⁻), pesticides, total organic carbon and suspended solids should be made before testing begins and/or, for example, every six months where a dilution water is known to be relatively constant in quality. Some chemical characteristics of acceptable dilution water are listed in Appendix 2.

Iodide concentration in test water

17. In order for the thyroid gland to synthesise thyroid hormones to support normal metamorphosis, sufficient iodide should be available to the larvae through a combination of aqueous and dietary sources. Currently, there are no empirically derived guidelines for minimum iodide concentrations in either food or water to ensure proper development. However, iodide availability may affect the responsiveness of the thyroid system to thyroid active agents and is known to modulate the basal activity of the thyroid gland which deserves attention when interpreting the results from thyroid histopathology. Based on previous work, successful performance of the assay has been demonstrated when dilution water iodide (I⁻) concentrations range between 0,5 and 10 µg/l. Ideally, the minimum iodide concentration in the dilution water throughout the test should be 0,5 µg/l (added as the sodium or potassium salt). If the test water is reconstituted from deionised water, iodine should be added at a minimum concentration of 0,5 µg/l. The measured iodide concentrations from the test water (i.e., dilution water) and the supplementation of the test water with iodine or other salts (if used) should be reported. Iodine content may also be measured in food(s) in addition to test water.

Exposure system

18. The test was developed using a flow-through diluter system. The system components should have water-contact components of glass, stainless steel, and/or other chemically inert materials. Exposure tanks should be glass or stainless steel aquaria and tank usable volume should be between 4,0 and 10,0 l (minimum water depth of 10 to 15 cm). The system should be capable of supporting all exposure concentrations, a control, and a solvent control, if necessary, with four replicates per treatment and eight in the controls. The flow rate to each tank should be constant in consideration of both the maintenance of biological conditions and chemical exposure. It is recommended that flow rates should be appropriate (e.g., at least 5 tank turnovers per day) to avoid chemical concentration declines due to metabolism by both the test organisms and aquatic microorganisms present in the aquaria or abiotic routes of degradation (hydrolysis, photolysis) or dissipation (volatilisation, sorption). The treatment tanks should be randomly assigned to a position in the exposure system to reduce potential positional effects, including slight variations in temperature, light intensity, etc. Further information on setting up flow-through exposure systems can be obtained from the ASTM Standard Guide for Conducting Acute Toxicity Tests on Test Materials with Fishes, Macroinvertebrates, and Amphibians (16).

Chemical delivery: preparation of test solutions

19. To make test solutions in the exposure system, stock solution of the test chemical should be dosed into the exposure system by an appropriate pump or other apparatus. The flow rate of the stock solution should be calibrated in accordance with analytical confirmation of the test solutions before the initiation of exposure, and checked volumetrically periodically during the test. The test solution in each chamber should be renewed at a minimum of 5 volume renewals/day.

20. The method used to introduce the test chemical to the system can vary depending on its physicochemical properties. Therefore, prior to the test, baseline information about the chemical that is relevant to determining its testability should be obtained. Useful information about test chemical-specific properties include the structural formula, molecular weight, purity, stability in water and light, pK_a and K_{ow} , water solubility (preferably in the test medium) and vapour pressure as well as results of a test for ready biodegradability (test method C.4 (17) or C.29 (18)). Solubility and vapour pressure can be used to calculate Henry's law constant, which will indicate whether losses due to evaporation of the test chemical may occur. Conduct of this test without the information listed above should be carefully considered as the study design will be dependent on the physicochemical properties of the test chemical and, without these data test results may be difficult to interpret or meaningless. A reliable analytical method for the quantification of the test chemical in the test solutions with known and reported accuracy and limit of detection should be available. Water soluble test chemicals can be dissolved in aliquots of dilution water at a concentration which allows delivery at the target test concentration in a flow-through system. Chemicals which are liquid or solid at room temperature and moderately soluble in water may require liquid:liquid or liquid:solid (e.g., glass wool column) saturators (19). While it may also be possible to dose very hydrophobic test chemicals via the feed, there has been little experience using that exposure route in this assay.
21. Test solutions of the chosen concentrations are prepared by dilution of a stock solution. The stock solution should preferably be prepared by simply mixing or agitating the test chemical in dilution water by mechanical means (e.g. stirring and/or ultrasonication). Saturation columns/systems or passive dosing methods (20) can be used for achieving a suitably concentrated stock solution. The preference is to use a co-solvent-free test system; however, different test chemicals will possess varied physicochemical properties that will likely require different approaches for preparation of chemical exposure water. All efforts should be made to avoid solvents or carriers because: (1) certain solvents themselves may result in toxicity and/or undesirable or unexpected responses, (2) testing chemicals above their water solubility (as can frequently occur through the use of solvents) can result in inaccurate determinations of effective concentrations, (3) the use of solvents in longer-term tests can result in a significant degree of "biofilming" associated with microbial activity which may impact environmental conditions as well as the ability to maintain exposure concentrations and (4) the absence of historical data that demonstrate that the solvent does not influence the outcome of the study, use of solvents requires a solvent control treatment which has significant animal welfare implications as additional animals are required to conduct the test. For difficult to test chemicals, a solvent may be employed as a last resort, and the OECD Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures should be consulted (21) to determine the best method. The choice of solvent will be determined by the chemical properties of the test chemical and the availability of historical control data on the solvent. In the absence of historical data, the suitability of a solvent should be determined prior to conducting the definitive study. In the event that use of a solvent is unavoidable, and microbial activity (biofilming) occurs, recommend recording/reporting of the biofilming per tank (at least weekly) throughout the test. Ideally, the solvent concentration should be kept constant in the solvent control and all test treatments. If the concentration of solvent is not kept constant, the highest concentration of solvent in the test treatment should be used in the solvent control. In cases where a solvent carrier is used, maximum solvent concentrations should not exceed 100 µl/l or 100 mg/l (21), and it is recommended to keep solvent concentration as low as possible (e.g. ≤20 µl/l) to avoid potential effects of the solvent on endpoints measured (22).

Test animals

Test species

22. The test species is *X. laevis* because this is: (1) routinely cultured in laboratories worldwide, (2) easily obtainable through commercial suppliers and (3) capable of having its genetic sex determined.

Adult care and breeding

23. Appropriate care and breeding of *X. laevis* is described by a standardised guideline (23). Housing and care of *X. laevis* are also described by Read (24). To induce breeding, three to five pairs of adult females and males are injected intraperitoneally with human chorionic gonadotropin (hCG). Female and male specimens are injected with e.g., approximately 800-1 000 IU and 500-800 IU, respectively, of hCG dissolved in 0,6-0,9 % saline solution (or frog Ringer's solution, an isotonic saline for use with amphibians). Injection volumes should be about 10 µl/g body weight (~1 000 µl). Afterwards, induced breeding pairs are held in large tanks, undisturbed and under static conditions

promote amplexus. The bottom of each breeding tank should have a false bottom of stainless steel mesh (e.g., 1,25 cm openings) which permits the eggs to fall to the bottom of the tank. Frogs injected with hCG in the late afternoon will usually deposit most of their eggs by mid-morning of the next day. After a sufficient quantity of eggs is released and fertilised, adults should be removed from the breeding tanks. Eggs are then collected and jelly coats are removed by L-cysteine treatment (23). A 2 % L-cysteine solution should be prepared and pH adjusted to 8.1 with 1 M NaOH. This 21 °C solution is added to a 500 ml Erlenmeyer flask containing the eggs from a single spawn and swirled gently for one to two minutes and then rinsed thoroughly 6-8 times with 21 °C culture water. The eggs are then transferred to a crystallising dish and determined to be > 70 % viable with minimal abnormalities in embryos exhibiting cell division.

TEST DESIGN

Test concentrations

24. It is recommended to use a minimum of four chemical concentrations and appropriate controls (including solvent controls, if necessary). Generally, a concentration separation (spacing factor) not exceeding 3.2 is recommended.

25. For the purposes of this test, results from existing amphibian studies should be used to the extent possible in determining the highest test concentration so as to avoid concentrations that are overtly toxic. Information from, for example, quantitative structure-activity relationships, read across and data from existing amphibian studies such as the Amphibian Metamorphosis Assay, test method C.38 (25) and the Frog Embryo Teratogenesis Assay - *Xenopus* (23) and/or fish tests such as test methods C.48, C.41 and C.49 (26) (27) (28) may contribute toward setting this concentration. Prior to running the LAGDA a range finding experiment may be conducted. It is recommended that the range-finding exposure is initiated within 24 hours of fertilisation and continued for 7-14 days (or more, if needed), and the test concentrations are set such that the intervals between test concentrations are no greater than a factor of 10. The results of the range finding experiment should serve to set the highest test concentration in the LAGDA. Note that if a solvent has to be used, then the suitability of the solvent (i.e. whether it may have an impact on the outcome of the study) could be determined as part of the range finding study.

Replicates within treatment groups and controls

26. A minimum of four replicate tanks per test concentration and a minimum of eight replicates for the controls (and solvent control, if needed) should be used (i.e., the number of replicates in the control and any solvent control should be twice as large as the number of replicates of each treatment group, to ensure appropriate statistical power). Each replicate should contain no more than 20 animals. The minimum number of animals processed would be 15 (5 for NF stage 62 sub-sample and 10 juveniles). However, additional animals are added to each replicate to factor in the possibility for mortality while maintaining the critical number of 15.

PROCEDURE

Assay overview

27. The assay is initiated with newly spawned embryos (NF stage 8-10) and continues into juvenile development. Animals are examined daily for mortality and any sign of abnormal behaviour. At NF stage 62, a larval sub-sample (up to 5 animals per replicate) is collected and various endpoints are examined (Table 1). After all animals have reached NF stage 66, i.e. completion of metamorphosis (or after 70 days from the assay initiation, whichever comes first), a cull is carried out at random (but without sub-sampling) to reduce the number of animals (10 per tank) (see paragraph 43), and the remaining animals continue exposure until 10 weeks after the median time to NF stage 62 in the control. At test termination (juvenile sampling) additional measurements are made (Table 1).

Exposure conditions

28. A complete summary of test parameters can be found in Appendix 3. During the exposure period, dissolved oxygen, temperature, and pH of test solutions should be measured daily. Conductivity, alkalinity, and hardness are measured once a month. For the water temperature of test solutions, the inter-replicate and inter-treatment differentials (within one day) should not exceed 1,0 °C. Also, for pH of test solutions, the inter-replicate and inter-treatment differentials should not exceed 0,5.
29. The exposure tanks may be siphoned on a daily basis to remove uneaten food and waste products, being careful to avoid cross-contamination of tanks. Care should be used to minimise stress and trauma to the animals, especially during movement, cleaning of aquaria, and manipulation. Stressful conditions/activities should be avoided such as loud and/or incessant noise, tapping on aquaria, vibrations in the tank.

Duration of exposure to the test chemical

30. The exposure is initiated with newly spawned embryos (NF stage 8-10) and continued until ten weeks after the median time to NF stage 62 (≤ 45 days from the assay initiation) in control group. Generally, the duration of the LAGDA is 16 weeks (maximum 17 weeks).

Initiation of assay

31. Parent animals used for the initiation of the assay should have previously been shown to produce offspring that can be genetically sexed (Appendix 5). After spawning of adults, embryos are collected, cysteine-treated to remove the jelly coat and screened for viability (23). Cysteine treatment allows the embryos to be handled during screening without sticking to surfaces. Screening takes place under a dissecting microscope using an appropriately sized eye dropper to remove non-viable embryos. It is preferred that a single spawn resulting in greater than 70 % viability be used for the test. Embryos at NF stage 8-10 are randomly distributed into exposure treatment tanks containing an appropriate volume of dilution water until each tank contains 20 embryos. Embryos should be carefully handled during this transfer in order to minimise handling stress and to avoid any injury. At 96 hours post fertilisation, the tadpoles should have moved up the water column and begun clinging to the sides of the tank.

Feeding regime

32. Feed and feeding rate change during different life stages of *X. laevis* are a very important aspect of the LAGDA protocol. Excessive feeding during the larval phase typically results in increased incidences and severity of scoliosis (Appendix 8) and should be avoided. Conversely, inadequate feeding during the larval phase results in highly variable developmental rates among controls potentially compromising statistical power or confounding test results. Appendix 4 provides recommended larval and juvenile diet and feeding regimes for *X. laevis* in flow-through conditions, but alternatives are permissible providing the test organisms grow and develop satisfactorily. It is important to note that if endocrine-specific endpoints are being measured, feed should be free of endocrine-active substances such as soy meal.

Larval feeding

33. The recommended larval diet consists of trout starter feeds, *Spirulina* algae discs and goldfish crisps (e.g., TetraFin® flakes, Tetra, Germany) blended together in culture (or dilution) water. This mixture is administered three times daily on weekdays and once daily on weekends. Tadpoles are also fed live brine shrimp, *Artemia* spp., 24-hour-old nauplii, twice daily on weekdays and once daily on the weekends starting on day 8 post-fertilisation. The larval feeding, which should be consistent in each test vessel, should allow appropriate growth and development for test animals in order to ensure reproducibility and transferability of the assay results: (1) the median time to NF stage 62 in controls should be ≤ 45 days and (2) a mean weight within $1,0 \pm 0,2$ g at NF stage 62 in controls is recommended.

Juvenile feeding

34. Once metamorphosis is complete, the feeding regime consists of premium sinking frog food, e.g., Sinking Frog Food -3/32 (Xenopus Express, FL, USA) (Appendix 4). For froglets (early juveniles), the pellets are briefly run in a coffee grinder, blender or crushed with a mortar and pestle in order to reduce their size. Once juveniles are large enough to consume full pellets, grinding or crushing is no longer necessary. The animals should be fed once per day. The juvenile feeding should allow appropriate growth and development of the organisms: a mean weight within $11,5 \pm 3$ g in control juveniles at the termination of the assay is recommended.

Analytical chemistry

35. Prior to initiation of the assay, the stability of the test chemical (e.g., solubility, degradability, and volatility) and all analytical methods needed should be established e.g., using existing information or knowledge. When dosing via the dilution water, it is recommended that test solutions from each replicate tank be analysed prior to test initiation to verify system performance. During the exposure period, the concentrations of the test chemical are determined at appropriate intervals, preferably every week for at least one replicate in each treatment group, rotating between replicates of the same treatment group every week. It is recommended that results be based on measured concentrations. However, if concentration of the test chemical in solution has been satisfactorily maintained within ± 20 % of the nominal concentration throughout the test, then the results can either be based on nominal or measured values. Also, the coefficient of variation (CV) of the measured test concentrations over the entire test period within a treatment should be maintained at 20 % or less in each concentration. When the measured concentrations do not remain within 80-120 % of the nominal concentration (for example, when testing highly biodegradable or adsorptive chemicals), the effect concentrations should be determined and expressed relative to the arithmetic mean concentration for flow-through tests.

36. The flow rates of dilution water and stock solution should be checked at appropriate intervals (e.g. three times a week) throughout the exposure duration. In the case of chemicals which cannot be detected at some or all of the nominal concentrations, (e.g., due to rapid degradation or adsorption in the test vessels, or by marked chemical accumulation in the bodies of exposed animals), it is recommended that the renewal rate of the test solution in each chamber be adapted to maintain test concentrations as constant as possible.

Observations and endpoint measurements

37. The endpoints evaluated during the course of the exposure are those indicative of toxicity including mortality, abnormal behaviour such as clinical signs of disease and/or general toxicities, and growth determinations (length and weight), as well as pathology endpoints which may respond to both general toxicity and endocrine modes of action targeting oestrogen-, androgen-, or thyroid-mediated pathways. In addition, plasma VTG concentration may be optionally measured at the termination of the assay. Measurement of VTG may be useful in understanding study results in the context of endocrine mechanisms for suspected EDCs. The endpoints and timing of measurements are summarised in Table 1.

Table 1

Endpoint overview of the LAGDA

Endpoints (*)	Daily	Interim Sampling (Larval sampling)	Test Termination (Juvenile sampling)
Mortality and abnormalities	X		
Time to NF stage 62		X	
Histo(patho)logy (thyroid gland)		X	
Morphometrics (growth in weight and length)		X	X
Liver-somatic index (LSI)			X
Genetic/phenotypic sex ratios			X
Histopathology (gonads, reproductive ducts, kidney and liver)			X
Vitellogenin (VTG) (optional)			X

(*) All endpoints are analysed statistically.

Mortality and daily observations

38. All test tanks should be checked daily for dead animals and mortalities recorded for each tank. Dead animals should be removed from the test tank as soon as observed. The developmental stage of dead animals should be categorised as either pre-NF stage 58 (pre-forelimb emergence), NF stage 58-NF stage 62, NF stage 63-NF stage 66 (between NF stage 62 and complete tail absorption), or post-NF stage 66 (post-larval). Mortality rates exceeding 20 % may indicate inappropriate test conditions or overtly toxic effects of the test chemical. The animals tend to be most sensitive to non-chemical induced mortality events during the first few days of development after the spawning event and during metamorphic climax. Such mortality could be apparent from the control data.
39. In addition, any observation of abnormal behaviour, grossly visible malformations (e.g., scoliosis), or lesions should be recorded. Observations of scoliosis should be counted (incidence) and graded with respect to severity (e.g., not remarkable – NR, minimal – 1, moderate – 2, severe – 3; Appendix 8). Efforts should be made to ensure that the prevalence of moderate and severe scoliosis is limited (e.g., below 10 % in controls) throughout the study, although greater prevalence of control abnormalities would not necessarily be a reason for stopping the test. Normal behaviour for larval animals is characterised by suspension in the water column with tail elevated above the head, regular rhythmic tail fin beating, periodic surfacing, operculating, and being responsive to stimuli. Abnormal behaviours would include, for example, floating on the surface, lying on the bottom of the tank, inverted or irregular swimming, lack of surfacing activity, and being nonresponsive to stimuli. For post-metamorphic animals, in addition to the above abnormal behaviours, gross differences in food consumption between treatments should be recorded. Gross malformations and lesions could include morphological abnormalities (e.g., limb deformities), haemorrhagic lesions, abdominal oedema, and bacterial or fungal infections, to name a few. The occurrences of lesions on the head of juveniles, just posterior to the nostrils, may be indications of insufficient humidity levels. These determinations are qualitative and should be considered akin to clinical signs of disease/stress and made in comparison to control animals. If the rate of occurrence is greater in exposed tanks than in the controls, then these should be considered as evidence for overt toxicity.

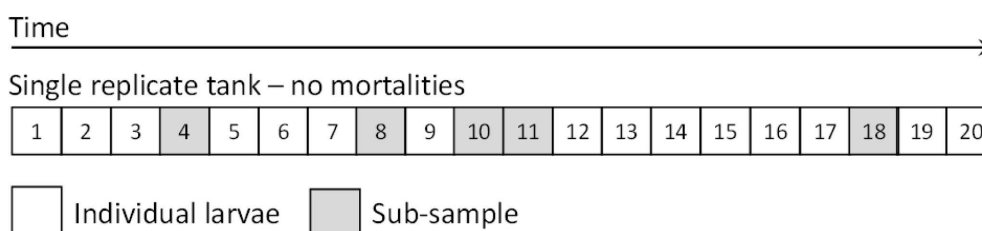
Larval sub-sampling

Outline of larval sub-sampling:

40. The tadpoles that have reached NF stage 62 should be removed from the tanks and either sampled or moved to the next part of the exposure in a new tank, or physically separated from the remaining tadpoles in the same tank with a divider. Tadpoles are checked daily, and the study day on which an individual tadpole reaches NF stage 62 is recorded. The defining characteristic for use in this assessment is the shape of the head. Once the head has become reduced in size such that it is visually approximately the same width as the trunk of the tadpole and forelimbs at the level of the middle of the heart, then that individual would be counted as having attained NF stage 62.
41. The goal is to sample a total of five NF stage 62 tadpoles per replicate tank. This should be performed entirely at random, but decided *a priori*. A hypothetical example of a replicate tank is provided in **Figure 1**. Should there be 20 surviving tadpoles in a particular tank when the first individual reaches NF stage 62, five random numbers should be chosen from 1-20. Tadpole #1 is the first individual to reach NF stage 62 and tadpole #20 is the last individual in a tank to reach NF stage 62. Likewise, if there are 18 surviving larvae in a tank, five random numbers should be chosen from 1-18. This should be performed for every replicate tank when the first individual on test reaches NF stage 62. If there are mortalities during the NF stage 62 sampling, the remaining samples need to be re-randomised based on how many larvae are left <NF stage 62 and how many more samples are needed to reach a total of five samples from that replicate. On the day a tadpole reaches NF stage 62, reference to the prepared sampling chart is made to determine whether that individual is sampled or physically separated from the remaining tadpoles for continued exposure. In the example provided (Figure 1), the first individual to reach NF stage 62 (i.e. box #1) is physically separated from the other larvae, continues exposure and the study day on which that individual reached NF stage 62 is recorded. Subsequently, individuals #2 and #3 are treated the same way as #1 and then individual #4 is sampled for growth and thyroid histology (according to this example). This procedure continues until the 20th individual either joins the rest of the post-NF stage 62 individuals or is sampled. The random procedure used must give each organism on test equal probability of being selected. This can be achieved by using any randomising method, but also requires that each tadpole be netted at some point throughout the NF stage 62 sub-sampling period.

Figure 1

Hypothetical example of NF stage 62 sampling regime for a single replicate tank



42. For the larval sub-sampling, the endpoints obtained are: (1) time to NF stage 62 (i.e., number of days between fertilisation and NF stage 62), (2) external abnormalities, (3) morphometrics (e.g., weight and length) and (4) thyroid histology.

Humane killing of tadpoles

43. The sub-sample of NF stage 62 tadpoles (5 individuals per replicate) should be euthanised by immersion for 30 minutes in appropriate amounts (e.g., 500 ml) of anaesthetic solution (e.g., 0,3 % solution of MS-222, tricaine methane sulfonate, CAS.886-86-2). MS-222 solution should be buffered with sodium bicarbonate to a pH of approximately 7,0 because unbuffered MS-222 solution is acidic and irritating to frog skin resulting in poor absorption and unnecessary additional stress to the organisms.
44. Using a mesh dip net, a tadpole is removed from the experimental chamber and transported (placed) into the euthanasia solution. The animal is properly euthanised and is ready for necropsy when it is unresponsive to external stimuli such as pinching the hind limb with a pair of forceps.

Morphometrics (weight and length)

45. Measurements of wet weight (nearest mg) and snout-to-vent length (SVL) (nearest 0,1 mm) for each tadpole should be made immediately after it becomes non-responsive by anaesthesia (Figure 2a). Image analysis software may be used to measure SVL from a photograph. Tadpoles should be blotted dry before weighing to remove excess adherent water. After measurements of body size (weight and SVL) are made, any gross morphological abnormalities and/or clinical signs of toxicity such as scoliosis (see Appendix 8), petechiae and haemorrhage should be recorded or noted, and digital documentation is recommended. Note that petechiae are small red or purple haemorrhages in skin capillaries.

Tissue Collection and Fixation

46. For the larval sub-sample, thyroid glands are assessed for histology. The lower torso posterior to the forelimbs is removed and discarded. The trimmed carcass is fixed in Davidson's fixative. The volume of fixative in the container should be at least 10 times the approximate volume of the tissues. Appropriate agitation or circulation of the fixative should be achieved to adequately fix the tissues of interest. All tissues remain in Davidson's fixative for at least 48 hours, but no longer than 96 hours, at which time they are rinsed in deionised water and stored in 10 % neutral buffered formalin (1) (29).

Thyroid histology

47. Each larval sub-sample (tissues fixed) is histologically assessed for thyroid glands, i.e., diagnosis and severity grading (29) (30).

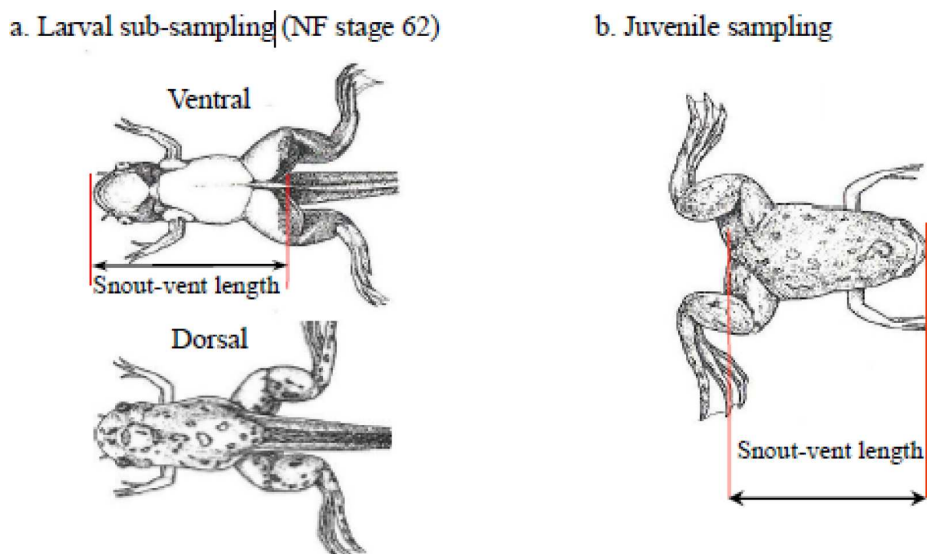


Figure 2: Landmarks for measuring snout-vent length for the LAGDA in NF Stage 62 (a) and juvenile frogs (b). The defining characteristics of NF stage 62 (a): the head is the same width as the trunk, the olfactory nerve length is shorter than the diameter of the olfactory bulb (dorsal view), and the forelimbs are at the level of the heart (ventral view). Images adapted from Nieuwkoop and Faber (1994).

End of larval exposure

48. Given the initial number of tadpoles, it is expected that there will likely be a small percentage of individuals that do not develop normally and do not complete metamorphosis (NF stage 66) in a reasonable amount of time. The larval portion of the exposure should not exceed 70 days. Any tadpoles remaining at the end of this period should be euthanised (see para. 43), their wet weight and SVL measured, staged according to Nieuwkoop and Faber, 1994, and any developmental abnormalities noted.

Cull after NF stage 66

49. Ten individuals per tank should continue from NF stage 66 (complete tail resorption) until termination of the exposure. Therefore, after all animals have reached NF stage 66 or after 70 days (whichever occurs first), a cull should be conducted. Post NF stage 66 animals that will not continue the exposure should be selected at random.

50. Animals that are not selected for continued exposure are euthanised (see para. 43). Measurements of developmental stage, wet weight and SVL (Figure 2b) and a gross necropsy are conducted for each animal. The phenotypic sex (based on gonad morphology) is noted as female, male, or indeterminate.

Juvenile Sampling

Outline of juvenile sampling

51. The remaining animals continue exposure until 10 weeks after the median time to NF stage 62 in the dilution water (and/or solvent control if relevant) control. At the end of the exposure period, the remaining animals (maximum 10 frogs per replicate) are euthanised, and the various endpoints are measured or evaluated and recorded: (1) morphometrics (weight and length), (2) phenotypic/genotypic sex ratios, (3) liver weight (Liver-Somatic Index), (4) histopathology (gonads, reproductive ducts, liver and kidney) and optionally (5) plasma VTG.

Humane killing of frogs

52. The juvenile samples, post-metamorphic frogs, are euthanised by an intraperitoneal injection of anaesthetic, e.g., 10 % MS-222 in an appropriate phosphate buffered solution. Frogs may be sampled after becoming unresponsive (usually around 2 min after injection, if 10 % MS-222 is used in a dosage of 0,01 ml per g of frog). While the juvenile frogs could be immersed in a higher concentration of anaesthetic (MS-222), experience has shown that it takes longer for them to be anaesthetised using this method and the duration may not be adequate to allow for sampling. Injection provides efficient, fast euthanasia prior to sampling. Sampling should not be started until lack of responsiveness of the frogs has been confirmed to ensure that the animals are dead. If frogs are showing signs of considerable suffering (very severe and death can be reliably predicted) and considered moribund, animals should be anaesthetised and euthanised and treated as mortality for data analysis. When a frog is euthanised due to morbidity, this should be noted and reported. Depending on when the frog is euthanised during the study, retaining the frog for histopathology analysis may be conducted (fixing the frog for possible histopathology).

Morphometrics (weight and length)

53. Measurements of wet weight and SVL (Figure 2b) are identical to those outlined for the larval sub-sampling.

Plasma VTG (option)

54. VTG is a widely accepted biomarker resulting from exposure to oestrogenic chemicals. For the LAGDA, plasma VTG optionally may be measured within juvenile samples (this may be particularly relevant if the test chemical is suspected of being an oestrogen).

55. The euthanised juvenile hind limbs are cut and blood is collected with a heparinised capillary (although alternative blood collection methods, such as cardiac puncture, may be suitable). The blood is expelled into a microcentrifuge tube (e.g., 1,5 ml volume) and centrifuged to obtain plasma. The plasma samples should be stored at -70 °C or below until VTG determination. Plasma VTG concentration can be measured by an enzyme-linked immunosorbent assay (ELISA) method (Appendix 6), or by an alternative method such as mass spectrometry (31). Species specific antibodies are preferred due to greater sensitivity.

Genetic sex determination

56. The genetic sex of each juvenile frog is assessed based on the markers developed by Yoshimoto *et al.* (11). To determine the genetic sex, a portion (or whole) of one hind limb (or any other tissue) removed during dissection is collected and stored in a microcentrifuge tube (tissue samples from frogs can be obtained from any tissue). Tissue can be stored at -20°C or below until isolation of deoxyribose nucleic acid (DNA). The isolation of DNA from tissues can be performed with commercially available kits and analysis for presence or absence of the marker is done by a polymerase chain reaction (PCR) method (Appendix 5). Generally, the concordance between histological sex and genotype across control animals at the juvenile sampling time point in control groups is more than 95 %.

Tissue collection and fixation for histopathology

57. Gonads, reproductive ducts, kidneys and livers are collected for histological analysis during the final sampling. The abdominal cavity is opened, and the liver is dissected out and weighed. Next, the digestive organs (*e.g.*, stomach, intestines) are carefully removed from the lower abdomen to reveal the gonads, kidneys and reproductive ducts. Any gross morphological abnormalities in the gonads should be noted. Finally, the hind limbs should be removed if they have not previously been removed for blood collection. Collected livers and the carcass with the gonads left *in situ* should be immediately placed into Davidson's fixative. The volume of fixative in the container should be at least 10 times the approximate volume of the tissues. All tissues remain in Davidson's fixative for at least 48 hours, but no longer than 96 hours at which time they are rinsed in de-ionised water and stored in 10 % neutral buffered formalin (1) (29).

Histopathology

58. Each juvenile sample is evaluated histologically for pathology in the gonads, reproductive ducts, kidneys and liver tissue, *i.e.*, diagnosis and severity grading (32). The gonad phenotype is also derived from this evaluation (*e.g.*, ovary, testis, intersex), and together with individual genetic sex measurements, these observations can be used to calculate phenotypic/genotypic sex ratios.

DATA AND REPORTING

Statistical analysis

59. The LAGDA generates three forms of data to be statistically analysed: (1) quantitative continuous data (weight, SVL, LSI, VTG), (2) time-to-event data for developmental rates (*i.e.*, days to NF stage 62 from assay initiation) and (3) ordinal data in the form of severity scores or developmental stages from histopathology evaluations.
60. It is recommended that the test design and selection of statistical test permit adequate power to detect changes of biological importance in endpoints where a NOEC or EC_x is to be reported. Statistical analyses of the data (generally, replicate mean basis) should preferably follow procedures described in the document *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application* (33). Appendix 7 of this test method provides the recommended statistical analysis decision tree and guidance for the treatment of data and in the choice of the most appropriate statistical test or model to use in the LAGDA.
61. The data from juvenile sampling (*e.g.*, growth, LSI) should be analysed for each genotypic sex separately since genotypic sex is determined for all frogs.

Data analysis considerations

Use of compromised replicates and treatments

62. Replicates and treatments may become compromised due to excess mortality from overt toxicity, disease, or technical error. If a treatment is compromised from disease or technical error, there should be three uncompromised treatments with three uncompromised replicates available for analysis. If overt toxicity occurs in the high treatment(s), it is preferable that at least three treatment levels with three uncompromised replicates are available for analysis (consistent with the Maximum Tolerated Concentration approach for OECD test guidelines (34)). In addition to mortality, signs of overt toxicity may include behavioural effects (e.g. floating on the surface, lying on the bottom of the tank, inverted or irregular swimming, lack of surfacing activity), morphological lesions (e.g. haemorrhagic lesions, abdominal oedema) or inhibition of normal feeding responses when compared qualitatively to control animals.

Solvent control

63. At the termination of the test, an evaluation of the potential effects of the solvent (if used) should be performed. This is done through a statistical comparison of the solvent control group and the dilution water control group. The most relevant endpoints for consideration in this analysis are growth determinants (weight and length), as these can be affected through generalised toxicities. If statistically significant differences are detected in these endpoints between the dilution water control and solvent control groups, best professional judgment should be used to determine if the validity of the test is compromised. If the two controls differ, the treatments exposed to the chemical should be compared to the solvent control unless it is known that comparison to the dilution water control is preferred. If there is no statistically significant difference between the two control groups it is recommended that the treatments exposed to the test chemical are compared with the pooled (solvent and dilution water control groups), unless it is known that comparison to either the dilution-water or solvent control group only is preferred.

Test report

64. The test report should include the following:

Test chemical:

— Physical nature and, where relevant, physicochemical properties;

— Mono-constituent substance:

physical appearance, water solubility, and additional relevant physicochemical properties;

chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc. (including the organic carbon content, if appropriate).

- Multi-constituent substance, UVCBs and mixtures:

characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physico-chemical properties of the constituents.

Test species:

- Scientific name, strain if available, source and method of collection of the fertilised eggs and subsequent handling.
- Incidence of scoliosis in historical controls for the stock culture used.

Test conditions:

- Photoperiod(s);
- Test design (*e.g.*, chamber size, material and water volume, number of test chambers and replicates, number of test organisms per replicate);
- Method of preparation of stock solutions and frequency of renewal (the solubilising agent and its concentration should be given, when used);
- Method of dosing the test chemical (*e.g.*, pumps, diluting systems);
- The recovery efficiency of the method and the nominal test concentrations, the limit of quantification, the means of the measured values and their standard deviations in the test vessels and the method by which these were attained and evidence that the measurements refer to the concentrations of the test chemical in true solution;
- Dilution water characteristics: pH, hardness, temperature, dissolved oxygen concentration, residual chlorine levels (if measured), total iodine, total organic carbon (if measured), suspended solids (if measured), salinity of the test medium (if measured) and any other measurements made;

- The nominal test concentrations, the means of the measured values and their standard deviations;
- Water quality within test vessels, pH, temperature (daily) and dissolved oxygen concentration;
- Detailed information on feeding (*e.g.*, type of foods, source, amount given and frequency).

Results:

- Evidence that controls met the validity criteria;
 - Data for the control (plus solvent control when used) and the treatment groups as follows: mortality and abnormality observed, time to NF stage 62, thyroid histology assessment (larval sample only), growth (weight and length), LSI (juvenile sample only), genetic/phenotypic sex ratios (juvenile sample only), histopathology assessment results for gonads, reproductive ducts, kidney and liver (juvenile sample only) and plasma VTG (juvenile sample only, if performed);
 - Approach for the statistical analysis and treatment of data (statistical test or model used);
 - No observed effect concentration (NOEC) for each response assessed;
 - Lowest observed effect concentration (LOEC) for each response assessed (at $\alpha = 0,05$); EC_x for each response assessed, if applicable, and confidence intervals (*e.g.*, 95 %) and a graph of the fitted model used for its calculation, the slope of the concentration-response curve, the formula of the regression model, the estimated model parameters and their standard errors.
 - Any deviation from the test method and deviations from the acceptance criteria, and considerations of potential consequences on the outcome of the test.
65. For the results of endpoint measurements, mean values and their standard deviations (on both replicate and concentration basis, if possible) should be presented.
66. Median time to NF stage 62 in controls should be calculated and presented as the mean of replicate medians and their standard deviation. Likewise, for treatments, a treatment median should be calculated and presented as the mean of replicate medians and their standard deviation.

LITERATURE

- (1) U.S. Environmental Protection Agency (2013). Validation of the Larval Amphibian Growth and Development Assay: Integrated Summary Report.
- (2) OECD (2012a). Guidance Document on Standardised Test Guidelines for Evaluating Endocrine Disrupters. Environment, Health and Safety Publications, Series on testing and assessment (No 150) Organisation for Economic Cooperation and Development, Paris.

- (3) Nieuwkoop PD and Faber J. (1994). Normal Table of *Xenopus laevis* (Daudin). Garland Publishing, Inc, New York, NY, USA.
- (4) Kloas W and Lutz I. (2006). Amphibians as Model to Study Endocrine Disrupters. *Journal of Chromatography A* 1 130: 16-27.
- (5) Chang C, Witschi E. (1956). Genic Control and Hormonal Reversal of Sex Differentiation in *Xenopus*. *Journal of the Royal Society of Medicine* 93: 140-144.
- (6) Gallien L. (1953). Total Inversion of Sex in *Xenopus laevis* Daud, Following Treatment with Estradiol Benzoate Administered During Larval Stage. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 237: 1 565.
- (7) Villalpando I and Merchant-Larios H. (1990). Determination of the Sensitive Stages for Gonadal Sex-Reversal in *Xenopus Laevis* Tadpoles. *International Journal of Developmental Biology* 34: 281-285.
- (8) Miyata S, Koike S and Kubo T. (1999). Hormonal Reversal and the Genetic Control of Sex Differentiation in *Xenopus*. *Zoological Science* 16: 335-340.
- (9) Mikamo K and Witschi E. (1963). Functional Sex-Reversal in Genetic Females of *Xenopus laevis*, Induced by Implanted Testes. *Genetics* 48: 1 411.
- (10) Olmstead AW, Kosian PA, Korte JJ, Holcombe GW, Woodis K and Degitz SJ. (2009)a. Sex reversal of the Amphibian, *Xenopus tropicalis*, Following Larval Exposure to an Aromatase Inhibitor. *Aquatic Toxicology* 91: 143-150.
- (11) Yoshimoto S, Okada E, Umemoto H, Tamura K, Uno Y, Nishida-Umehara C, Matsuda Y, Takamatsu N, Shiba T and Ito M. (2008). A W-linked DM-Domain Gene, DM-W, Participates in Primary Ovary Development in *Xenopus Laevis*. *Proceedings of the National Academy of Sciences of the United States of America* 105: 2 469-2 474.
- (12) Olmstead AW, Korte JJ, Woodis KK, Bennett BA, Ostazeski S and Degitz SJ. (2009)b. Reproductive Maturation of the Tropical Clawed Frog: *Xenopus tropicalis*. *General and Comparative Endocrinology* 160: 117-123.
- (13) Tobias ML, Tomasson J and Kelley DB. (1998). Attaining and Maintaining Strong Vocal Synapses in Female *Xenopus laevis*. *Journal of Neurobiology* 37: 441-448.
- (14) Qin ZF, Qin XF, Yang L, Li HT, Zhao XR and Xu XB. (2007). Feminizing/Demasculinizing Effects of Polychlorinated Biphenyls on the Secondary Sexual Development of *Xenopus Laevis*. *Aquatic Toxicology* 84: 321-327.

- (15) Porter KL, Olmstead AW, Kumsher DM, Dennis WE, Sprando RL, Holcombe GW, Korte JJ, Lindberg-Livingston A and Degitz SJ. (2011). Effects of 4-Tert-Octylphenol on *Xenopus Tropicalis* in a Long Term Exposure. *Aquatic Toxicology* 103: 159-169.
- (16) ASTM. (2002). Standard Guide for Conducting Acute Toxicity Tests on Test Materials with Fishes, Macro-invertebrates, and Amphibians. ASTM E729-96, Philadelphia, PA, USA.
- (17) Chapter C.4 of this Annex, Ready Biodegradability Test.
- (18) Chapter C.29 of this Annex, Ready Biodegradability - CO₂ in sealed vessels (Headspace Test).
- (19) Kahl MD, Russom CL, DeFoe DL and Hammermeister DE (1999). Saturation Units for Use in Aquatic Bioassays. *Chemosphere* 39: 539-551.
- (20) Adolfsson-Erici M, Åkerman G, Jahnke A, Mayer P, McLachlan MS (2012). A flow-through passive dosing system for continuously supplying aqueous solutions of hydrophobic chemicals to bioconcentration and aquatic toxicity tests. *Chemosphere*, 86(6): 593-9.
- (21) OECD (2000). Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. Environment, Health and Safety Publications, Series on testing and assessment (No 23), Organisation for Economic Cooperation and Development, Paris.
- (22) Hutchinson TH, Shillabeer N, Winter MJ and Pickford DB. (2006). Acute and Chronic Effects of Carrier Solvents in Aquatic Organisms: A Critical Review. *Aquatic Toxicology* 76: 69-92.
- (23) ASTM (2004). Standard Guide for Conducting the Frog Embryo Teratogenesis Assay - *Xenopus* (FETAX). ASTM E1439 - 98, Philadelphia, PA, USA.
- (24) Read BT (2005). Guidance on the Housing and Care of the African Clawed Frog *Xenopus Laevis*. Royal Society for the Prevention of Cruelty to Animals (RSPCA), Horsham, Sussex, U.K., 84 pp.
- (25) Chapter C.38 of this Annex, Amphibian Metamorphosis Assay.
- (26) Chapter C.48 of this Annex, Fish Short Term Reproduction Assay.

- (27) Chapter C.41 of this Annex, Fish Sexual Development Test.
- (28) Chapter C.49 of this Annex, Fish Embryo Acute Toxicity (FET) Test.
- (29) OECD (2007). Guidance Document on Amphibian Thyroid Histology. Environment, Health and Safety Publications, Series on Testing and Assessment. (No 82) Organisation for Economic Cooperation and Development, Paris.
- (30) Grim KC, Wolfe M, Braunbeck T, Iguchi T, Ohta Y, Tooi O, Touart L, Wolf DC and Tietge J. (2009). Thyroid Histopathology Assessments for the Amphibian Metamorphosis Assay to Detect Thyroid-Active Substances, *Toxicological Pathology* 37: 415-424.
- (31) Luna LG and Coady K.(2014). Identification of *X. laevis* Vitellogenin Peptide Biomarkers for Quantification by Liquid Chromatography Tandem Mass Spectrometry. *Analytical and Bioanalytical Techniques* 5(3): 194.
- (32) OECD (2015). Guidance on histopathology techniques and evaluation. Environment, Health and Safety Publications, Series on Testing and Assessment (No 228), Organisation for Economic Cooperation and Development, Paris.
- (33) OECD (2006). Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application. Environment, Health and Safety Publications, Series on testing and assessment (No 54), Organisation for Economic Cooperation and Development, Paris.
- (34) Hutchinson TH, Bögi C, Winter MJ, Owens JW, 2009. Benefits of the Maximum Tolerated Dose (MTD) and Maximum Tolerated concentration (MTC) Concept in Aquatic Toxicology. *Aquatic Toxicology* 91(3): 197-202.

Appendix 1

DEFINITIONS

Apical endpoint: Causing effect at population level.

Chemical: A substance or a mixture

ELISA: Enzyme-Linked Immunosorbent Assay

EC_x: (Effect concentration for x % effect) is the concentration that causes an x % of an effect on test organisms within a given exposure period when compared with a control. For example, an EC₅₀ is a concentration estimated to cause an effect on a test end point in 50 % of an exposed population over a defined exposure period.

dpf: Days post fertilization

Flow-through test: A test with continued flow of test solutions through the test system during the duration of exposure.

HPG axis: hypothalamic-pituitary-gonadal axis

IUPAC: International Union of Pure and Applied Chemistry.

Lowest observed effect concentration (LOEC): is the lowest tested concentration of a test chemical at which the chemical is observed to have a statistically significant effect (at $p < 0.05$) when compared with the control. However, all test concentrations above the LOEC should have a harmful effect equal to or greater than those observed at the LOEC. When these two conditions cannot be satisfied, a full explanation should be given for how the LOEC (and hence the NOEC) has been selected. Appendix 7 provides guidance.

Median Lethal Concentration (LC₅₀): is the concentration of a test chemical that is estimated to be lethal to 50 % of the test organisms within the test duration.

No observed effect concentration (NOEC): is the test concentration immediately below the LOEC, which when compared with the control, has no statistically significant effect ($p < 0.05$), within a stated exposure period.

SMILES: Simplified Molecular Input Line Entry Specification.

Test chemical: Any substance or mixture tested using this Test Method.

UVCB: Substances of unknown or variable composition, complex reaction products or biological materials.

VTG: Vitellogenin is a phospholipoglycoprotein precursor to egg yolk protein that normally occurs in sexually active females of all oviparous species.

Appendix 2

SOME CHEMICAL CHARACTERISTICS OF AN ACCEPTABLE DILUTION WATER

Substance	Limit concentration
Particulate matter	5 mg/l
Total organic carbon	2 mg/l
Un-ionised ammonia	1 µg/l
Residual chlorine	10 µg/l
Total organophosphorous pesticides	50 ng/l
Total organochlorine pesticides plus polychlorinated biphenyls	50 ng/l
Total organic chlorine	25 ng/l
Aluminium	1 µg/l
Arsenic	1 µg/l
Chromium	1 µg/l
Cobalt	1 µg/l
Copper	1 µg/l
Iron	1 µg/l
lead	1 µg/l
Nickel	1 µg/l
Zinc	1 µg/l
Cadmium	100 ng/l
Mercury	100 ng/l
Silver	100 ng/l

Appendix 3

TEST CONDITIONS FOR THE LAGDA

1. Test species	<i>Xenopus laevis</i>
2. Test type	Continuous flow-through,
3. Water temperature	The nominal temperature is 21 °C. The mean temperature over the duration of the test is 21 ± 1 °C (the inter-replicate and the inter-treatment differentials should not exceed 1,0 °C)
4. Illumination quality	Fluorescent bulbs (wide spectrum) 600-2000 lux (lumens/m ²) at the water surface
5. Photoperiod	12 h light:12 h dark
6. Test solution volume and test vessel (tank)	4-10 l (minimum 10–15 cm water depth) Glass or stainless steel tank
7. Volume exchanges of test solutions	Constant, in consideration of both the maintenance of biological conditions and chemical exposure (e.g., 5 tank volume renewal per day)
8. Age of test organisms at initiation	Nieuwkoop and Faber (NF) stage 8-10
9. Number. of organisms per replicate	20 animals (embryos)/tank (replicate) at exposure initiation and 10 animals (juveniles)/tank (replicate) after NF stage 66 to exposure termination
10. Number of treatments	Minimum 4 test chemical treatments plus appropriate control(s)
11. Number of replicates per treatment	4 replicates per treatment for test chemical and 8 replicates for control(s)
12. Number of organisms per test concentration	Minimum 80 animals per treatment for test chemical and minimum 160 animals for control(s)
13. Dilution water	Any water that permits normal growth and development of <i>X. laevis</i> (e.g., spring water or charcoal-filtered tap water)
14. Aeration	None required, but aeration of the tanks may be necessary if dissolved oxygen levels drop below recommended limits and increases in flow of test solution is maximised.
15. Dissolved oxygen of test solution	Dissolved oxygen: ≥ 40 % of air saturation value or ≥ 3,5 mg/l

-
- | | |
|--|--|
| 16. pH of test solution | 6.5-8.5 (the inter-replicate and the inter-treatment differentials should not exceed 0,5) |
| 17. Hardness and alkalinity of test solution | 10-250 mg CaCO ₃ /l |
| 18. Feeding regime | (See Appendix 4) |
| 19. Exposure period | From NF stage 8-10 to ten weeks after the median time to NF stage 62 in water and/or solvent control group (maximum 17 weeks) |
| 20. Biological endpoints | Mortality (and abnormal appearances), time to NF stage 62 (larval sample), thyroid histology assessment (larval sample), growth (weight and length), liver-somatic index (juvenile sample), genetic/phenotypic sex ratios (juvenile sample), histopathology for gonads, reproductive ducts, kidney and liver (juvenile sample) and plasma vitellogenin (juvenile sample, optional) |
| 21. Test validity criteria | Dissolved oxygen should be > 40 % air saturation value; mean water temperature should be 21 ± 1 °C and the inter-replicate and -treatment differentials should be < 1.0 °C; pH of test solution should be ranged between 6,5 and 8,5; the mortality in control should be ≤ 20 % in each replicate, and the mean time to NF stage 62 in control should be ≤ 45 days; the mean weight of test organisms at NF stage 62 and at the termination of the assay in controls and solvent controls (if used) should reach 1,0 ± 0,2 and 11,5 ± 3 g, respectively; evidence should be available to demonstrate that the concentrations of the test chemical in solution have been satisfactorily maintained within ± 20 % of the mean measured values. |

Appendix 4

FEEDING REGIME

It should be noted that although this feeding regime is recommended, alternatives are permissible providing the test organisms grow and develop at an appropriate rate.

Larval feeding*Preparation for larval diet*

A. 1:1 (v/v) Trout Starter: algae/TetraFin® (or equivalent);

1. Trout Starter: blend 50 g of Trout Starter (fine granules or powder) and 300 ml of suitable filtered water on a high blender setting for 20 seconds
2. Algae/TetraFin® (or equivalent) mixture: blend 12 g spirulina algae disks and 500 ml filtered water on a high blender setting for 40 seconds, blend 12 g Tetrafin® (or equivalent) with 500 ml filtered water and then combine these to make up 1 l of 12 g/l spirulina algae and 12 g/l Tetrafin® (or equivalent)
3. Combine equal volumes of the blended Trout Starter and the algae/TetraFin® (or equivalent) mixture

B. Brine shrimp:

15 ml brine shrimp eggs are hatched in 1 l of salt water (prepared by adding 20 ml of NaCl to 1 l deionised water). After aerating 24 hours at room temperature under constant light, the brine shrimp are harvested. Briefly, the brine shrimp are allowed to settle for 30 min by stopping aeration. Cysts that float to the top of the canister are poured off and discarded, and the shrimp are poured through the appropriate filters and brought up to 30 ml with filtered water.

Feeding Protocol

Table 1 provides a reference regarding the type and amount of feed used during the larval stages of the exposure. The animals should be fed three times per day Monday through Friday and once per day on the weekends.

Table 1

Feeding regime for *X. laevis* larvae in flow-through conditions

Time (*) (Post Fertilisation)	Trout Starter: algae/TetraFin®(or equivalent)		Brine Shrimp	
	Weekday (3 times per day)	Weekend (once per day)	Weekday (twice per day)	Weekend (once per day)
Days 4-14 (in Weeks 0-1)	0,33 ml	1,2 ml	0.5 ml (from Day 8 to 15) 1 ml (from Day 16)	0.5 ml (from Day 8 to 15) 1 ml (from Day 16)
Week 2	0,67 ml	2,4 ml		
Week 3	1,3 ml	4,0 ml	1 ml	1 ml
Week 4	1,5 ml	4,0 ml	1 ml	1 ml

Time (*) (Post Fertilisation)	Trout Starter: algae/TetraFin®(or equivalent)		Brine Shrimp	
	Weekday (3 times per day)	Weekend (once per day)	Weekday (twice per day)	Weekend (once per day)
Week 5	1,6 ml	4,4 ml	1 ml	1 ml
Week 6	1,6 ml	4,6 ml	1 ml	1 ml
Week 7	1,7 ml	4,6 ml	1 ml	1 ml
Weeks 8-10	1,7 ml	4,6 ml	1 ml	1 ml

(*) Day 0 is defined as the day hCG injection is done.

Larval to juvenile diet transition

As larvae complete metamorphosis, they transition to a juvenile diet formulation explained below. While this transition is taking place, the larval diet should be reduced as the juvenile feed increases. This can be accomplished by proportionally decreasing the larval feed while proportionally increasing the juvenile feed as each group of five tadpoles surpass NF stage 62 and approach completion of metamorphosis at NF stage 66.

Juvenile feeding

Juvenile diet

Once metamorphosis is complete (stage 66), the feeding regime changes to 3/32 inch premium sinking frog food alone (Xenopus Express™, FL, USA), or equivalent.

Preparation of crushed pellet for larval to juvenile transition

Sinking frog food pellets are briefly run in a coffee grinder, blender or mortar and pestle in order to reduce the size of the pellets by approximately 1/3. Processing too long results in powder and is discouraged.

Feeding protocol

Table 2 provides a reference regarding the type and amount of feed used during juvenile and adult life stages. The animals should be fed once per day. It should be noted that as animals metamorphose, they continue receiving a portion of the brine shrimp until > 95 % of animals complete metamorphosis.

The animals should not be fed on the day of test termination so feed does not confound weight measurements.

Table 2

Feeding regime for *X. laevis* juveniles in flow-through conditions. It should be noted that unmetamorphosed animals, including those whose metamorphosis has been delayed by the chemical treatment, cannot eat uncrushed pellets

Time (*) (Weeks post-median metamorphosis date)	Crushed pellet (mg per froglet)	Whole pellet (mg per froglet)
As animals complete metamorphosis	25	0
Weeks 0-1	25	28
Weeks 2-3	0	110
Weeks 4-5	0	165
Weeks 6-9	0	220

(*) The first day of Week 0 is the median metamorphosis date in control animals.

Appendix 5

GENETIC SEX DETERMINATION (GENETIC SEXING)

The method of genetic sexing for *Xenopus laevis* is based on Yoshimoto *et al.*, 2008. Procedures in detail on the genotyping can be obtained from this publication, if needed. Alternative methods (e.g. high-throughput qPCR) may be used if considered suitable.

X. laevis primers*DM-W marker*

Forward: 5'-CCACACCCAGCTCATGTAAAG-3'

Reverse: 5'-GGGCAGAGTCACATATACTG-3'

Positive Control

Forward: 5'-AACAGGAGCCCAATTCTGAG-3'

Reverse: 5'-AACTGCTTGACCTCTAATGC-3'

DNA purification

Purify DNA from muscle or skin tissue using e.g., Qiagen DNeasy Blood and Tissue Kit (cat # 69 506) or similar product according to kit instructions. DNA can be eluted from the spin columns using less buffer to yield more concentrated samples if deemed necessary for PCR. Note that DNA is quite stable, so care should be taken to avoid cross-contamination that could lead to mischaracterisation of males as females, or vice versa.

PCR

A sample protocol using JumpStart™ Taq from Sigma is outlined in **Table 1**.

Table 1

Sample protocol using JumpStart™ Taq from Sigma

Master Mix	1x (µl)	[Final]
NFW (*)	11	—
10X Buffer	2,0	—
MgCl ₂ (25mM)	2,0	2,5 mM
dNTP's (10mM each)	0,4	200 µM
Marker for primer (8 µM)	0,8	0,3 µM
Marker rev primer (8 µM)	0,8	0,3 µM
Control for primer (8 µM)	0,8	0,3 µM
Control rev primer (8 µM)	0,8	0,3 µM

Master Mix	1x (µl)	[Final]
JumpStart™ Taq	0,4	0,05 units/µl
DNA template	1,0	~200 pg/µl

(*) Nuclease-free water

Note: When preparing Master Mixes, prepare extra to account for any loss that may occur while pipetting (example: 25x should be used for only 24 reactions).

Reaction:

Master Mix	19,0 µl
Template	1,0 µl
Total	20.0 µl

Thermocycler Profile:

Step 1.	94 °C	1 min
Step 2.	94 °C	30 sec
Step 3.	60 °C	30 sec
Step 4.	72 °C	1 min
Step 5.	Go to step 2.	35 cycles
Step 6.	72 °C	1 min
Step 7.	4 °C	hold

PCR products can be run immediately in a gel or stored at 4 °C.

Agarose Gel Electrophoresis (3 %)(sample protocol)

50X TAE

Tris	24,2 g
Glacial acetic acid	5,71 ml
Na ₂ (EDTA)·2H ₂ O	3,72 g

Add water to 100 ml

1X TAE

H ₂ O	392 ml
50X TAE	8 ml

3:1 Agarose

3 parts NuSieve™ GTG™ agarose

1 part Fisher agarose low electroendosmosis (EEO)

Method

1. Prepare a 3 % gel by adding 1,2 g agarose mix to 43 ml 1X TAE. Swirl to disassociate large clumps.
2. Microwave agarose mixture until completely dissolved (avoid boiling over). Let cool slightly.
3. Add 1,0 µL ethidium bromide (10 mg/ml). Swirl flask. Note that ethidium bromide is mutagenic, so alternative chemicals should, in so far as is technically possible, be used for this step to minimise health risks to workers ⁽¹⁾.
4. Pour gel into mould with comb. Cool completely.
5. Add gel to apparatus. Cover gel with 1X TAE.
6. Add 1 µl of 6x loading dye to each 10 µl PCR product.
7. Pipette samples into wells.
8. Run at 160 constant volts for ~20 minutes.

An agarose gel image showing the band patterns indicative of male and female individuals is shown in **Figure 1**.

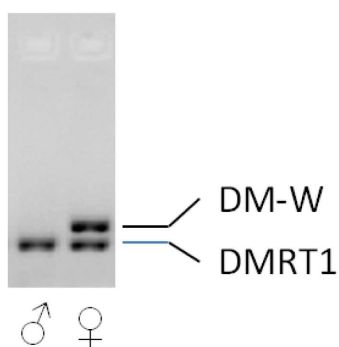


Figure 1: Agarose gel image showing the band pattern indicative of a male (♂) individual (single band ~203 bp: DMRT1) and of a female (♀) individual (two bands at ~259 bp: DM-W and 203 bp:DMRT1).

LITERATURE

Yoshimoto S, Okada E, Umemoto H, Tamura K, Uno Y, Nishida-Umehara C, Matsuda Y, Takamatsu N, Shiba T, Ito M. 2008. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. Proceedings of the National Academy of Sciences of the United States of America 105: 2469-2474.

⁽¹⁾ In accordance to Article 4.1 of Directive 2004/37/EC of the European Parliament and of the Council of 29 April 2004 on the protection of workers from the risks related to exposure to carcinogens or mutagens at work (Sixth individual Directive within the meaning of Article 16(1) of Council Directive 89/391/EEC) (OJ L 158, 30.4.2004, p. 50).

Appendix 6

MEASUREMENT OF VITELLOGENIN

The measurement of vitellogenin (VTG) is made using an enzyme-linked immunosorbent assay (ELISA) method which was originally developed for fathead minnow VTG (Parks *et al.*, 1999). Currently there are no commercially available antibodies for *X. laevis*. However, given the wealth of information for this protein and the availability of cost-effective commercial antibody production services, it is reasonable that laboratories can easily develop an ELISA to make this measurement (Olmstead *et al.*, 2009). Also Olmstead *et al.* (2009) provide a description of the assay as modified for VTG in *X. tropicalis*, as shown below. The method uses an antibody made against *X. tropicalis* VTG, but it is known also to work for *X. laevis* VTG. It should be noted that non-competitive ELISAs can also be used, and that these may have lower detection limits than the method described below.

Materials and Reagents

- Preadsorbed 1st Antibody (Ab) serum

- Mix 1 part anti-*X. tropicalis* VTG 1st Ab serum with 2 parts control male plasma and leave at RT for ~ 75 minutes, put on ice for 30 min, centrifuge > 20K x G for 1 hour at 4 °C, remove supernatant, aliquot, store at -20 °C.

- 2nd Antibody

- Goat Anti-Rabbit IgG-HRP conjugate (e.g., Bio-Rad 172-1019)

- VTG Standard

- purified *X. laevis* VTG at 3,3 mg/ml.

- TMB (3,3',5,5' Tetramethyl-benzidine) (e.g., KPL 50-76-00; or Sigma T0440)

- Normal Goat Serum (NGS) (e.g., Chemicon® S26-100ml)

- 96 well EIA polystyrene microtiter plates (e.g., ICN: 76-381-04, Costar: 53590, Fisher: 07-200-35)

- 37 °C hybridization oven (or fast equilibrating air incubator) for plates, water bath for tubes

- Other common laboratory equipment, chemicals, and supplies.

Recipes

Coating Buffer (50 mM Carbonate Buffer, pH 9.6):

NaHCO ₃	1,26 g
Na ₂ CO ₃	0,68 g
water	428 ml

10X PBS (0.1 M phosphate, 1,5 M NaCl):

NaH ₂ PO ₄ ·H ₂ O	0,83 g
Na ₂ HPO ₄ ·7 H ₂ O	20,1 g
NaCl	71 g
water	810 ml

Wash Buffer (PBST):

10X PBS	100 ml
Water	900 ml

Adjust pH to 7.3 with 1 M HCl, then add 0,5 ml Tween-20

Assay Buffer:

Normal Goat Serum (NGS)	3,75 ml
Wash Buffer	146,25 ml

Sample collection

Blood is collected with a heparinised microhematocrit tube and placed on ice. After centrifugation for 3 minutes, the tube is scored, broken open, and the plasma expelled into 0,6 ml microcentrifuge tubes which contain 0,13 units of lyophilised aprotinin. (These tubes are prepared in advance by adding the appropriate amount of aprotinin, freezing, and lyophilising in a speed-vac at low heat until dry.) Store plasma at -80 °C until analysed.

Procedure for one plate

Coating the plate

Mix 20 µl of purified VTG with 22 ml of carbonate buffer (final 3 µg/ml). Add 200 µl to each well of a 96-well plate. Cover the plate with adhesive sealing film and allow to incubate at 37 °C for 2 hours (or 4 °C overnight).

Blocking the plate

Blocking solution is prepared by adding 2 ml of Normal Goat Serum (NGS) to 38 ml of carbonate buffer. Remove coating solution and shake dry. Add 350 µl of the blocking solution to each well. Cover with adhesive sealing film and incubate at 37 °C for 2 hours (or at 4 °C overnight).

Preparation of standards

5.8 µl of purified VTG standard is mixed with 1,5 ml of assay buffer in a 12 x 75 mm borosilicate disposable glass test tube. This yields 12 760 ng/ml. Then a serial dilution is performed by adding 750 µl of the previous dilution to 750 µl of assay buffer to yield final concentrations of 12 760, 6 380, 3 190, 1 595, 798, 399, 199, 100, and 50 ng/ml.

Preparation of Samples

Start with a 1:300 (e.g., combine 1 μl plasma with 299 μl of assay buffer) or 1:30 dilution of plasma into assay buffer. If a large amount of VTG is expected, additional or greater dilutions may be needed. Try to keep B/B_0 within the range of standards. For samples without appreciable VTG, e.g., control males and females (which are all immature), use the 1:30 dilution. Samples diluted less than this may show unwanted matrix effects.

Additionally, it is recommended to run a positive control sample on each plate. This comes from a pool of plasma containing high induced levels of VTG. The pool is initially diluted in NGS, divided in aliquots and stored at -80 C . For each plate, an aliquot is thawed, diluted further in assay buffer and run similar to a test sample.

Incubation with 1st antibody

The 1st Ab is prepared by making a 1:2000 dilution of preadsorbed 1st Ab serum in assay buffer (e.g., 8 μl to 16 ml of assay buffer). Combine 300 μl of 1st Ab solution with 300 μl of sample/standard in a glass tube. The B_0 tube is prepared similarly with 300 μl of assay buffer and 300 μl of antibody. Also, a NSB tube should be prepared using 600 μl of assay buffer only, i.e., no Ab. Cover the tubes with Parafilm and vortex gently to mix. Incubate in a $37\text{ }^\circ\text{C}$ water bath for 1 hour.

Washing the plate

Just before the 1st Ab incubation is complete, wash the plate. This is done by shaking out the contents and patting dry on absorbent paper. Then fill wells with 350 μl of wash solution, dump out, and pat dry. A multi-channel repeater pipette or plate washer is useful here. The wash step is repeated two more times for a total of three washes.

Loading the plate

After the plate has been washed, remove the tubes from the water bath and vortex lightly. Add 200 μl from each sample, standard, B_0 , and NSB tube to duplicate wells of the plate. Cover plate with adhesive sealing film and allow to incubate for 1 hour at $37\text{ }^\circ\text{C}$.

Incubation with the 2nd antibody

At the end of the incubation from the previous step, the plate should be washed three times again, like above. The diluted 2nd Ab is prepared by mixing 2,5 μl of 2nd Ab with 50 ml of assay buffer. Add 200 μl of diluted 2nd Ab to each well, seal like above, and incubate for 1 hour at $37\text{ }^\circ\text{C}$.

Addition of substrate

After the incubation with the 2nd Ab is complete, wash the plate three times as described earlier. Then add 100 μl of TMB substrate to each well. Allow the reaction to proceed for 10 minutes, preferably out of bright light. Stop the reaction by adding 100 μl of 1 M phosphoric acid. This will change the colour from blue to an intense yellow. Measure the absorbance at 450 nm using a plate reader.

Calculate B/Bo

Subtract the average NSB value from all measurements. The B/B_0 for each sample and standard is calculated by dividing the absorbance value (B) by the average absorbance of the B_0 sample.

Obtain the standard curve and determine unknown amounts

Generate a standard curve with the aid of some computer graphing software (e.g., SlidewriteTM or Sigma Plot[®]) that will extrapolate quantity from B/B_0 of sample based on B/B_0 of standards. Typically, the amount is plotted on a log scale and the curve has a sigmoid shape. However, it may appear linear when using a narrow range of standards. Correct sample amounts for dilution factor and report as mg VTG/ml of plasma.

Determination of minimum detection limits (MDL)

Often, particularly in normal males, it will not be clear how to report results from low values. In these cases, the 95 % "Confidence limits" should be used to determine if the value should be reported as zero or as some other number. If the sample result is within the confidence interval of the zero standard (B_0), the result should be reported as zero. The minimum detection level will be the lowest standard which is consistently different from the zero standard; that is, the two confidence intervals don't overlap. For any sample result which is within the confidence limit of the minimum detection level, or above, the calculated value will be reported. If a sample falls between the zero standard and the minimum detection level confidence intervals, one half of the minimum detection level should be reported for the value of that sample.

LITERATURE

Olmstead AW, Korte JJ, Woodis KK, Bennett BA, Ostazeski S, Degitz SJ. 2009. Reproductive maturation of the tropical clawed frog: *Xenopus tropicalis*. *General and Comparative Endocrinology* 160: 117-123.

Parks LG, Cheek AO, Denslow ND, Heppell SA, McLachlan JA, LeBlanc GA, Sullivan CV. 1999. Fathead minnow (*Pimephales promelas*) vitellogenin: purification, characterisation and quantitative immunoassay for the detection of estrogenic compounds. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 123: 113-125.

Appendix 7

STATISTICAL ANALYSIS

The LAGDA generates three forms of data to be statistically analysed: (1) Quantitative continuous data, (2) Time-to-event data for developmental rates (Time to NF stage 62) and (3) Ordinal data in the form of severity scores or developmental stages from histopathology evaluations. The recommended statistical analysis decision tree for the LAGDA is shown in Figure 1. Also, some annotations which might be needed to conduct statistical analysis for the measurements from the LAGDA are indicated below. For the analysis decision tree, the results of measurements for mortality, growth (weight and length) and liver-somatic-index (LSI) should be analysed according to the “Other endpoints” branch.

Continuous data

Data for continuous endpoints should first be checked for monotonicity by rank transforming the data, fitting to an ANOVA model and comparing linear and quadratic contrasts. If the data are monotonic, a step-down Jonckheere-Terpstra trend test should be performed on replicate medians and no subsequent analyses should be applied. An alternative for data that are normally distributed with homogeneous variances is the step-down Williams' test. If the data are non-monotonic (quadratic contrast is significant and linear is not significant), they should be analysed using a mixed effects ANOVA model. The data should then be assessed for normality (preferably using the Shapiro-Wilk or Anderson-Darling test) and variance homogeneity (preferably using Levene's test). Both tests are performed on the residuals from the mixed effects ANOVA model. Expert judgment can be used in lieu of these formal tests for normality and variance homogeneity, though formal tests are preferred. If the data are normally distributed with homogeneous variance, then the assumptions of a mixed effect ANOVA are met and a significant treatment effect is determined from Dunnett's test. Where non-normality or variance heterogeneity is found, then the assumptions of Dunnett's test are violated and a normalising, variance stabilising transform is sought. If no such transform is found, then a significant treatment effect is determined with a Dunn's test. Whenever possible, a one-tailed test should be performed as opposed to a two-tailed test, but it requires expert judgment to determine which is appropriate for a given endpoint.

Mortality

Mortality data should be analysed for the time period encompassing the full test and should be expressed as proportion that died in any particular tank. Tadpoles that do not complete metamorphosis in the given time frame, those tadpoles that are in the larval sub-sample cohort, those juvenile frogs that are culled, and any animal that dies due to experimenter error should be treated as censored data and not included in the denominator of the percent calculation. Prior to any statistical analyses, mortality proportions should be arcsin-square root transformed. An alternative is to use the step-down Cochran-Armitage test, possibly with a Rao-Scott adjustment in the presence of overdispersion.

Weight and length (growth data)

Males and females are not sexually-dimorphic during metamorphosis so larval sub-sampling growth data should be analysed independent of gender. However, juvenile growth data should be analysed separately based on genetic sex. A log-transformation may be needed for these endpoints since log-normality of size data is not uncommon.

Liver-somatic-index (LSI)

Liver weights should be normalised as proportions of whole body weights (i.e., LSI) and analysed separately based on genetic sex.

Time to NF stage 62

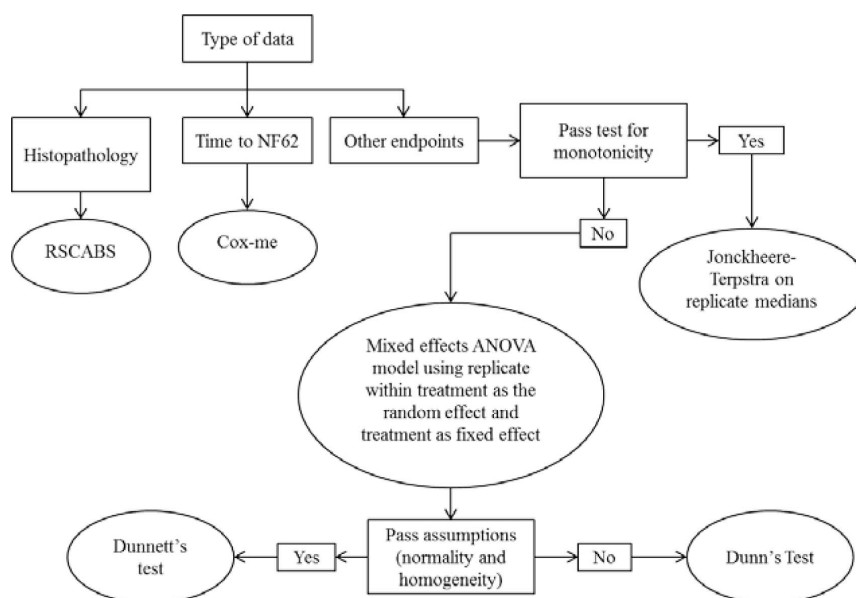
Time to metamorphosis data should be treated as time-to-event data, with any mortalities or individuals not reaching NF stage 62 in 70 days treated as right-censored data (i.e. the true value is greater than 70 days but the study ends before the animals had reached NF stage 62 in 70 days). Median time to NF stage 62 completion of metamorphosis in dilution water controls should be used to determine the test termination date. Median time to completion of metamorphosis could be determined by Kaplan-Meier product-limit estimators. This endpoint should be analysed using a mixed-effects Cox proportional hazard model that takes account of the replicate structure of the study.

Histopathology data (severity scores and developmental stages)

Histopathology data are in the form of severity scores or developmental stages. A test termed RSCABS (Rao-Scott Cochran-Armitage by Slices) uses a step-down Rao-Scott adjusted Cochran-Armitage trend test on each level of severity in a histopathology response (Green *et al.*, 2014). The Rao-Scott adjustment incorporates the replicate vessel experimental design into the test. The “by Slices” procedure incorporates the biological expectation that severity of effect tends to increase with increasing doses or concentrations, while retaining the individual subject scores and revealing the severity of any effect found. The RSCABS procedure not only determines which treatments are statistically different from controls (i.e., have more severe pathology than controls), but it also determines at which severity score the difference occurs thereby providing much needed context to the analysis. In the case of developmental staging of gonads and reproductive ducts, an additional manipulation should be applied to the data since an assumption of RSCABS is that severity of effect increases with dose. The effect observed could be a delay or acceleration of development. Therefore, developmental staging data should be analysed as reported to detect acceleration in development and then manually inverted prior to a second analysis to detect a delay in development.

Figure 1

Statistical analysis decision tree for LAGDA data



LITERATURE

Green JW, Springer TA, Saulnier AN, Swintek J. 2014. Statistical analysis of histopathology endpoints. *Environmental Toxicology and Chemistry* 33, 1 108-1 116.

Appendix 8

CONSIDERATIONS FOR TRACKING AND MINIMISING THE OCCURRENCE OF SCOLIOSIS

Idiopathic scoliosis, usually manifesting as “bent tail” in *Xenopus laevis* tadpoles, may complicate morphological and behavioural observations in test populations. Efforts should be made to minimise or eliminate the incidence of scoliosis, both in stock and under test conditions. In the definitive test, it is recommended that the prevalence of moderate and severe scoliosis be less than 10 %, to improve confidence that the test can detect treatment-related developmental effects in otherwise healthy amphibian larvae.

Daily observations during the definitive test should record both the incidence (individual count) and severity of scoliosis, when present. The nature of the abnormality should be described with respect to location (*e.g.*, anterior or posterior to the vent) and direction of curvature (*e.g.*, lateral or dorsal-to-ventral). Severity may be graded as follows:

(NR) Not remarkable: no curvature present

(1) Minimal: slight, lateral curvature posterior to the vent; apparent only at rest

(2) Moderate: lateral curvature posterior to the vent; visible at all times but does not inhibit movement

(3) Severe: lateral curvature anterior to the vent; OR any curvature that inhibits movement; OR any dorsal-to-ventral curvature

A US EPA FIFRA Scientific Advisory Panel (FIFRA SAP 2013) reviewed summary data for scoliosis in fifteen Amphibian Metamorphosis Assays with *X. laevis* (NF stage 51 through 60+) and provided general recommendations for reducing the prevalence of this abnormality in test populations. The recommendations are relevant to the LAGDA even though this test encompasses a longer developmental timeline.

Historical Spawning Performance

Generally, high quality, healthy adults should be used as breeding pairs; eliminating breeding pairs that produce offspring with scoliosis may minimise its occurrence over time. Specifically, minimising the use of wild-caught breeding stock may be beneficial. The LAGDA exposure period begins with NF stage 8-to-10 embryos, and it is not feasible to determine at the test outset whether given individuals will exhibit scoliosis. Thus, in addition to tracking the incidence of scoliosis in animals that are placed on test, historical clutch performance (including the prevalence of scoliosis in any larvae allowed to develop) should be documented. It may be useful to further monitor the portion of each clutch not used in a given study and to report these observations (FIFRA SAP 2013).

Water Quality

It is important to ensure adequate water quality, both in laboratory stock and during the test. In addition to water quality criteria routinely evaluated for aquatic toxicity tests, it may be useful to monitor for and to correct any nutrient deficiencies (*e.g.*, deficiency of vitamin C, calcium, phosphorus) or excess levels of selenium and copper, which are reported to cause scoliosis to varying degrees in laboratory-reared *Rana* sp. and *Xenopus* sp. (Marshall *et al.* 1980; Leibovitz *et al.* 1992; Martinez *et al.* 1992; as reported in FIFRA SAP 2013). The use of an appropriate dietary regimen (see Appendix 4), and regular tank cleaning, will generally improve water quality and health of the test specimens.

Diet

Specific recommendations for a dietary regimen, found to be successful in the LAGDA, are detailed in Appendix 4. It is recommended that feed sources be screened for biological toxins, herbicides, and other pesticides which are known to cause scoliosis in *X. laevis* or other aquatic animals (Schlenk and Jenkins 2013). For example, exposure to certain cholinesterase inhibitors has been associated with scoliosis in fish (Schultz *et al.* 1985) and frogs (Bacchetta *et al.* 2008).

LITERATURE

Bacchetta, R., P. Mantecca, M. Andrioletti, C. Vismara, and G. Vailati. 2008. Axial-skeletal defects caused by carbaryl in *Xenopus laevis* embryos. *Science of the Total Environment* 392: 110 – 118.

Schultz, T.W., J.N. Dumont, and R.G. Epler. 1985. The embryotoxic and osteolathrogenic effects of semicarbazide. *Toxicology* 36: 185-198.

Leibovitz, H.E., D.D. Culley, and J.P. Geaghan. 1982. Effects of vitamin C and sodium benzoate on survival, growth and skeletal deformities of intensively culture bullfrog larvae (*Rana catesbeiana*) reared at two pH levels. *Journal of the World Aquaculture Society* 13: 322-328.

Marshall, G.A., R.L. Amborski, and D.D. Culley. 1980. Calcium and pH requirements in the culture of bullfrog (*Rana catesbeiana*) larvae. *Journal of the World Aquaculture Society* 11: 445-453.

Martinez, I., R. Alvarez, I. Herraiez, and P. Herraiez. 1992. Skeletal malformations in hatchery reared *Rana perezi* tadpoles. *Anatomical Records* 233(2): 314-320.

Schlenk, D., and Jenkins, F. 2013. Endocrine Disruptor Screening Prog (EDSP) Tier 1 Screening Assays and Battery Performance. US EPA FIFRA SAP Minutes No. 2013-03. May 21-23, 2013. Washington, DC."
