

II

(Non-legislative acts)

REGULATIONS

COMMISSION REGULATION (EU) 2017/735

of 14 February 2017

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) It is necessary to update Regulation (EC) No 440/2008 to include new and updated test methods recently adopted by the Organisation for Economic Cooperation and Development (OECD) in order to take into account technical progress, and to ensure the reduction in the number of animals to be used for experimental purposes, in accordance with Directive 2010/63/EU of the European Parliament and of the Council ⁽³⁾. Stakeholders have been consulted on this draft.
- (3) The adaptation to technical progress contains 20 test methods: one new method for the determination of a physicochemical property, five new and one updated test methods for the assessment of ecotoxicity, two updated test methods to assess the environmental fate and behaviour, and four new and seven updated test methods for the determination of effects on human health.
- (4) The OECD regularly reviews its test guidelines in order to identify those which are scientifically obsolete. This adaptation to technical progress deletes six test methods for which the corresponding OECD test guidelines have been cancelled.

⁽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).

⁽³⁾ Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes (OJ L 276, 20.10.2010, p. 33).

- (5) Regulation (EC) No 440/2008 should therefore be amended accordingly.
- (6) 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, 14 February 2017.

For the Commission

The President

Jean-Claude JUNKER

ANNEX

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

- (1) In Part A, the following Chapter is added:

'A.25 DISSOCIATION CONSTANTS IN WATER (TITRATION METHOD — SPECTROPHOTOMETRIC METHOD — CONDUCTOMETRIC METHOD)

INTRODUCTION

This test method is equivalent to OECD test guideline 112 (1981)

Prerequisites

- Suitable analytical method
- Water solubility

Guidance information

- Structural formula
- Electrical conductivity for conductometric method

Qualifying statements

- All test methods may be carried out on pure or commercial grade substances. The possible effects of impurities on results should be considered.
- The titration method is not suitable for low solubility substances (see Test solutions, below).
- The spectrophotometric method is only applicable to substances having appreciably different UV/VIS-absorption spectra for the dissociated and undissociated forms. This method may also be suitable for low solubility substances and for non-acid/base dissociations, e.g. complex formation.
- In cases where the Onsager equation holds, the conductometric method may be used, even at moderately low concentrations and even in cases for non-acid/base equilibria.

Standard documents

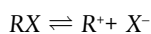
This test method is based on methods given in the references listed in the section "Literature" and on the Preliminary Draft Guidance for Premanufacture Notification EPA, August 18, 1978.

METHOD — INTRODUCTION, PURPOSE, SCOPE, RELEVANCE, APPLICATION AND LIMITS OF TEST

The dissociation of a substance in water is of importance in assessing its impact upon the environment. It governs the form of the substance which in turn determines its behaviour and transport. It may affect the adsorption of the chemical on soils and sediments and absorption into biological cells.

Definitions and units

Dissociation is the reversible splitting into two or more chemical species which may be ionic. The process is indicated generally by



and the concentration equilibrium constant governing the reaction is

$$K = \frac{[R^{+}][X^{-}]}{[RX]}$$

For example, in the particular case where R is hydrogen (the substance is an acid), the constant is

$$K_a = [H^+] \cdot \frac{[X^-]}{[HX]}$$

or

$$pK_a = pH - \log \frac{[X^-]}{[HX]}$$

Reference substances

The following reference substances need not be employed in all cases when investigating a new substance. They are provided primarily so that calibration of the method may be performed from time to time and to offer the chance to compare the results when another method is applied.

	pK _a ⁽¹⁾	Temp. in °C
p-Nitrophenol	7,15	25 ⁽¹⁾
Benzoic acid	4,12	20
p-Chloroaniline	3,93	20

⁽¹⁾ No value for 20 °C is available, but it can be assumed that the variability of measurement results is higher than the temperature dependence to be expected.

It would be useful to have a substance with several pKs as indicated in Principle of the method, below. Such a substance could be:

Citric acid	pK _a (8)	Temp. in °C
	(1) 3,14	20
	(2) 4,77	20
	(3) 6,39	20

Principle of the test method

The chemical process described is generally only slightly temperature dependent in the environmentally relevant temperature range. The determination of the dissociation constant requires a measure of the concentrations of the dissociated and undissociated forms of the chemical substance. From the knowledge of the stoichiometry of the dissociation reaction indicated in Definitions and units, above, the appropriate constant can be determined. In the particular case described in this test method the substance is behaving as an acid or a base, and the determination is most conveniently done by determining the relative concentrations of the ionised and unionised forms of the substance and the pH of the solution. The relationship between these terms is given in the equation for pK_a in Definitions and units, above. Some substances exhibit more than one dissociation constant and similar equations can be developed. Some of the methods described herein are also suitable for non-acid/base dissociation.

Quality criteria

Repeatability

The dissociation constant should be replicated (a minimum of three determinations) to within ± 0,1 log units.

DESCRIPTION OF THE TEST PROCEDURES

There are two basic approaches to the determination of pK_a . One involves titrating a known amount of substance with standard acid or base, as appropriate; the other involves determining the relative concentration of the ionised and unionised forms and its pH dependence.

Preparations

Methods based on those principles may be classified as titration, spectrophotometric and conductometric procedures.

Test solutions

For the titration method and conductometric method the chemical substance should be dissolved in distilled water. For spectrophotometric and other methods buffer solutions are used. The concentration of the test substance should not exceed the lesser of 0,01 M or half the saturation concentration, and the purest available form of the substance should be employed in making up the solutions. If the substance is only sparingly soluble, it may be dissolved in a small amount of a water-miscible solvent prior to adding to the concentrations indicated above.

Solutions should be checked for the presence of emulsions using a Tyndall beam, especially if a co-solvent has been used to enhance solubility. Where buffer solutions are used, the buffer concentration should not exceed 0,05 M.

Test conditions

Temperature

The temperature should be controlled to at least ± 1 °C. The determination should preferably be carried out at 20 °C.

If a significant temperature dependence is suspected, the determination should be carried out at least at two other temperatures. The temperature intervals should be 10 °C in this case and the temperature control $\pm 0,1$ °C.

Analyses

The method will be determined by the nature of the substance being tested. It must be sufficiently sensitive to allow the determination of the different species at each test solution concentration.

Performance of the test

Titration method

The test solution is determined by titration with the standard base or acid solution as appropriate, measuring the pH after each addition of titrant. At least 10 incremental additions should be made before the equivalence point. If equilibrium is reached sufficiently rapidly, a recording potentiometer may be used. For this method both the total quantity of substance and its concentration need to be accurately known. Precautions must be taken to exclude carbon dioxide. Details of procedure, precautions, and calculation are given in standard tests, e.g. references (1), (2), (3), (4).

Spectrophotometric method

A wavelength is found where the ionised and unionised forms of the substance have appreciably different extinction coefficients. The UV/VIS absorption spectrum is obtained from solutions of constant concentration under a pH condition where the substance is essentially unionised and fully ionised and at several intermediate pHs. This may be done, either by adding increments of concentrated acid (base) to a relatively large volume of a solution of the substance in a multicomponent buffer, initially at high (low) pH (ref. 5), or by adding equal volumes of a stock solution of the substance in e.g. water, methanol, to constant volumes of various buffer solutions covering the desired pH range. From the pH and absorbance values at the chosen wavelength, a sufficient number of values for the pK_a is calculated using data from at least 5 pHs where the substance is at least 10 per cent and less than 90 per cent ionised. Further experimental details and method of calculation are given in reference (1).

Conductometric method

Using a cell of small, known cell constant, the conductivity of an approximately 0,1 M solution of the substance in conductivity water is measured. The conductivities of a number of accurately-made dilutions of this solution are also measured. The concentration is halved each time, and the series should cover at least an order of magnitude in concentration. The limiting conductivity at infinite dilution is found by carrying out a similar experiment with the Na salt and extrapolating. The degree of dissociation may then be calculated from the conductivity of each solution using the Onsager equation, and hence using the Ostwald Dilution Law the dissociation constant may be calculated as $K = \alpha^2 C / (1 - \alpha)$ where C is the concentration in moles per litre and α is the fraction dissociated. Precautions must be taken to exclude CO₂. Further experimental details and method of calculation are given in standard texts and references (1), (6) and (7).

DATA AND REPORTING

Treatment of results

Titration method

The pK_a is calculated for 10 measured points on the titration curve. The mean and standard deviation of such pK_a values are calculated. A plot of pH versus volume of standard base or acid should be included along with a tabular presentation.

Spectrophotometric methods

The absorbance and pH are tabulated from each spectrum. At least five values for the pK_a are calculated from the intermediate spectra data points, and the mean and standard deviation of these results are also calculated.

Conductometric method

The equivalent conductivity Λ is calculated for each acid concentration and for each concentration of a mixture of one equivalent of acid, plus 0,98 equivalent of carbonate-free sodium hydroxide. The acid is in excess to prevent an excess of OH⁻ due to hydrolysis. $1/\Lambda$ is plotted against \sqrt{C} and Λ_0 of the salt can be found by extrapolation to zero concentration.

Λ_0 of the acid can be calculated using literature values for H⁺ and Na⁺. The pK_a can be calculated from $\alpha = \Lambda_i / \Lambda_0$ and $K_a = \alpha^2 C / (1 - \alpha)$ for each concentration. Better values for K_a can be obtained by making corrections for mobility and activity. The mean and standard deviations of the pK_a values should be calculated.

Test report

All raw data and calculated pK_a values should be submitted together with the method of calculation (preferably in a tabulated format, such as suggested in ref. 1) as should the statistical parameters described above. For titration methods, details of the standardisation of titrants should be given.

For the spectrophotometric method, all spectra should be submitted. For the conductometric method, details of the cell constant determination should be reported. Information on technique used, analytical methods and the nature of any buffers used should be given.

The test temperature(s) should be reported.

LITERATURE:

- (1) Albert, A. & Sergeant, E.P.: *Ionization Constants of Acids and Bases*, Wiley, Inc., New York, 1962.
- (2) Nelson, N.H. & Faust, S.D.: Acidic dissociation constants of selected aquatic herbicides, *Env. Sci. Tech.* 3, II, pp. 1186-1188 (1969).

- (3) ASTM D 1293 — Annual ASTM Standards, Philadelphia, 1974.
 - (4) Standard Method 242. APHA/AWWA/WPCF, *Standard Methods for the Examination of Water and Waste Water*, 14th Edition, American Public Health Association, Washington, D.C., 1976.
 - (5) Clark, J. & Cunliffe, A.E.: Rapid spectrophotometric measurement of ionisation constants in aqueous solution. *Chem. Ind. (London)* 281, (March 1973).
 - (6) ASTM D 1125 — Annual ASTM Standards, Philadelphia, 1974.
 - (7) Standard Method 205 — APHA/AWWA/NPCF (see above (4)).
 - (8) *Handbook of Chemistry and Physics*, 60th ed. CRC-Press, Boca Raton, Florida, 33431 (1980).'
- (2) In Part B, Chapter B.5 is replaced by the following:

B.5 ACUTE EYE IRRITATION/CORROSION

INTRODUCTION

This test method is equivalent to OECD test guideline (TG) 405 (2012). OECD test guidelines for Testing of Chemicals are periodically reviewed to ensure that they reflect the best available science. In previous reviews of this test guideline, special attention was given to possible improvements through the evaluation of all existing information on the test chemical in order to avoid unnecessary testing in laboratory animals and thereby address animal welfare concerns. TG 405 (adopted in 1981 and updated in 1987, 2002, and 2012) includes the recommendation that prior to undertaking the described *in vivo* test for acute eye irritation/corrosion, a weight-of-the-evidence analysis should be performed (1) on the existing relevant data. Where insufficient data are available, it is recommended that they should be developed through application of sequential testing (2) (3). The testing strategy includes the performance of validated and accepted *in vitro* tests and is provided as a supplement to this test method. For the purpose of Regulation (EC) No 1907/2006 concerning the registration, evaluation, authorization and restriction of chemicals (REACH) ⁽¹⁾, an integrated testing strategy is also included in the relevant ECHA Guidance (21). Testing in animals should only be conducted if determined to be necessary after consideration of available alternative methods, and use of those determined to be appropriate. At the time of drafting of this updated test method, there are instances where using this test method is still necessary or required under some regulatory frameworks.

The latest update mainly focused on the use of analgesics and anesthetics without impacting the basic concept and structure of the test guideline. ICCVAM ⁽²⁾ and an independent international scientific peer review panel reviewed the usefulness and limitations of routinely using topical anesthetics, systemic analgesics, and humane endpoints during *in vivo* ocular irritation safety testing (12). The review concluded that the use of topical anesthetics and systemic analgesics could avoid most or all pain and distress without affecting the outcome of the test, and recommended that these substances should always be used. This test method takes this review into account. Topical anesthetics, systemic analgesics, and humane endpoints should be routinely used during acute eye irritation and corrosion *in vivo* testing. Exceptions to their use should be justified. The refinements described in this method will substantially reduce or avoid animal pain and distress in most testing situations where *in vivo* ocular safety testing is still necessary.

Balanced preemptive pain management should include (i) routine pretreatment with a topical anesthetic (e.g. proparacaine or tetracaine) and a systemic analgesic (e.g. buprenorphine), (ii) routine post-treatment schedule of systemic analgesia (e.g. buprenorphine and meloxicam), (iii) scheduled observation, monitoring, and recording of animals for clinical signs of pain and/or distress, and (iv) scheduled observation, monitoring, and recording of the nature, severity, and progression of all eye injuries. Further detail is provided in the updated procedures described below. Following test chemical administration, no additional topical anesthetics or analgesics should be applied in order to avoid interference with the study. Analgesics with anti-inflammatory activity (e.g. meloxicam) should not be applied topically, and doses used systemically should not interfere with ocular effects.

Definitions are set out in the Appendix to the test method.

⁽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).

⁽²⁾ The US Interagency Coordinating Committee on the Validation of Alternative Methods.

INITIAL CONSIDERATIONS

In the interest of both sound science and animal welfare, *in vivo* testing should not be considered until all available data relevant to the potential eye corrosivity/irritation of the chemical have been evaluated in a weight-of-the-evidence analysis. Such data include evidence from existing studies in humans and/or laboratory animals, evidence of eye corrosivity/irritation of one or more structurally related substances or mixtures of such substances, data demonstrating high acidity or alkalinity of the chemical (4) (5), and results from validated and accepted *in vitro* or *ex vivo* tests for skin corrosion and eye corrosion/irritation (6) (13) (14) (15) (16) (17). The studies may have been conducted prior to, or as a result of, a weight-of-the-evidence analysis.

For certain chemical, such an analysis may indicate the need for *in vivo* studies of the ocular corrosion/irritation potential of the chemical. In all such cases, before considering the use of the *in vivo* eye test, preferably a study of the *in vitro* and/or *in vivo* skin corrosion effects of the chemical should be conducted first and evaluated in accordance with the sequential testing strategy in test method B.4 (7) or the integrated testing strategy described in ECHA Guidance (21).

A sequential testing strategy, which includes the performance of validated *in vitro* or *ex vivo* eye corrosion/irritation tests, is included as a Supplement to this test method, and, for the purpose of REACH, in ECHA Guidance (21). It is recommended that such a testing strategy be followed prior to undertaking *in vivo* testing. For new chemicals, a stepwise testing approach is recommended for developing scientifically sound data on the corrosivity/irritation of the chemical. For existing chemicals with insufficient data on skin and eye corrosion/irritation, the strategy can be used to fill missing data gaps. The use of a different testing strategy or procedure or the decision not to use a stepwise testing approach, should be justified.

PRINCIPLE OF THE *IN VIVO* TEST

Following pretreatment with a systemic analgesic and induction of appropriate topical anesthesia, the chemical to be tested is applied in a single dose to one of the eyes of the experimental animal; the untreated eye serves as the control. The degree of eye irritation/corrosion is evaluated by scoring lesions of conjunctiva, cornea, and iris, at specific intervals. Other effects in the eye and adverse systemic effects are also described 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.

Animals showing signs of severe distress and/or pain at any stage of the test or lesions consistent with the humane endpoints described in this test method (see Paragraph 26) should be humanely killed, and the chemical assessed accordingly. Criteria for making the decision to humanely kill moribund and severely suffering animals are the subject of an OECD Guidance document (8).

PREPARATIONS FOR THE *IN VIVO* TEST

Selection of species

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

Preparation of animals

Both eyes of each experimental animal provisionally selected for testing should be examined within 24 hours before testing starts. Animals showing eye irritation, ocular defects, or pre-existing corneal injury should not be used.

Housing and feeding conditions

Animals should be individually housed. The temperature of the experimental animal room should be 20 °C (\pm 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. Excessive light intensity should be avoided. For feeding, conventional laboratory diets may be used with an unrestricted supply of drinking water.

TEST PROCEDURE

Use of topical anesthetics and systemic analgesics

The following procedures are recommended to avoid or minimize pain and distress in ocular safety testing procedures. Alternate procedures that have been determined to provide as good or better avoidance or relief of pain and distress may be substituted.

- Sixty minutes prior to test chemical application (TCA), buprenorphine 0,01 mg/kg is administered by subcutaneous injection (SC) to provide a therapeutic level of systemic analgesia. Buprenorphine and other similar opiod analgesics administered systemically are not known or expected to alter ocular responses (12).
- Five minutes prior to TCA, one or two drops of a topical ocular anesthetic (e.g. 0,5 % proparacaine hydrochloride or 0,5 % tetracaine hydrochloride) are applied to each eye. In order to avoid possible interference with the study, a topical anesthetic that does not contain preservatives is recommended. The eye of each animal that is not treated with a test chemical, but which is treated with topical anesthetics, serves as a control. If the test chemical is anticipated to cause significant pain and distress, it should not normally be tested *in vivo*. However, in case of doubt or where testing is necessary, consideration should be given to additional applications of the topical anesthetic at 5-minute intervals prior to TCA. Users should be aware that multiple applications of topical anesthetics could potentially cause a slight increase in the severity and/or time required for chemically-induced lesions to clear.
- Eight hours after TCA, buprenorphine 0,01 mg/kg SC and meloxicam 0,5 mg/kg SC are administered to provide a continued therapeutic level of systemic analgesia. While there are no data to suggest that meloxicam has anti-inflammatory effects on the eye when administered SC once daily, meloxicam should not be administered until at least 8 hours after TCA in order to avoid any possible interference with the study (12).
- After the initial 8-hour post-TCA treatment, buprenorphine 0,01 mg/kg SC should be administered every 12 hours, in conjunction with meloxicam 0,5 mg/kg SC every 24 hours, until the ocular lesions resolve and no clinical signs of pain and distress are present. Sustained-release preparations of analgesics are available that could be considered to decrease the frequency of analgesic dosing.
- “Rescue” analgesia should be given immediately after TCA if pre-emptive analgesia and topical anesthesia are inadequate. If an animal shows signs of pain and distress during the study, a “rescue” dose of buprenorphine 0,03 mg/kg SC would be given immediately and repeated as often as every 8 hours, if necessary, instead of 0,01 mg/kg SC every 12 hours. Meloxicam 0,5 mg/kg SC would be administered every 24 hours in conjunction with the “rescue” dose of buprenorphine, but not until at least 8 hours post-TCA.

Application of the test chemical

The test chemical should be placed in the conjunctival sac of one eye of each animal after gently pulling the lower lid away from the eyeball. The lids are then gently held together for about one second in order to prevent loss of the material. The other eye, which remains untreated, serves as a control.

Irrigation

The eyes of the test animals should not be washed for at least 24 hours following instillation of the test chemical, except for solids (see paragraph 18), and in case of immediate corrosive or irritating effects. At 24 hours a washout may be used if considered appropriate.

Use of a satellite group of animals to investigate the influence of washing is not recommended unless it is scientifically justified. If a satellite group is needed, two rabbits should be used. Conditions of washing should be carefully documented, e.g. time of washing; composition and temperature of wash solution; duration, volume, and velocity of application.

Dose level**(1) Testing of liquids**

For testing liquids, a dose of 0,1 ml is used. Pump sprays should not be used for instilling the chemical directly into the eye. The liquid spray should be expelled and collected in a container prior to instilling 0,1 mL into the eye.

(2) Testing of solids

When testing solids, pastes, and particulate chemicals, the amount used should have a volume of 0,1 ml or a weight of not more than 100 mg. The test chemical should be ground to a fine dust. The volume of solid material should be measured after gently compacting it, e.g. by tapping the measuring container. If the solid test chemical has not been removed from the eye of the test animal by physiological mechanisms at the first observation time point of 1 hour after treatment, the eye may be rinsed with saline or distilled water.

(3) Testing of aerosols

It is recommended that all pump sprays and aerosols be collected prior to instillation into the eye. The one exception is for chemicals in pressurised aerosol containers, which cannot be collected due to vaporisation. In such cases, the eye should be held open, and the test chemical administered to the eye in a simple burst of about one second, from a distance of 10 cm directly in front of the eye. This distance may vary depending on the pressure of the spray and its contents. Care should be taken not to damage the eye from the pressure of the spray. In appropriate cases, there may be a need to evaluate the potential for “mechanical” damage to the eye from the force of the spray.

An estimate of the dose from an aerosol can be made by simulating the test as follows: the chemical is sprayed on to weighing paper through an opening the size of a rabbit eye placed directly before the paper. The weight increase of the paper is used to approximate the amount sprayed into the eye. For volatile chemicals, the dose may be estimated by weighing a receiving container before and after removal of the test chemical.

Initial test (in vivo eye irritation/corrosion test using one animal)

It is strongly recommended that the *in vivo* test be performed initially using one animal (see Supplement to this test method: A Sequential Testing Strategy for Eye Irritation and Corrosion). Observations should allow for determination of severity and reversibility before proceeding to a confirmatory test in a second animal.

If the results of this test indicate the chemical to be corrosive or a severe irritant to the eye using the procedure described, further testing for ocular irritancy should not be performed.

Confirmatory test (in vivo eye irritation test with additional animals)

If a corrosive or severe irritant effect is not observed in the initial test, the irritant or negative response should be confirmed using up to two additional animals. If an irritant effect is observed in the initial test, it is recommended that the confirmatory test be conducted in a sequential manner in one animal at a time, rather than exposing the two additional animals simultaneously. If the second animal reveals corrosive or severe irritant effects, the test is not continued. If results from the second animal are sufficient to allow for a hazard classification determination, then no further testing should be conducted.

Observation period

The duration of the observation period should be sufficient to evaluate fully the magnitude and reversibility of the effects observed. However, the experiment should be terminated at any time that the animal shows signs of severe pain or distress (8). To determine reversibility of effects, the animals should be observed normally for 21 days post administration of the test chemical. If reversibility is seen before 21 days, the experiment should be terminated at that time.

Clinical observations and grading of eye reactions

The eyes should be comprehensively evaluated for the presence or absence of ocular lesions one hour post-TCA, followed by at least daily evaluations. Animals should be evaluated several times daily for the first 3 days to ensure that termination decisions are made in a timely manner. Test animals should be routinely evaluated for the entire duration of the study for clinical signs of pain and/or distress (e.g. repeated pawing or rubbing of the eye, excessive blinking, excessive tearing) (9) (10) (11) at least twice daily, with a minimum of 6 hours between observations, or more often if necessary. This is necessary to (i) adequately assess animals for evidence of pain and distress in order to make informed decisions on the need to increase the dosage of analgesics and (ii) assess animals for evidence of established humane endpoints in order to make informed decisions on whether it is appropriate to humanely euthanize animals, and to ensure that such decisions are made in a timely manner. Fluorescein staining should be routinely used and a slit lamp biomicroscope used when considered appropriate (e.g. assessing depth of injury when corneal ulceration is present) as an aid in the detection and measurement of ocular damage, and to evaluate if established endpoint criteria for humane euthanasia have been met. Digital photographs of observed lesions may be collected for reference and to provide a permanent record of the extent of ocular damage. Animals should be kept on test no longer than necessary once definitive information has been obtained. Animals showing severe pain or distress should be humanely killed without delay, and the chemical assessed accordingly.

Animals with the following eye lesions post-instillation should be humanely killed (refer to Table 1 for a description of lesion grades): corneal perforation or significant corneal ulceration including staphyloma; blood in the anterior chamber of the eye; grade 4 corneal opacity; absence of a light reflex (iridial response grade 2) which persists for 72 hours; ulceration of the conjunctival membrane; necrosis of the conjunctivae or nictitating membrane; or sloughing. This is because such lesions generally are not reversible. Furthermore, it is recommended that the following ocular lesions be used as humane endpoints to terminate studies before the end of the scheduled 21-day observation period. These lesions are considered predictive of severe irritant or corrosive injuries and injuries that are not expected to fully reverse by the end of the 21-day observation period: severe depth of injury (e.g. corneal ulceration extending beyond the superficial layers of the stroma), limbus destruction > 50 % (as evidenced by blanching of the conjunctival tissue), and severe eye infection (purulent discharge). A combination of: vascularisation of the cornea surface (i.e., pannus); area of fluorescein staining not diminishing over time based on daily assessment; and/or lack of re-epithelialisation 5 days after test chemical application could also be considered as potentially useful criteria to influence the clinical decision on early study termination. However, these findings individually are insufficient to justify early study termination. Once severe ocular effects have been identified, an attending or qualified laboratory animal veterinarian or personnel trained to identify the clinical lesions should be consulted for a clinical examination to determine if the combination of these effects warrants early study termination. The grades of ocular reaction (conjunctivae, cornea and iris) should be obtained and recorded at 1, 24, 48, and 72 hours following test chemical application (Table 1). Animals that do not develop ocular lesions may be terminated not earlier than 3 days post instillation. Animals with ocular lesions that are not severe should be observed until the lesions clear, or for 21 days, at which time the study is terminated. Observations should be performed and recorded at a minimum of 1 hour, 24 hours, 48 hours, 72 hours, 7 days, 14 days, and 21 days in order to determine the status of the lesions, and their reversibility or irreversibility. More frequent observations should be performed if necessary in order to determine whether the test animal should be euthanized out of humane considerations or removed from the study due to negative results

The grades of ocular lesions (Table 1) should be recorded at each examination. Any other lesions in the eye (e.g. pannus, staining, anterior chamber changes) or adverse systemic effects should also be reported.

Examination of reactions can be facilitated by use of a binocular loupe, hand slit-lamp, biomicroscope, or other suitable device. After recording the observations at 24 hours, the eyes may be further examined with the aid of fluorescein.

The grading of ocular responses is necessarily subjective. To promote harmonisation of grading of ocular 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.

DATA AND REPORTING

Evaluation of results

The ocular 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 chemical, as other effects of the test chemical are also evaluated. Instead, individual scores should be viewed as reference values and are only meaningful when supported by a full description and evaluation of all observations.

Test report

The test report should include the following information:

Rationale for in vivo testing: weight-of-the-evidence analysis of pre-existing test data, including results from sequential testing strategy:

- description of relevant data available from prior testing;
- data derived in each step of testing strategy;
- description of in vitro tests performed, including details of procedures, results obtained with test/reference chemicals;
- description of in vivo dermal irritation / corrosion study performed, including results obtained;
- weight-of-the-evidence analysis for performing in vivo study.

Test chemical:

- identification data (e.g. chemical name and if available CAS number, purity, known impurities, source, lot number);
- physical nature and physicochemical properties (e.g. pH, volatility, solubility, stability, reactivity with water);
- in case of a mixture, components should be identified including identification data of the constituent substances (e.g. chemical names and if available CAS numbers) and their concentrations;
- dose applied.

Vehicle:

- identification, concentration (where appropriate), volume used;
- justification for choice of vehicle.

Test animals:

- species/strain used, rationale for using animals other than albino rabbit;
- age of each animal at start of study;
- number of animals of each sex in test and control groups (if required);
- individual animal weights at start and conclusion of test;
- source, housing conditions, diet, etc.

Anaesthetics and analgesics

- doses and times when topical anaesthetics and systemic analgesics were administered;
- if local anaesthetic is used, identification, purity, type, and potential interaction with test chemical.

Results:

- description of method used to score irritation at each observation time (e.g. hand slitlamp, biomicroscope, fluorescein);
- tabulation of irritant/corrosive response data for each animal at each observation time up to removal of each animal from the test;
- narrative description of the degree and nature of irritation or corrosion observed;
- description of any other lesions observed in the eye (e.g. vascularisation, pannus formation, adhesions, staining);
- description of non-ocular local and systemic adverse effects, record of clinical signs of pain and distress, digital photographs, and histopathological findings, if any.

*Discussion of results***Interpretation of the results**

Extrapolation of the results of eye irritation studies in laboratory animals to humans is valid only to a limited degree. In many cases the albino rabbit is more sensitive than humans to ocular irritants or corrosives.

Care should be taken in the interpretation of data to exclude irritation resulting from secondary infection.

LITERATURE:

- (1) Barratt, M.D., *et al.* (1995), The Integrated Use of Alternative Approaches for Predicting Toxic Hazard, ECVAM Workshop Report 8, ATLA 23, 410 - 429.
- (2) de Silva, O., *et al.* (1997), Evaluation of Eye Irritation Potential: Statistical Analysis and Tier Testing Strategies, Food Chem. Toxicol 35, 159 - 164.
- (3) Worth A.P. and Fentem J.H. (1999), A general approach for evaluating stepwise testing strategies ATLA 27, 161-177.
- (4) Young, J.R., *et al.* (1988), Classification as Corrosive or Irritant to Skin of Preparations Containing Acidic or Alkaline Substance Without Testing on Animals, Toxicol. In Vitro, 2, 19 - 26.
- (5) Neun, D.J. (1993), Effects of Alkalinity on the Eye Irritation Potential of Solutions Prepared at a Single pH, J. Toxicol. Cut. Ocular Toxicol. 12, 227 - 231.
- (6) Fentem, J.H., *et al.* (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, pp.483 - 524.
- (7) Chapter B.4 of this Annex, *Acute Dermal Irritation/Corrosion*.
- (8) OECD (2000), Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. OECD Environmental Health and Safety Publications, Series on Testing and Assessment No. 19. (<http://www.oecd.org/ehs/test/monos.htm>).
- (9) Wright EM, Marcella KL, Woodson JF. (1985), Animal pain: evaluation and control, Lab Animal, May/June, 20-36.
- (10) National Research Council (NRC) (2008), Recognition and Alleviation of Distress in Laboratory Animals, Washington, DC: The National Academies Press.
- (11) National Research Council (NRC) (2009), Recognition and Alleviation of Pain in Laboratory Animals, Washington, DC: The National Academies Press.

- (12) ICCVAM (2010), ICCVAM Test Method Evaluation Report: Recommendations for Routine Use of Topical Anesthetics, Systemic Analgesics, and Humane Endpoints to Avoid or Minimize Pain and Distress in Ocular Safety Testing, NIH Publication No. 10-7514, Research Triangle Park, NC, USA: National Institute of Environmental Health Sciences.

<http://iccvam.niehs.nih.gov/methods/ocutox/OcuAnest-TMER.htm>

- (13) Chapter B.40 of this Annex, *In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test (TER)*.
- (14) Chapter B.40bis of this Annex, *In Vitro Skin Corrosion: Human Skin Model Test*.
- (15) OECD (2006), *Test No. 435: In vitro Membrane Barrier Test Method for Skin corrosion*, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Paris.
- (16) 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*.
- (17) 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 for Eye Irritation or Serious Eye Damage*.
- (18) U.S. EPA (2003), *Label Review Manual: 3rd Edition*, EPA737-B-96-001, Washington, DC: U.S., Environmental Protection Agency.
- (19) UN (2011), *Globally Harmonized System of Classification and Labelling of Chemicals (GHS)*, Fourth revised edition, New York & Geneva: United Nations Publications.
- (20) EC (2008), 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, 31.12.2008, p. 1).
- (21) ECHA Guidance on information requirements and chemical safety assessment, Chapter R.7a: Endpoint specific guidance.

http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf

Table 1

Grading of ocular lesions

Cornea	Grade
Opacity: degree of density (readings should be taken from most dense area) (*)	
No ulceration or opacity	0
Scattered or diffuse areas of opacity (other than slight dulling of normal lustre); details of iris clearly visible	1
Easily discernible translucent area; details of iris slightly obscured	2
Nacrous area; no details of iris visible; size of pupil barely discernible	3

Cornea	Grade
Opaque cornea; iris not discernible through the opacity	4
Maximum possible: 4	
Iris	
Normal	0
Markedly deepened rugae, congestion, swelling, moderate circumcorneal hyperaemia; or injection; iris reactive to light (a sluggish reaction is considered to be an effect)	1
Hemorrhage, gross destruction, or no reaction to light	2
Maximum possible: 2	
Conjunctivae	
Redness (refers to palpebral and bulbar conjunctivae; excluding cornea and iris)	
Normal	0
Some blood vessels hyperaemic (injected)	1
Diffuse, crimson colour; individual vessels not easily discernible	2
Diffuse beefy red	3
Maximum possible: 3	
Chemosis	
Swelling (refers to lids and/or nictating membranes)	
Normal	0
Some swelling above normal	1
Obvious swelling, with partial eversion of lids	2
Swelling, with lids about half closed	3
Swelling, with lids more than half closed	4
Maximum possible: 4	

(*) The area of corneal opacity should be noted.

Appendix

DEFINITIONS:

Acid/alkali reserve: For acidic preparations, this is the amount (g) of sodium hydroxide/100 g of preparation required to produce a specified pH. For alkaline preparations, it is the amount (g) of sodium hydroxide equivalent to the g sulphuric acid/100 g of preparation required to produce a specified pH (Young *et al.* 1988).

Chemical: A substance or a mixture.

Non irritants: Substances that are not classified as EPA Category I, II, or III ocular irritants; or GHS eye irritants Category 1, 2, 2A, or 2B; or EU Category 1 or 2 (17) (18) (19).

Ocular corrosive: (a) A chemical that causes irreversible tissue damage to the eye; (b) Chemicals that are classified as GHS eye irritants Category 1, or EPA Category I ocular irritants, or EU Category 1 (17) (18) (19).

Ocular irritant: (a) A chemical that produces a reversible change in the eye; (b) Chemicals that are classified as EPA Category II or III ocular irritants; or GHS eye irritants Category 2, 2A or 2B; or EU Category 2 (17) (18) (19).

Ocular severe irritant: (a) A chemical that causes tissue damage in the eye that does not resolve within 21 days of application or causes serious physical decay of vision; (b) Chemicals that are classified as GHS eye irritant Category 1, or EPA Category I ocular irritants, or EU Category 1 (17) (18) (19).

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

Tiered approach: 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.

Weight-of-the-evidence (process): The strengths and weaknesses of a collection of information are used as the basis for a conclusion that may not be evident from the individual data.

SUPPLEMENT TO TEST METHOD B.5 ⁽¹⁾**A SEQUENTIAL TESTING STRATEGY FOR EYE IRRITATION AND CORROSION****General considerations**

In the interests of sound science and animal welfare, it is important to avoid the unnecessary use of animals, and to minimise testing that is likely to produce severe responses in animals. All information on a chemical relevant to its potential ocular irritation/corrosivity should be evaluated prior to considering *in vivo* testing. Sufficient evidence may already exist to classify a test chemical as to its eye irritation or corrosion potential without the need to conduct testing in laboratory animals. Therefore, utilizing a weight-of-the-evidence analysis and sequential testing strategy will minimise the need for *in vivo* testing, especially if the chemical is likely to produce severe reactions.

It is recommended that a weight-of-the-evidence analysis be used to evaluate existing information pertaining to eye irritation and corrosion of chemicals and to determine whether additional studies, other than *in vivo* eye studies, should be performed to help characterise such potential. Where further studies are needed, it is recommended that the sequential testing strategy be utilised to develop the relevant experimental data. For substances which have no testing history, the sequential testing strategy should be utilised to develop the data needed to evaluate its eye corrosion/irritation. The initial testing strategy described in this Supplement was developed at an OECD workshop (1). It was subsequently affirmed and expanded in the Harmonised 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, in November 1998 (2), and updated by an OECD expert group in 2011.

Although this testing strategy is not an integrated part of test method B.5, it expresses the recommended approach for the determination of eye irritation/corrosion properties. This approach represents both best practice and an ethical benchmark for *in vivo* testing for eye irritation/corrosion. The test method provides guidance for the conduct of the *in vivo* test and summarises the factors that should be addressed before considering such a test. The sequential testing strategy provides a weight-of-the-evidence approach for the evaluation of existing data on the eye irritation/corrosion properties of chemicals and a tiered approach for the generation of relevant data on chemicals for which additional studies are needed or for which no studies have been performed. The strategy includes the performance first of validated and accepted *in vitro* or *ex vivo* tests and then of TM B.4 studies under specific circumstances (3) (4).

Description of the stepwise testing strategy

Prior to undertaking tests as part of the sequential testing strategy (Figure), all available information should be evaluated to determine the need for *in vivo* eye testing. Although significant information might be gained from the evaluation of single parameters (e.g. extreme pH), the totality of existing information should be assessed. All relevant data on the effects of the chemical in question, and its structural analogues, should be evaluated in making a weight-of-the-evidence decision, and a rationale for the decision should be presented. Primary emphasis should be placed upon existing human and animal data on the chemical, followed by the outcome of *in vitro* or *ex vivo* testing. *In vivo* studies of corrosive chemicals should be avoided whenever possible. The factors considered in the testing strategy include:

Evaluation of existing human and/or animal data and/or *in vitro* data from validated and internationally accepted methods (Step 1)

⁽¹⁾ For the use of an integrated testing strategy for eye irritation under the REACH see also the ECHA Guidance on information requirements and chemical safety assessment, Chapter R.7a: Endpoint specific guidance http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf

Existing human data, e.g. clinical and occupational studies, and case reports, and/or animal test data from ocular studies and/or *in vitro* data from validated and internationally accepted methods for eye irritation/corrosion should be considered first, because they provide information directly related to effects on the eyes. Thereafter, available data from human and/or animal studies investigating dermal corrosion/irritation, and/or *in vitro* studies from validated and internationally accepted methods for skin corrosion should be evaluated. Chemicals with known corrosivity or severe irritancy to the eye should not be instilled into the eyes of animals, nor should chemicals showing corrosive or severe irritant effects to the skin; such chemicals should be considered to be corrosive and/or irritating to the eyes as well. Chemicals with sufficient evidence of non-corrosivity and non-irritancy from previously performed ocular studies should also not be tested in *in vivo* eye studies.

Analysis of structure activity relationships (SAR) (Step 2)

The results of testing of structurally related chemicals should be considered, if available. When sufficient human and/or animal data are available on structurally related substances or mixtures of such substances to indicate their eye corrosion/irritancy potential, it can be presumed that the test chemical will produce the same responses. In those cases, the chemical may not need to be tested. Negative data from studies of structurally related substances or mixtures of such substances do not constitute sufficient evidence of non-corrosivity/non-irritancy of a chemical under the sequential testing strategy. Validated and accepted SAR approaches should be used to identify the corrosion and irritation potential for both dermal and ocular effects.

Physicochemical properties and chemical reactivity (Step 3)

Chemicals exhibiting pH extremes such as $\leq 2,0$ or $\geq 11,5$ may have strong local effects. If extreme pH is the basis for identifying a chemical as corrosive or irritant to the eye, then its acid/alkaline reserve (buffering capacity) may also be taken into consideration (5)(6)(7). If the buffering capacity suggests that a chemical may not be corrosive to the eye (i.e., chemicals with extreme pH and low acid/alkaline reserve), then further testing should be undertaken to confirm this, preferably by the use of a validated and accepted *in vitro* or *ex vivo* test (see paragraph 10).

Consideration of other existing information (Step 4)

All available information on systemic toxicity via the dermal route should be evaluated at this stage. The acute dermal toxicity of the test chemical should also be considered. If the test chemical has been shown to be highly toxic by the dermal route, it may not need to be tested in the eye. Although there is not necessarily a relationship between acute dermal toxicity and eye irritation/corrosion, it can be assumed that if an agent is highly toxic via the dermal route, it will also exhibit high toxicity when instilled into the eye. Such data may also be considered between Steps 2 and 3.

Assessment of dermal corrosivity of the chemical if also required for regulatory purposes (Step 5)

The skin corrosion and severe irritation potential should be evaluated first in accordance with test method B.4 (4) and the accompanying Supplement (8), including the use of validated and internationally accepted *in vitro* skin corrosion test methods (9) (10) (11). If the chemical is shown to produce corrosion or severe skin irritation, it may also be considered to be a corrosive or severely irritant to the eye. Thus, no further testing would be required. If the chemical is not corrosive or severely irritating to the skin, an *in vitro* or *ex vivo* eye test should be performed.

Results from *in vitro* or *ex vivo* tests (Step 6).

Chemicals that have demonstrated corrosive or severe irritant properties in an *in vitro* or *ex vivo* test (12) (13) that has been validated and internationally accepted for the assessment specifically of eye corrosivity/irritation, need not be tested in animals. It can be presumed that such chemicals will produce similar severe effects *in vivo*. If validated and accepted *in vitro/ex vivo* tests are not available, one should bypass Step 6 and proceed directly to Step 7.

In vivo test in rabbits (Steps 7 and 8)

In vivo ocular testing should begin with an initial test using one animal. If the results of this test indicate the chemical to be a severe irritant or corrosive to the eyes, further testing should not be performed. If that test does not reveal any corrosive or severe irritant effects, a confirmatory test is conducted with two additional animals. Depending upon the results of the confirmatory test, further tests may be needed. [see test method B.5]

TESTING AND EVALUATION STRATEGY FOR EYE IRRITATION/CORROSION

	Activity	Finding	Conclusion
1	<div>Existing human and/or animal data, and/or <i>in vitro</i> data from validated and internationally accepted methods showing effects on eyes</div> <div>Existing human and/or animal data and/or <i>in vitro</i> data from validated and internationally accepted methods showing corrosive effects on skin</div> <div>Existing human and/or animal data and/or <i>in vitro</i> data from validated and internationally accepted methods showing severe irritant effects on skin</div>	<div>Severe damage to eyes</div> <div>Eye irritant</div> <div>Not corrosive/not irritating to eyes</div> <div>Skin corrosive</div> <div>Severe skin irritant</div>	<div>Apical endpoint; consider corrosive to eyes. No testing is needed.</div> <div>Apical endpoint; consider irritating to eyes. No testing is needed.</div> <div>Apical endpoint; considered non-corrosive and non-irritating to eyes. No testing required.</div> <div>Assume corrosivity to eyes. No testing is needed.</div> <div>Assume irritating to eyes. No testing is needed</div>
	↓		
	<i>no information available, or available information is not conclusive</i>		
	↓		
2	<div>Perform SAR for eye corrosion/irritation</div> <div>Consider SAR for skin corrosion</div>	<div>Predict severe damage to eyes</div> <div>Predict irritation to eyes</div> <div>Predict skin corrosivity</div>	<div>Assume corrosivity to eyes. No testing is needed.</div> <div>Assume irritating to eyes. No testing is needed.</div> <div>Assume corrosivity to eyes. No testing is needed.</div>
	↓		
	<i>No predictions can be made, or predictions are not conclusive or negative</i>		
	↓		
3	Measure pH (buffering capacity, if relevant)	pH ≤ 2 or ≥ 11,5 (with high buffering capacity, if relevant)	Assume corrosivity to eyes. No testing is needed.
	↓		
	<i>2 < pH < 11,5, or pH ≤ 2,0 or ≥ 11,5 with low/no buffering capacity, if relevant</i>		
	↓		

	Activity	Finding	Conclusion
4	Consider existing systemic toxicity data via the dermal route	Highly toxic at concentrations that would be tested in the eye.	Chemical would be too toxic for testing. No testing is needed.
	↓		
	<i>Such information is not available, or chemical is not highly toxic</i>		
	↓		
5	Experimentally assess skin corrosion potential according to the testing strategy in chapter B.4 of this Annex if also required for regulatory purposes	Corrosive or severe irritant response	Assume corrosive to eyes. No further testing is needed.
	↓		
	<i>Chemical is not corrosive or severely irritating to skin</i>		
	↓		
6	Perform validated and accepted <i>in vitro</i> or <i>ex vivo</i> ocular test(s)	Corrosive or severe irritant response	Assume corrosive or severe irritant to eyes, provided the test performed can be used to identify corrosives/severe irritants and the chemical is within the applicability domain of the test. No further testing is needed.
		Irritant response	Assume irritant to eyes, provided the test(s) performed can be used to correctly identify corrosive, severe irritants, and irritants, and the chemical is within the applicability domain of the test(s). No further testing is needed.
		Non-irritant response	Assume non-irritant to eyes, provided the test (s) performed can be used to correctly identify non-irritants, correctly distinguish these from chemicals that are irritants, severe irritants, or ocular corrosives, and the chemical is within the applicability domain of the test. No further testing is needed.
	↓		
	<i>Validated and accepted in vitro or ex vivo ocular test(s) cannot be used to reach a conclusion</i>		
	↓		
7	Perform initial <i>in vivo</i> rabbit eye test using one animal	Severe damage to eyes	Consider corrosive to eyes. No further testing is needed.
	↓		
	<i>No severe damage, or no response</i>		
	↓		

	Activity	Finding	Conclusion
8	Perform confirmatory test using one or two additional animals	Corrosive or irritating	Consider corrosive or irritating to eyes. No further testing is needed
		Not corrosive or irritating	Consider non-irritating and non-corrosive to eyes. No further testing is needed.

LITERATURE:

- (1) OECD (1996) OECD Test Guidelines Programme: Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods. Held in Solna, Sweden, 22 - 24 January 1996 (<http://www.oecd.org/ehs/test/background.htm>).
- (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 (<http://www.oecd.org/ehs/Class/HCL6.htm>).
- (3) Worth, A.P. and Fentem J.H. (1999). A General Approach for Evaluating Stepwise Testing Strategies. ATLA 27, 161-177.
- (4) Chapter B.4 of this Annex, Acute Dermal Irritation/Corrosion.
- (5) Young, J.R., How, M.J., Walker, A.P., Worth W.M.H. (1988) Classification as Corrosive or Irritant to Skin of Preparations Containing Acidic or Alkaline Substance Without Testing on Animals. *Toxicol. In Vitro*, 2, 19 - 26.
- (6) Fentem, J.H., Archer, G.E.B., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Edsail, 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, pp.483 - 524.
- (7) Neun, D.J. (1993) Effects of Alkalinity on the Eye Irritation Potential of Solutions Prepared at a Single pH. *J. Toxicol. Cut. Ocular Toxicol.* 12, 227 - 231.
- (8) Supplement to Chapter B.4 of this Annex, A Sequential Testing Strategy for Skin Irritation and Corrosion.
- (9) Chapter B.40 of this Annex, *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance Test (TER).
- (10) Chapter B.40bis of this Annex, *In Vitro* Skin Corrosion: Human Skin Model Test.
- (11) OECD (2006), *Test No. 435: In vitro Membrane Barrier Test Method for Skin corrosion*, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Paris.
- (12) 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.
- (13) 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 for Eye Irritation or Serious Eye Damage.'

- (3) In Part B, Chapter B.10 is replaced by the following:

B.10 *In Vitro* Mammalian Chromosomal Aberration Test

INTRODUCTION

This test method is equivalent to OECD test guideline 473 (2016). It is part of a series of test methods on genetic toxicology. An OECD document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to these Test Guidelines has been developed (1).

The purpose of the *in vitro* chromosomal aberration test is to identify chemicals that cause structural chromosomal aberrations in cultured mammalian cells (2) (3) (4). Structural aberrations may be of two types, chromosome or chromatid. Polyploidy (including endoreduplication) could arise in chromosome aberration assays *in vitro*. While aneugens can induce polyploidy, polyploidy alone does not indicate aneugenic potential and can simply indicate cell cycle perturbation or cytotoxicity (5). This test is not designed to measure aneuploidy. An *in vitro* micronucleus test (6) would be recommended for the detection of aneuploidy.

The *in vitro* chromosomal aberration test may employ cultures of established cell lines or primary cell cultures of human or rodent origin. The cells used should be selected on the basis of growth ability in culture, stability of the karyotype (including chromosome number) and spontaneous frequency of chromosomal aberrations (7). At the present time, the available data do not allow firm recommendations to be made but suggest it is important, when evaluating chemical hazards to consider the p53 status, genetic (karyotype) stability, DNA repair capacity and origin (rodent *versus* human) of the cells chosen for testing. The users of this test method are thus encouraged to consider the influence of these and other cell characteristics on the performance of a cell line in detecting the induction of chromosomal aberrations, as knowledge evolves in this area.

Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

Tests conducted *in vitro* generally require the use of an exogenous source of metabolic activation unless the cells are metabolically competent with respect to the test chemicals. The exogenous metabolic activation system does not entirely mimic *in vivo* conditions. Care should be taken to avoid conditions that could lead to artifactual positive results, i.e. chromosome damage not caused by direct interaction between the test chemicals and chromosomes; such conditions include changes in pH or osmolality (8) (9) (10), interaction with the medium components (11) (12) or excessive levels of cytotoxicity (13) (14) (15) (16).

This test is used to detect chromosomal aberrations that may result from clastogenic events. The analysis of chromosomal aberration induction should be done using cells in metaphase. It is thus essential that cells should reach mitosis both in treated and in untreated cultures. For manufactured nanomaterials, specific adaptations of this test method may be needed but are not described in this test method.

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

Cell cultures of human or other mammalian origin are exposed to the test chemical both with and without an exogenous source of metabolic activation unless cells with an adequate metabolizing capability are used (see paragraph 13). At appropriate predetermined intervals after the start of exposure of cell cultures to the test chemical, they are treated with a metaphase-arresting chemical (e.g. colcemid or colchicine), harvested, stained and metaphase cells are analysed microscopically for the presence of chromatid-type and chromosome-type aberrations.

DESCRIPTION OF THE METHOD

Preparations

Cells

A variety of cell lines (e.g. Chinese Hamster Ovary (CHO), Chinese Hamster lung V79, Chinese Hamster Lung (CHL)/IU, TK6) or primary cell cultures, including human or other mammalian peripheral blood lymphocytes, can be used (7). The choice of the cell lines used should be scientifically justified. When primary cells are used, for animal welfare reasons, the use of primary cells from human origin should be considered where feasible and sampled in accordance with the human ethical principles and regulations. Human peripheral blood lymphocytes should be obtained from young (approximately 18-35 years of age), non-smoking individuals with no known illness or recent exposures to genotoxic agents (e.g. chemicals, ionizing radiations) at levels that would increase the background incidence of chromosomal aberrations. This would ensure the background incidence of chromosomal aberrations to be low and consistent. The baseline incidence of chromosomal aberrations increases with age and this trend is more marked in females than in males (17) (18). If cells from more than one donor are pooled for use, the number of donors should be specified. It is necessary to demonstrate that the cells have divided from the beginning of treatment with the test chemical to cell sampling. Cell cultures are maintained in an exponential cell growth phase (cell lines) or stimulated to divide (primary cultures of lymphocytes), to expose the cells at different stages of the cell cycle, since the sensitivity of cell stages to the test chemicals may not be known. The primary cells that need to be stimulated with mitogenic agents in order to divide are generally no longer synchronized during exposure to the test chemical (e.g. human lymphocytes after a 48-hour mitogenic stimulation). The use of synchronized cells during treatment is not recommended, but can be acceptable if justified.

Media and culture conditions

Appropriate culture medium and incubation conditions (culture vessels, humidified atmosphere of 5 % CO₂ if appropriate, incubation temperature of 37 °C) should be used for maintaining cultures. Cell lines should be checked routinely for the stability of the modal chromosome number and the absence of *Mycoplasma* contamination (7) (19), and cells should not be used if contaminated or if the modal chromosome number has changed. The normal cell cycle time of cell lines or primary cultures used in the testing laboratory should be established and should be consistent with the published cell characteristics (20).

Preparation of cultures

Cell lines: cells 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 until harvest time (e.g. confluence should be avoided for cells growing in monolayers).

Lymphocytes: whole blood treated with an anti-coagulant (e.g. heparin) or separated lymphocytes are cultured (e.g. for 48 hours for human lymphocytes) in the presence of a mitogen [e.g. phytohaemagglutinin (PHA) for human lymphocytes] in order to induce cell division prior to exposure to the test chemical.

Metabolic activation

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 (21) (22) (23) or a combination of phenobarbital and β -naphthoflavone (24) (25) (26) (27) (28) (29). The latter combination does not conflict with the Stockholm Convention on Persistent Organic Pollutants (30) and has been shown to be as effective as Aroclor 1254 for inducing mixed-function oxidases (24) (25) (26) (28). 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 use of products that reduce the mitotic index, especially calcium complexing products (31) should be avoided during treatment. The choice of type and concentration of exogenous metabolic activation system or metabolic inducer employed may be influenced by the class of chemicals being tested.

Test chemical preparation

Solid test chemicals should be prepared in appropriate solvents and diluted, if appropriate, prior to treatment of the cells (see paragraph 23). 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 (32) (33) (34). Preparations of the test chemical should be made just prior to treatment unless stability data demonstrate the acceptability of storage.

Test conditions

Solvents

The solvent should be chosen to optimize the solubility of the test chemicals without adversely impacting the conduct of the assay, 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 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 not 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 include untreated controls (see Appendix 1) to demonstrate that no deleterious or clastogenic effects are induced by the chosen solvent.

Measuring cell proliferation and cytotoxicity and choosing treatment concentrations

When determining the highest test chemical concentration, concentrations that have the capability of producing artifactual positive responses, such as those producing excessive cytotoxicity (see paragraph 22), precipitation in the culture medium (see paragraph 23), 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 artifactual positive results and to maintain appropriate culture conditions.

Measurements of cell proliferation are made to assure that a sufficient number of treated cells have reached mitosis during the test and that the treatments are conducted at appropriate levels of cytotoxicity (see paragraphs 18 and 22). Cytotoxicity should be determined with and without metabolic activation in the main experiment using an appropriate indication of cell death and growth. 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 mandatory. If performed, it should not replace the measurement of cytotoxicity in the main experiment.

Relative Population Doubling (RPD) or Relative Increase in Cell Count (RICC) are appropriate methods for the assessment of cytotoxicity in cytogenetic tests (13) (15) (35) (36) (55) (see Appendix 2 for formulas). In case of long-term treatment and sampling times after the beginning of treatment longer than 1,5 normal cell cycle lengths (i.e. longer than 3 cell cycle lengths in total), RPD might underestimate cytotoxicity (37). Under these circumstances RICC might be a better measure or the evaluation of cytotoxicity after 1,5 normal cell cycle lengths would be a helpful estimate using RPD.

For lymphocytes in primary cultures, while the mitotic index (MI) is a measure of cytotoxic/cytostatic effects, it is influenced by the time after treatment it is measured, the mitogen used and possible cell cycle disruption. However, the MI is acceptable because other cytotoxicity measurements may be cumbersome and impractical and may not apply to the target population of lymphocytes growing in response to PHA stimulation.

While RICC and RPD for cell lines and MI for primary culture of lymphocytes are the recommended cytotoxicity parameters, other indicators (e.g. cell integrity, apoptosis, necrosis, cell cycle) could provide useful additional information.

At least three test concentrations (not including the solvent and positive controls) that meet the acceptability criteria (appropriate cytotoxicity, number of cells, etc) should be evaluated. Whatever the types of cells (cell lines or primary cultures of lymphocytes), either replicate or single treated cultures may be used at each concentration tested. While the use of duplicate cultures is advisable, single cultures are also acceptable provided that the same total number of cells are scored for either single or duplicate cultures. The use of single cultures is particularly relevant when more than 3 concentrations are assessed (see paragraph 31). The results obtained in the independent replicate cultures at a given concentration can be pooled for the data analysis (38). 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 as described in paragraph 22 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 obtain data at low and moderate cytotoxicity or to study the dose response relationship in detail, it will be necessary to use more closely spaced concentrations and/or more than three concentrations (single cultures or replicates), in particular in situations where a repeat experiment is required (see paragraph 47).

If the maximum concentration is based on cytotoxicity, the highest concentration should aim to achieve $55 \pm 5\%$ cytotoxicity using the recommended cytotoxicity parameters (i.e. reduction in RICC and RPD for cell lines and reduction in MI for primary cultures of lymphocytes to $45 \pm 5\%$ of the concurrent negative control). Care should be taken in interpreting positive results only to be found in the higher end of this $55 \pm 5\%$ cytotoxicity range (13).

For poorly soluble test chemicals that are not cytotoxic at concentrations lower than 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. At the concentration producing a precipitate, care should be taken to assure that the precipitate does not interfere with the conduct of the test (e.g. staining or scoring). The determination of solubility in the culture medium prior to the experiment may be useful.

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) (41). When the test chemical is not of defined composition, e.g. a substance of unknown or variable composition, complex reaction products or biological material (UVCB) (42), environmental extract 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 (43).

Controls

Concurrent negative controls (see paragraph 15), consisting of solvent alone in the treatment medium and treated in the same way as the treatment cultures, should be included for every harvest time.

Concurrent positive controls are needed to demonstrate the ability of the laboratory to identify clastogens 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 the table 1 below. Alternative positive control chemicals can be used, if justified. Because *in vitro* mammalian cell tests for genetic toxicity are sufficiently standardized, the use of positive controls may be confined to a clastogen requiring metabolic activation. Provided it is done concurrently with the non-activated test using the same treatment duration, this single positive control response will demonstrate both the activity of the metabolic activation system and the responsiveness of the test system. Long term treatment (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 (i.e. the effects are clear but do not immediately reveal the identity of the coded slides to the reader), and the response should not be compromised by cytotoxicity exceeding the limits specified in the test method.

Table 1

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

Category	Chemical	CASRN
1. Clastogens active without metabolic activation		
	Methyl methanesulphonate	66-27-3
	Mitomycin C	50-07-7
	4-Nitroquinoline-N-Oxide	56-57-5
	Cytosine arabinoside	147-94-4
2. Clastogens requiring metabolic activation		
	Benzo(a)pyrene	50-32-8
	Cyclophosphamide	50-18-0

PROCEDURE

Treatment with test chemical

Proliferating cells are treated with the test chemical in the presence and absence of a metabolic activation system.

Culture harvest time

For thorough evaluation, which would be needed to conclude a negative outcome, all three of the following experimental conditions should be conducted using a short term treatment with and without metabolic activation and long term treatment without metabolic activation (see paragraphs 43, 44 and 45):

- Cells should be exposed to the test chemical without metabolic activation for 3-6 hours, and sampled at a time equivalent to about 1,5 normal cell cycle lengths after the beginning of treatment (18),
- Cells should be exposed to the test chemical with metabolic activation for 3-6 hours, and sampled at a time equivalent to about 1,5 normal cell cycle lengths after the beginning of treatment (18),
- Cells should be continuously exposed without metabolic activation until sampling at a time equivalent to about 1,5 normal cell cycle lengths. Certain chemicals (e.g. nucleoside analogues) may be more readily detected by treatment/sampling times longer than 1,5 normal cell cycle lengths (24).

In the event that any of the above experimental conditions lead to a positive response, it may not be necessary to investigate any of the other treatment regimens.

Chromosome preparation

Cell cultures are treated with colcemid or colchicine usually for one to three hours prior to harvesting. Each cell culture is harvested and processed separately for the preparation of chromosomes. Chromosome preparation involves hypotonic treatment of the cells, fixation and staining. In monolayers, mitotic cells (identifiable as being round and detaching from the surface) may be present at the end of the 3-6 hour treatment. Because these mitotic cells are easily detached, they can be lost when the medium containing the test chemical is removed. If there is evidence for a substantial increase in the number of mitotic cells compared with controls, indicating likely mitotic arrest, then the cells should be collected by centrifugation and added back to cultures, to avoid losing cells that are in mitosis, and at risk for chromosome aberration, at the time of harvest.

Analysis

All slides, including those of the positive and negative controls, should be independently coded before microscopic analysis for chromosomal aberrations. Since fixation procedures often result in a proportion of metaphase cells which have lost chromosomes, the cells scored should, therefore, contain a number of centromeres equal to the modal number ± 2 .

At least 300 well-spread metaphases should be scored per concentration and control to conclude a test chemical as clearly negative (see paragraph 45). The 300 cells should be equally divided among the replicates, when replicate cultures are used. When single cultures are used per concentration (see paragraph 21), at least 300 well spread metaphases should be scored in this single culture. Scoring 300 cells has the advantage of increasing the statistical power of the test and in addition, zero values will be rarely observed (expected to be only 5 %) (44). The number of metaphases scored can be reduced when high numbers of cells with chromosome aberrations are observed and the test chemical considered as clearly positive.

Cells with structural chromosomal aberration(s) including and excluding gaps should be scored. Breaks and gaps are defined in Appendix 1 according to (45) (46). Chromatid- and chromosome-type aberrations should be recorded separately and classified by sub-types (breaks, exchanges). Procedures in use in the laboratory should ensure that analysis of chromosomal aberrations is performed by well-trained scorers and peer-reviewed if appropriate.

Although the purpose of the test is to detect structural chromosomal aberrations, it is important to record polyploidy and endoreduplication frequencies when these events are seen. (See paragraph 2).

Proficiency of the laboratory

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 chemicals acting via different mechanisms and various negative controls (using various solvents/vehicle). 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 paragraph 37.

A selection of positive control chemicals (see Table 1 in paragraph 26) should be investigated with short and 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 clastogenic chemicals and determine the effectiveness of the metabolic activation system. A range of concentrations of the selected chemicals 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

The laboratory should establish:

- A historical positive control range and distribution,
- A historical negative (untreated, solvent) control range and distribution.

When first acquiring data for an historical negative control distribution, concurrent negative controls should be consistent with published control data, where they exist. As more experimental data are added to the control distribution, concurrent negative controls should ideally be within the 95 % control limits of that distribution (44) (47). 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 (48)), to identify how variable their positive and negative control data are, and to show that the methodology is 'under control' in their laboratory (44). 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 (47).

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.

Negative control data should consist of the incidence of cells with chromosome aberrations from a single culture or the sum of replicate cultures as described in paragraph 21. Concurrent negative controls should ideally be within the 95 % control limits of the distribution of the laboratory's historical negative control database (44) (47). Where concurrent negative control data fall outside the 95 % control limits they may be acceptable for inclusion in the historical control distribution as long as these data are not extreme outliers and there is evidence that the test system is 'under control' (see paragraph 37) and evidence of absence of technical or human failure.

DATA AND REPORTING

Presentation of the results

The percentage of cells with structural chromosomal aberration(s) 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 cultures. Gaps are recorded and reported separately but not included in the total aberration frequency. Percentage of polyploidy and/or endoreduplicated cells are reported when seen.

Concurrent measures of cytotoxicity for all treated, negative and positive control cultures in the main aberration experiment(s) should be recorded.

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

Acceptability Criteria

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 39.
- Concurrent positive controls (see paragraph 26) 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.
- Cell proliferation criteria in the solvent control should be fulfilled (paragraphs 17 and 18).
- All three experimental conditions were tested unless one resulted in positive results (see paragraph 28).
- Adequate number of cells and concentrations are analysable (paragraphs 31 and 21).
- The criteria for the selection of top concentration are consistent with those described in paragraphs 22, 23 and 24.

Evaluation and interpretation of results

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 28):

- (a) at least one of the test concentrations exhibits a statistically significant increase compared with the concurrent negative control,
- (b) the increase is dose-related when evaluated with an appropriate trend test,
- (c) any of the results are outside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits; see paragraph 39).

When all of these criteria are met, the test chemical is then considered able to induce chromosomal aberrations in cultured mammalian cells in this test system. Recommendations for the most appropriate statistical methods can be found in the literature (49) (50) (51).

Providing that all acceptability criteria are fulfilled, a test chemical is considered clearly negative if, in all experimental conditions examined (see paragraph 28):

- (a) none of the test concentrations exhibits a statistically significant increase compared with the concurrent negative control,

- (b) there is no concentration-related increase when evaluated with an appropriate trend test,
- (c) all results are inside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits; see paragraph 39).

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

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

In case 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. Scoring additional cells (where appropriate) or 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.

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

An increase in the number of polyploid cells may indicate that the test chemicals have the potential to inhibit mitotic processes and to induce numerical chromosomal aberrations (52). An increase in the number of cells with endoreduplicated chromosomes may indicate that the test chemicals have the potential to inhibit cell cycle progress (53) (54) (see paragraph 2). Therefore, incidence of polyploid cells and cells with endoreduplicated chromosomes should be recorded separately.

Test report

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 should also be indicated.

Cells:

- type and source of cells
- karyotype features and suitability of the cell type used;
- absence of mycoplasma, for cell lines;
- for cell lines, information on cell cycle length, doubling time or proliferation index;
- sex of blood donors, age and any relevant information on the donor, whole blood or separated lymphocytes, mitogen used;
- number of passages, if available, for cell lines;
- methods for maintenance of cell cultures, for cell lines;
- modal number of chromosomes, for cell lines.

Test conditions:

- identity of the metaphase-arresting chemical, its concentration and duration of cell exposure;
- concentration of test chemical expressed as final concentration in the culture medium (e.g. µg or mg/mL or mM of culture medium).
- rationale for selection of concentrations and number of cultures including, e.g. cytotoxicity data and solubility limitations;
- composition of media, CO₂ concentration if applicable, humidity level;
- concentration (and/or volume) of solvent and test chemical added in the culture medium;
- incubation temperature;
- incubation time;
- duration of treatment;
- harvest time after treatment;
- cell density at seeding, if appropriate;
- 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 chemicals, final concentrations for each conditions of treatment;
- methods of slide preparation and staining technique used;
- criteria for acceptability of assays;
- criteria for scoring aberrations;
- number of metaphases analysed;
- methods for the measurements of cytotoxicity;
- any supplementary information relevant to cytotoxicity and method used;
- criteria for considering studies as positive, negative or equivocal;
- methods used to determine pH, osmolality and precipitation.

Results:

- the number of cells treated and the number of cells harvested for each culture when cell lines are used
- cytotoxicity measurements, e.g. RPD, RICC, MI, other observations if any;
- information on cell cycle length, doubling time or proliferation index in case of cell lines;
- signs of precipitation and time of the determination;
- definition for aberrations, including gaps;
- Number of cells scored, number of cells with chromosomal aberrations and type of chromosomal aberrations given separately for each treated and control culture, including and excluding gaps;
- changes in ploidy (polyploid cells and cells with endoreduplicated chromosomes, given separately) if seen;
- 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 95 % control limits for the distribution, as well as the number of data;
- statistical analyses, p-values if any.

*Discussion of the results.**Conclusions.*

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) Evans, H.J. (1976), "Cytological Methods for Detecting Chemical Mutagens", in *Chemical Mutagens, Principles and Methods for their Detection*, Vol. 4, Hollaender, A. (ed.), Plenum Press, New York and London, pp. 1-29
- (3) Ishidate, M. Jr., T. Sofuni (1985), "The *in vitro* Chromosomal Aberration Test Using Chinese Hamster Lung (CHL) Fibroblast Cells in Culture" in *Progress in Mutation Research*, Vol. 5, Ashby, J. *et al.* (eds.), Elsevier Science Publishers, Amsterdam-New York- Oxford, pp. 427-432.
- (4) Galloway, S.M. *et al.* (1987), Chromosomal aberration and sister chromatid exchanges in Chinese hamster ovary cells: Evaluation of 108 chemicals, *Environmental and Molecular Mutagenesis*, Vol. 10/suppl. 10, pp. 1-175.
- (5) Muehlbauer, P.A. *et al.* (2008), "Improving dose selection and identification of aneugens in the *in vitro* chromosome aberration test by integration of flow cytometry-based methods", *Environmental and Molecular Mutagenesis*, Vol. 49/4, pp. 318-327.
- (6) Chapter B.49 of this Annex: *In Vitro Mammalian Cell Micronucleus Test*.
- (7) ILSI paper (draft), Lorge, E., M. Moore, J. Clements, M. O Donovan, F. Darroudi, M. Honma, A. Czich, J van Benthem, S. Galloway, V. Thybaud, B. Gollapudi, M. Aardema, J. Kim, D.J. Kirkland, Recommendations for good cell culture practices in genotoxicity testing.
- (8) Scott, D. *et al.* (1991), Genotoxicity under Extreme Culture Conditions. A report from ICPEMC Task Group 9, *Mutation Research/Reviews in Genetic Toxicology*, Vol.257/2, pp. 147-204.

- (9) Morita, T. *et al.* (1992), Clastogenicity of Low pH to Various Cultured Mammalian Cells, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 268/2, pp. 297-305.
- (10) Brusick, D. (1986), Genotoxic effects in cultured mammalian cells produced by low pH treatment conditions and increased ion concentrations, *Environmental and Molecular Mutagenesis*, Vol. 8/6, pp. 789-886.
- (11) Long, L.H. *et al.* (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 Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 634/1-2, pp. 177-183.
- (12) Nessler, F. *et al.* (2008), Characterization of the Genotoxicity of Nitrilotriacetic Acid, *Environmental and Molecular Mutagenesis*, Vol. 49/6, pp. 439-452.
- (13) Galloway, S. (2000), Cytotoxicity and chromosome aberrations *in vitro*: Experience in industry and the case for an upper limit on toxicity in the aberration assay, *Environmental and Molecular Mutagenesis*, Vol. 35/3, pp. 191-201.
- (14) Kirkland, D. *et al.* (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 Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 584/1-2, pp. 1-256.
- (15) Greenwood, S. *et al.* (2004), Population doubling: a simple and more accurate estimation of cell growth suppression in the *in vitro* assay for chromosomal aberrations that reduces irrelevant positive results, *Environmental and Molecular Mutagenesis*, Vol. 43/1, pp. 36-44.
- (16) Hilliard, C.A. *et al.* (1998), Chromosome aberrations *in vitro* related to cytotoxicity of nonmutagenic chemicals and metabolic poisons, *Environmental and Molecular Mutagenesis*, Vol. 31/4, pp. 316-326.
- (17) Hedner K. *et al.* (1982), Sister chromatid exchanges and structural chromosomal aberrations in relation to age and sex, *Human Genetics*, Vol. 62, pp. 305-309.
- (18) Ramsey M.J. *et al.* (1995), The effects of age and lifestyle factors on the accumulation of cytogenetic damage as measured by chromosome painting, *Mutation Research*, Vol. 338, pp. 95-106.
- (19) Coecke S. *et al.* (2005), Guidance on Good Cell Culture Practice. A Report of the Second ECVAM Task Force on Good Cell Culture Practice, *ATLA*, Vol. 33/3, pp. 261-287.
- (20) Henderson, L. *et al.* (1997), Industrial Genotoxicology Group collaborative trial to investigate cell cycle parameters in human lymphocyte cytogenetics studies, *Mutagenesis*, Vol.12/3, pp.163-167.
- (21) Ames, B.N., J. McCann, E. Yamasaki (1975), Methods for Detecting Carcinogens and Mutagens with the Salmonella/Mammalian Microsome Mutagenicity Test, *Mutation Research/Environmental Mutagenesis and Related Subjects*, Vol. 31/6, pp. 347-363.
- (22) Maron, D.M., B.N. Ames (1983), Revised Methods for the Salmonella Mutagenicity Test, *Mutation Research/Environmental Mutagenesis and Related Subjects*, Vol. 113/3-4, pp. 173-215.
- (23) Natarajan, A.T. *et al.* (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 Research*, Vol. 37/1, pp. 83-90.
- (24) Matsuoka, A., M. Hayashi, M. Jr. Ishidate (1979), Chromosomal Aberration Tests on 29 Chemicals Combined with S9 Mix *in vitro*, *Mutation Research/Genetic Toxicology*, Vol. 66/3, pp. 277-290.
- (25) 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, *Journal of Environmental Pathology and Toxicology*, Vol. 4/1, pp. 55-65.

- (26) Elliot, B.M. *et al.* (1992), Report of UK Environmental Mutagen Society Working Party. Alternatives to Aroclor 1254-induced S9 in *in vitro* Genotoxicity Assays, *Mutagenesis*, Vol. 7/3, pp. 175-177.
- (27) Matsushima, T. *et al.* (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.
- (28) Galloway, S.M. *et al.* (1994). Report from Working Group on *in vitro* Tests for Chromosomal Aberrations, *Mutation Research/Environmental Mutagenesis and Related Subjects*, Vol. 312/3, pp. 241-261.
- (29) Johnson, T.E., D.R. Umbenhauer, S.M. Galloway (1996), Human liver S-9 metabolic activation: proficiency in cytogenetic assays and comparison with phenobarbital/beta-naphthoflavone or Aroclor 1254 induced rat S-9, *Environmental and Molecular Mutagenesis*, Vol. 28/1, pp. 51-59.
- (30) UNEP (2001), Stockholm Convention on Persistent Organic Pollutants, United Nations Environment Programme (UNEP). Available at: <http://www.pops.int/>.
- (31) Tucker, J.D., M.L. Christensen (1987), Effects of anticoagulants upon sister-chromatid exchanges, cell-cycle kinetics, and mitotic index in human peripheral lymphocytes, *Mutation Research*, Vol. 190/3, pp. 225-8.
- (32) Krahn, D.F., F.C. Barsky, K.T. McCooey (1982), "CHO/HGPRT Mutation Assay: Evaluation of Gases and Volatile Liquids", in *Genotoxic Effects of Airborne Agents*, Tice, R.R., D.L. Costa, K.M. Schaich (eds.), Plenum, New York, pp. 91-103.
- (33) Zamora, P.O. *et al.* (1983), Evaluation of an Exposure System Using Cells Grown on Collagen Gels for Detecting Highly Volatile Mutagens in the CHO/HGPRT Mutation Assay, *Environmental and Molecular Mutagenesis*, Vol. 5/6, pp. 795-801.
- (34) Asakura, M. *et al.* (2008), An improved system for exposure of cultured mammalian cells to gaseous compounds in the chromosomal aberration assay, *Mutation Research*, Vol. 652/2, pp. 122-130.
- (35) Lorge, E. *et al.* (2008), Comparison of different methods for an accurate assessment of cytotoxicity in the *in vitro* micronucleus test. I. Theoretical aspects, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 655/1-2, pp. 1-3.
- (36) Galloway, S. *et al.* (2011), Workshop summary: Top concentration for *in vitro* mammalian cell genotoxicity assays; and Report from working group on toxicity measures and top concentration for *in vitro* cytogenetics assays (chromosome aberrations and micronucleus), *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 723/2, pp. 77-83.
- (37) Honma, M. (2011), Cytotoxicity measurement in *in vitro* chromosome aberration test and micronucleus test, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, Vol. 724/1-2, pp. 86-87.
- (38) Richardson, C. *et al.* (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.
- (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) ENV/JM/TG(2014)17. Available upon request.
- (40) Morita, T., M. Honma, K. Morikawa (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 Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 741/1-2, pp. 32-56.
- (41) Brookmire, L., J.J. Chen, D.D. Levy (2013), Evaluation of the Highest Concentrations Used in the In Vitro Chromosome Aberrations Assay, *Environmental and Molecular Muagenesis*, Vol. 54/1, pp. 36-43.

- (42) EPA, Office of Chemical Safety and Pollution Prevention (2011), Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials: UVCB Substances, <http://www.epa.gov/opptintr/newchemicals/pubs/uvcb.txt>.
 - (43) 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>.
 - (44) OECD (2014), "Statistical analysis supporting the revision of the genotoxicity Test Guidelines, OECD Environment, Health and Safety Publications (EHS)", Series on Testing and Assessment, No. 198, OECD Publishing, Paris.
 - (45) ISCN (2013), *An International System for Human Cytogenetic Nomenclature*, Schaffer, L.G., J. MacGowan-Gordon, M. Schmid (eds.), Karger Publishers Inc., Connecticut.
 - (46) Scott, D. *et al.* (1990), "Metaphase chromosome aberration assays *in vitro*", in *Basic Mutagenicity Tests: UKEMS Recommended Procedures*, Kirkland, D.J. (ed.), Cambridge University Press, Cambridge, pp. 62-86.
 - (47) Hayashi, M. *et al.* (2011), Compilation and use of genetic toxicity historical control Data, *Mutation Research*, Vol. 723/2, pp. 87-90.
 - (48) Ryan, T. P. (2000), *Statistical Methods for Quality Improvement*, 2nd Edition, John Wiley and Sons, New York.
 - (49) Fleiss, J. L., B. Levin, M.C. Paik (2003), *Statistical Methods for Rates and Proportions*, 3rd ed., John Wiley & Sons, New York.
 - (50) Galloway, S.M. *et al.* (1987), Chromosome aberration and sister chromatid exchanges in Chinese hamster ovary cells: Evaluation of 108 chemicals, *Environmental and Molecular Mutagenesis*, Vol. 10/suppl. 10, pp. 1-175.
 - (51) Richardson, C. *et al.* (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.
 - (52) Warr, T.J., E.M. Parry, J.M. Parry (1993), A comparison of two *in vitro* mammalian cell cytogenetic assays for the detection of mitotic aneuploidy using 10 known or suspected aneugens, *Mutation Research*, Vol. 287/1, pp. 29-46.
 - (53) Locke-Huhle, C. (1983), Endoreduplication in Chinese hamster cells during alpha-radiation induced G2 arrest, *Mutation Research*, Vol. 119/3, pp. 403-413.
 - (54) Huang, Y., C. Change, J.E. Trosko (1983), Aphidicolin — induced endoreduplication in Chinese hamster cells, *Cancer Research*, Vol. 43/3, pp. 1362-1364.
 - (55) Soper, K.A., S.M. Galloway (1994), Cytotoxicity measurement in *in vitro* chromosome aberration test and micronucleus test, *Mutation Research*, Vol. 312, pp. 139-149.
-

Appendix 1

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).

Apoptosis: programmed cell death characterised by a series of steps leading to a disintegration of cells into membrane-bound particles that are then eliminated by phagocytosis or by shedding.

Cell proliferation: increase in cell number as a result of mitotic cell division.

Chemical: a substance or a mixture.

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

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

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

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

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

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

Cytotoxicity: For the assays covered in this test method using cell lines, cytotoxicity is identified as a reduction in relative population doubling (RPD) or relative increase in cell count (RICC) of the treated cells as compared to the negative control (see paragraph 17 and Appendix 2). For the assays covered in this test method using primary cultures of lymphocytes, cytotoxicity is identified as a reduction in mitotic index (MI) of the treated cells as compared to the negative control (see paragraph 18 and Appendix 2).

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

Genotoxic: a general term encompassing all types of DNA or chromosome damage, including breaks, deletions, adducts, nucleotides modifications and linkages, rearrangements, gene 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 sequences(s) in genes or of the structure of chromosomes (chromosome aberrations).

Numerical aberration: a change in the number of chromosomes from the normal number characteristic of the cells utilised.

Polyploidy: numerical chromosomal aberrations in cells or organisms involving entire set(s) of chromosomes, as opposed to an individual chromosome or chromosomes (aneuploidy).

p53 status: p53 protein is involved in cell cycle regulation, apoptosis and DNA repair. Cells deficient in functional p53 protein, unable to arrest cell cycle or to eliminate damaged cells via apoptosis or other mechanisms (e.g. induction of DNA repair) related to p53 functions in response to DNA damage, should be theoretically more prone to gene mutations or chromosomal aberrations.

Relative Increase in Cell Counts (RICC): the increase in the number of cells in chemically-exposed cultures versus increase in non-treated cultures, a ratio expressed as a percentage.

Relative Population Doubling (RPD): the increase in the number of population doublings in chemically-exposed cultures versus increase in non-treated cultures, a ratio expressed as a percentage.

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

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

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

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

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

Untreated controls: cultures that receive no treatment (i.e. no test chemical nor solvent) but are processed concurrently in the same way as the cultures receiving the test chemical.

Appendix 2

FORMULAS FOR CYTOTOXICITY ASSESSMENT

Mitotic index (MI):

$$\text{MI}(\%) = \frac{\text{Number of mitotic cells}}{\text{Total number of Cells scored}} \times 100$$

Relative Increase in Cell Counts (RICC) or **Relative Population Doubling (RPD)** is recommended, as both take into account the proportion of the cell population which has divided.

$$\text{RICC}(\%) = \frac{(\text{Increase in number of cells in treated cultures}(\text{final-starting}))}{(\text{Increase in numbers of cells in control cultures}(\text{final-starting}))} \times 100$$

$$\text{RPD}(\%) = \frac{(\text{No. of Population doublings in treated cultures})}{(\text{No. of population doublings in control cultures})} \times 100$$

where:

$$\text{Population Doubling} = [\log (\text{Post-treatment cell number} \div \text{Initial cell number})] \div \log 2$$

For example, a RICC, or a RPD of 53 % indicates 47 % cytotoxicity/cytostasis and 55 % cytotoxicity/cytostasis measured by MI means that the actual MI is 45 % of control.

In any case, the number of cells before treatment should be measured and the same for treated and negative control cultures.

While RCC (i.e. Number of cells in treated cultures/Number of cells in control cultures) had been used as cytotoxicity parameter in the past, is no longer recommended because it can underestimate cytotoxicity

In the negative control cultures, population doubling should be compatible with the requirement to sample cells after treatment at a time equivalent to about 1,5 normal cell cycle length and mitotic index should be higher enough to get a sufficient number of cells in mitosis and to reliably calculate a 50 % reduction.'

- (4) In Part B, Chapter B.11 is replaced by the following:

'B.11 Mammalian Bone Marrow Chromosomal Aberration Test

INTRODUCTION

This test method is equivalent to OECD test guideline 475 (2016). It is part of a series of test methods on genetic toxicology. An OECD document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to these Test Guidelines has been developed (1).

The mammalian *in vivo* bone marrow chromosomal aberration test is especially relevant for 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 responses. An *in vivo* assay is also useful for further investigation of genotoxicity detected by an *in vitro* system.

The mammalian *in vivo* chromosomal aberration test is used for the detection of structural chromosome aberrations induced by test chemicals in bone marrow cells of animals, usually rodents (2) (3) (4) (5). Structural chromosomal aberrations may be of two types, chromosome or chromatid. While the majority of genotoxic chemical-induced aberrations are of the chromatid-type, chromosome-type aberrations also occur. Chromosomal damage and related events are the cause of many human genetic diseases and there is substantial evidence that, when these lesions and related events cause alterations in oncogenes and tumour suppressor genes, they are involved in cancer in humans and experimental systems. Polyploidy (including endoreduplication) could arise in chromosome aberration assays *in vivo*. However, an increase in polyploidy per se does not indicate aneugenic potential and can simply indicate cell cycle perturbation or cytotoxicity. This test is not designed to measure aneuploidy. An *in vivo* mammalian erythrocyte micronucleus test (Chapter B.12 of this Annex) or the *in vitro* mammalian cell micronucleus test (Chapter B.49 of this Annex) would be the *in vivo* and *in vitro* tests, respectively, recommended for the detection of aneuploidy.

Definitions of terminology used are set out in Appendix 1.

INITIAL CONSIDERATIONS

Rodents are routinely used in this test, but other species may in some cases be appropriate if scientifically justified. Bone marrow is the target tissue in this test since it is a highly vascularised tissue and it contains a population of rapidly cycling cells that can be readily isolated and processed. The scientific justification for using species other than rats and mice should be provided in the report. If species other than rodents are used, it is recommended that the measurement of bone marrow chromosomal aberration be integrated into another appropriate toxicity test.

If there is evidence that the test chemical(s), or its metabolite(s), will not reach the target tissue, it may not be appropriate to use this test.

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 METHOD

Animals are exposed to the test chemical by an appropriate route of exposure and are humanely euthanised at an appropriate time after treatment. Prior to euthanasia, animals are treated with a metaphase-arresting agent (e.g. colchicine or colcemid). Chromosome preparations are then made from the bone marrow cells and stained, and metaphase cells are analysed for chromosomal aberrations.

VERIFICATION OF LABORATORY PROFICIENCY

Proficiency Investigations

In order to establish sufficient experience with the conduct of the assay prior to using it for routine testing, the laboratory should have demonstrated the ability to reproduce expected results from published data (e.g. (6)) for chromosomal aberration frequencies with a minimum of two positive control chemicals (including weak responses induced by low doses of positive controls), such as those listed in Table 1 and with compatible vehicle/solvent controls (see paragraph 22). These experiments should use doses that give reproducible and dose related increases and demonstrate the sensitivity and dynamic range of the test system in the tissue of interest (bone marrow) and using the scoring method to be employed within the laboratory. This requirement is not applicable to laboratories that have experience, i.e. that have a historical database available as defined in paragraphs 10-14.

Historical Control Data

During the course of the proficiency investigations, the laboratory should establish:

- A historical positive control range and distribution, and
- A historical negative control range and distribution.

When first acquiring data for a historical negative control distribution, concurrent negative controls should be consistent with published control data, where they exist. As more experimental data are added to the historical control distribution, concurrent negative controls should ideally be within the 95 % control limits of that distribution. The laboratory's historical negative control database should be statistically robust to ensure the ability of the laboratory to assess the distribution of their negative control data. The literature suggests that a minimum of 10 experiments may be necessary 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 (7)), to identify how variable their data are, and to show that the methodology is 'under control' in their laboratory. Further recommendations on how to build and use the historical data (i.e. criteria for inclusion and exclusion of data in historical data and the acceptability criteria for a given experiment) can be found in the literature (8).

Where the laboratory does not complete a sufficient number of experiments to establish a statistically robust negative control distribution (see paragraph 11) during the proficiency investigations (described in paragraph 9), it is acceptable that the distribution can be built during the first routine tests. This approach should follow the recommendations set out in the literature (8) and the negative control results obtained in these experiments should remain consistent with published negative control data.

Any changes to the experimental protocol should be considered in terms of their impact on the resulting data remaining consistent with the laboratory's existing historical control database. Only major inconsistencies should result in the establishment of a new historical control database, where expert judgement determines that it differs from the previous distribution (see paragraph 11). During the re-establishment, a full negative control database may not be needed to permit the conduct of an actual test, provided that the laboratory can demonstrate that their concurrent negative control values remain either consistent with their previous database or with the corresponding published data.

Negative control data should consist of the incidence of structural chromosomal aberration (excluding gaps) in each animal. Concurrent negative controls should ideally be within the 95 % control limits of the distribution of the laboratory's historical negative control database. Where concurrent negative control data fall outside the 95 % control limits, they may be acceptable for inclusion in the historical control distribution as long as these data are not extreme outliers and there is evidence that the test system is 'under control' (see paragraph 11) and no evidence of technical or human failure.

DESCRIPTION OF THE METHOD

Preparations

Selection of animal species

Commonly used laboratory strains of healthy young adult animals should be employed. Rats are commonly used, although mice 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

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) of the same sex and treatment group if no aggressive behaviour is expected, preferably in solid floor cages with appropriate environmental enrichment. Animals may be housed individually only if scientifically justified.

Preparation of the animals

Healthy young adult animals (for rodents, ideally 6-10 weeks old at start of treatment, though slightly older animals are also acceptable) 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 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

Solid test chemicals should be dissolved or suspended in appropriate solvents or vehicles or admixed in diet or drinking water prior to dosing the animals. Liquid test chemicals may be dosed directly or diluted prior to dosing. For inhalation exposures, test chemicals can be administered as a 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.

Solvent/vehicle

The solvent/vehicle should not produce toxic effects at the dose levels used, and should not be suspected 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 aberrations or 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.

Controls

Positive controls

A group of animals treated with a positive control chemical should normally be included with each test. This may be waived when the testing laboratory has demonstrated proficiency in the conduct of the test and has established a historical positive control range. 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.

Positive control chemicals should reliably produce a detectable increase in the frequency of cells with structural chromosomal aberrations over the spontaneous level. 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. It is acceptable that the positive control be administered by a route different from the test chemical, using a different treatment schedule, and for sampling to occur only at a single time point. In addition, the use of chemical class-related positive control chemicals may be considered, when appropriate. Examples of positive control chemicals are included in Table 1.

Table 1

Examples of positive control chemicals

Chemical	CASRN
Ethyl methanesulphonate	62-50-0
Methyl methanesulphonate	66-27-3
Ethyl nitrosourea	759-73-9
Mitomycin C	50-07-7
Cyclophosphamide (monohydrate)	50-18-0 (6055-19-2)
Triethylenemelamine	51-18-3

Negative controls

Negative control group animals should be included at every sampling time and otherwise handled in the same way as the treatment groups, except for not receiving treatment with the test chemical. If a solvent/vehicle is used in administering the test chemical, the control group should receive this solvent/vehicle. However, if consistent inter-animal variability and frequencies of cells with structural aberrations are demonstrated by historical negative control data at each sampling time for the testing laboratory, only a single sampling for the negative control may be necessary. Where a single sampling is used for negative controls, it should be the first sampling time used in the study.

PROCEDURE

Number and sex of animals

In general, the micronucleus response is similar between male and female animals (9) and it is expected that this will be true also for structural chromosomal aberrations; therefore, most studies could be performed in either sex. Data demonstrating relevant differences between males and females (e.g. differences in systemic toxicity, metabolism, bioavailability, bone marrow toxicity, etc. including e.g. a range-finding study) would encourage the use of both sexes. In this case, it may be appropriate to perform a study in both sexes, e.g. as part of a repeated dose toxicity study. It might be appropriate to use the factorial design in case both sexes are used. Details on how to analyse the data using this design are given in Appendix 2.

Group sizes at study initiation should be established with the aim of providing a minimum of 5 analysable animals of one sex, or of each sex if both are used, per group. Where human exposure to chemicals may be sex-specific, as for example with some pharmaceuticals, the test should be performed with the appropriate sex. As a guide to maximum typical animal requirements, a study in bone marrow at two sampling times with three dose groups and a concurrent negative control group, plus a positive control group (each group composed of five animals of a single sex), would require 45 animals.

Dose levels

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 (10). 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, by inducing body weight depression or hematopoietic system cytotoxicity), but not death or evidence of pain, suffering or distress necessitating humane euthanasia (11).

The highest dose may also be defined as a dose that produces some indication of toxicity to the bone marrow.

Chemicals that exhibit saturation of toxicokinetic properties, or induce detoxification processes that may lead to a decrease in exposure after long-term treatment may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis.

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. When target tissue (bone marrow) 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.

Limit test

If dose range-finding experiments, or existing data from related animal strains, indicate that a treatment regime of at least the limit dose (described below) produces no observable toxic effects, (including no depression of bone marrow proliferation or other evidence of target tissue cytotoxicity), and if genotoxicity would not be expected based upon *in vitro* genotoxicity studies or data from structurally related chemicals, then a full study using three dose levels may not be considered necessary, provided it has been demonstrated that the test chemical(s) reach(es) the target tissue (bone marrow). In such cases, a single dose level, at the limit dose, may be sufficient. For an administration period of > 14 days, the limit dose is 1 000 mg/kg body weight/day. For administration periods of 14 days or less, the limit dose is 2 000 mg/kg/body weight/day.

Administration of doses

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, intratracheal, 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 generally not 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 sampling should be sufficient to allow detection of the effects (see paragraphs 33-34). The maximum volume of liquid that can be administered by gavage or injection at one time depends on the size of the test animal. The volume should not normally exceed 1 ml/100 g body weight except in the case of aqueous solutions where a maximum of 2 ml/100 g may be used. The use of volumes greater than this should be justified. Except for irritating or corrosive test chemicals, which will normally produce exacerbated effects at higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure administration of a constant volume in relation to body weight at all dose levels.

Treatment schedule

Test chemicals are normally administered as a single treatment, but may be administered as a split dose (i.e. two or more treatments on the same day separated by no more than 2-3 hours) to facilitate administering a large volume. Under these circumstances, or when administering the test chemical by inhalation, the sampling time should be scheduled based on the time of the last dosing or the end of exposure.

There are little data available on the suitability of a repeated-dose protocol for this test. However, in circumstances where it is desirable to integrate this test with a repeated-dose toxicity test, care should be taken to avoid loss of chromosomally damaged mitotic cells as may occur with toxic doses. Such integration is acceptable when the highest dose is greater or equal to the limit dose (see paragraph 29) and a dose group is administered the limit dose for the duration of the treatment period. The micronucleus test (test method B.12) should be viewed as the *in vivo* test of choice for chromosomal aberrations when integration with other studies is desired.

Bone marrow samples should be taken at two separate times following single treatments. For rodents, the first sampling interval should be the time necessary to complete 1,5 normal cell cycle lengths (the latter being normally 12-18 hours following the treatment period). 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, a later sample collection 24 hours after the first sampling time is recommended. At the first sampling time, all dose groups should be treated and samples collected for analysis; however, at the later sampling time(s), only the highest dose needs to be administered. If dose regimens of more than one day are used based on scientific justification, one sampling time at up to approximately 1,5 normal cell cycle lengths after the final treatment should generally be used.

Following treatment and prior to sample collection, animals are injected intraperitoneally with an appropriate dose of a metaphase-arresting agent (e.g. colcemid or colchicine), and samples are collected at an appropriate interval thereafter. For mice this interval is approximately 3-5 hours prior to collection and for rats it is 2-5 hours. Cells are harvested from the bone marrow, swollen, fixed and stained, and analysed for chromosomal aberrations (12).

Observations

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 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 excessive toxicity should be humanely euthanised prior to completion of the test period (11).

Target tissue exposure

A blood sample should be taken at appropriate time(s) in order to permit investigation of the plasma levels of the test chemicals for the purposes of demonstrating that exposure of the bone marrow occurred, where warranted and where other exposure data do not exist (see paragraph 44).

Bone marrow and chromosome preparations

Immediately after humane euthanasia, bone marrow cells are obtained from the femurs or tibias of the animals, exposed to hypotonic solution and fixed. The metaphase cells are then spread on slides and stained using established methods (see (3) (12)).

Analysis

All slides, including those of positive and negative controls, should be independently coded before analysis and should be randomised so the scorer is unaware of the treatment condition.

The mitotic index should be determined as a measure of cytotoxicity in at least 1 000 cells per animal for all treated animals (including positive controls), untreated or vehicle/solvent negative control animals.

At least 200 metaphases should be analysed for each animal for structural chromosomal aberrations including and excluding gaps (6). However, if the historical negative control database indicates the mean background structural chromosomal aberration frequency is $< 1\%$ in the testing laboratory, consideration should be given to scoring additional cells. Chromatid and chromosome-type aberrations should be recorded separately and classified by sub-types (breaks, exchanges). Procedures in use in the laboratory should ensure that analysis of chromosomal aberrations is performed by well-trained scorers and peer-reviewed if appropriate. 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.

DATA AND REPORTING

Treatment of Results

Individual animal data should be presented in tabular form. The mitotic index, the number of metaphase cells scored, the number of aberrations per metaphase cell and the percentage of cells with structural chromosomal aberration(s) should be evaluated for each animal. Different types of structural chromosomal aberrations should be listed with their numbers and frequencies for treated and control groups. Gaps, as well as polyploid cells and cells with endoreduplicated chromosomes are recorded separately. The frequency of gaps is reported but generally not included in the analysis of the total structural aberration frequency. If there is no evidence for a difference in response between the sexes, the data may be combined for statistical analysis. Data on animal toxicity and clinical signs should also be reported.

Acceptability Criteria

The following criteria determine the acceptability of the test:

- (a) The concurrent negative control data are considered acceptable for addition to the laboratory historical control database (see paragraphs 11-14);
- (b) The concurrent positive controls or scoring controls should induce responses that are compatible with those generated in the historical positive control database and produce a statistically significant increase compared with the negative control (see paragraphs 20-21);
- (c) The appropriate number of doses and cells has been analysed;
- (d) The criteria for the selection of highest dose are consistent with those described in paragraphs 25-28.

Evaluation and Interpretation of Results

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

- (a) At least one of the treatment groups exhibits a statistically significant increase in the frequency of cells with structural chromosomal aberrations (excluding gaps) compared with the concurrent negative control,
- (b) This increase is dose-related at least at one sampling time when evaluated with an appropriate trend test, and
- (c) Any of these results are outside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits).

If only the highest dose is examined at a particular sampling time, a test chemical is considered clearly positive if there is a statistically significant increase compared with the concurrent negative control and the results are outside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits). Recommendations for appropriate statistical methods can be found in the literature (13). When conducting a dose-response analysis, at least three treated dose groups should be analysed. Statistical tests should use the animal as the experimental unit. Positive results in the chromosomal aberration test indicate that a test chemical induces structural chromosomal aberrations in the bone marrow of the species tested.

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

- (a) None of the treatment groups exhibits a statistically significant increase in the frequency of cells with structural chromosomal aberrations (excluding gaps) compared with the concurrent negative control,

- (b) There is no dose-related increase at any sampling time when evaluated by an appropriate trend test,
- (c) All results are inside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits), and
- (d) Bone marrow exposure to the test chemical(s) occurred.

Recommendations for the most appropriate statistical methods can be found in the literature (13). Evidence of exposure of the bone marrow to a test chemical may include a depression of the mitotic index or measurement of the plasma or blood levels of the test chemical(s). In the case of intravenous administration, evidence of exposure is not needed. Alternatively, ADME data, obtained in an independent study using the same route and same species can be used to demonstrate bone marrow exposure. Negative results indicate that, under the test conditions, the test chemical does not induce structural chromosomal aberrations in the bone marrow of the species tested.

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

In cases where 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 judgement and/or further investigations of the existing experiments completed. In some cases, analysing more cells or performing a repeat experiment using modified experimental conditions could be useful.

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

The frequencies of polyploid and endoreduplicated metaphases among total metaphases should be recorded separately. An increase in the number of polyploid/endoreduplicated cells may indicate that the test chemical has the potential to inhibit mitotic processes or cell cycle progression (see paragraph 3).

Test Report

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, 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.

Test chemical preparation:

- justification for choice of vehicle;
- solubility and stability of the test chemical in 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 use;
- number, age and sex 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) controls;
- data from range-finding study, if conducted;
- rationale for dose level selection;
- details of test chemical preparation;
- details of the administration of the test chemical;
- rationale for route and duration of administration;
- methods for verifying that the test chemical(s) reached the general circulation or bone marrow;
- 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;
- method of euthanasia;
- method of analgesia (where used);
- detailed description of treatment and sampling schedules and justifications for the choices;
- methods of slide preparation;
- methods for measurement of toxicity;
- identity of metaphase arresting chemical, its concentration, dose and time of administration before sampling;
- procedures for isolating and preserving samples;
- criteria for scoring aberrations;

- number of metaphase cells analysed per animal and the number of cells analysed for mitotic index determination;
- criteria for acceptability of the study;
- criteria for considering studies as positive, negative or inconclusive.

Results:

- animal condition prior to and throughout the test period, including signs of toxicity;
- mitotic index, given separately for each animal;
- type and number of aberrations and of aberrant cells, 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;
- changes in ploidy, if seen, including frequencies of polyploid and/or endoreduplicated cells;
- dose-response relationship, where possible;
- statistical analyses and method applied;
- data supporting that exposure of the bone marrow occurred;
- concurrent negative control and positive control data with ranges, means and standard deviations;
- historical negative and positive control data with ranges, means, standard deviations, and 95 % control limits for the distribution, as well as the time period covered and number of observations;
- criteria met for a positive or negative response.

Discussion of the results.

Conclusion.

References.

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*, Venittand, S., J.M. Parry (eds.), IRL Press, Washington, DC, pp. 275-306.
- (3) Preston, R.J. *et al.* (1987), *Mammalian in vivo cytogenetic assays. Analysis of chromosome aberrations in bone marrow cells*, Mutation Research, Vol. 189/2, pp. 157-165.
- (4) Richold, M. *et al.* (1990), "In Vivo Cytogenetics Assays", in *Basic Mutagenicity Tests, UKEMS Recommended Procedures. UKEMS Subcommittee on Guidelines for Mutagenicity Testing. Report. Part I revised*, Kirkland, D. J. (ed.), Cambridge University Press, Cambridge, pp. 115-141.
- (5) Tice, R.R. *et al.* (1994), *Report from the working group on the in vivo mammalian bone marrow chromosomal aberration test*, Mutation Research, Vol. 312/3, pp. 305-312.

- (6) Adler, I.D. *et al.* (1998), Recommendations for statistical designs of *in vivo* mutagenicity tests with regard to subsequent statistical analysis, Mutation Research/Genetic Toxicology and Environmental Mutagenesis, Vol. 417/1, pp. 19-30.
 - (7) Ryan, T.P. (2000), Statistical Methods for Quality Improvement, 2nd ed., John Wiley and Sons, New York.
 - (8) Hayashi, M. *et al.* (2011), Compilation and use of genetic toxicity historical control data, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, Vol. 723/2, pp. 87-90.
 - (9) Hayashi, M. *et al.* (1994), *in vivo* rodent erythrocyte micronucleus assay, Mutation Research/Environmental Mutagenesis and Related Subjects, Vol. 312/3, pp. 293-304.
 - (10) Fielder, R.J. *et al.* (1992), Report of British Toxicology Society/UK Environmental Mutagen Society Working Group. Dose setting in *in vivo* mutagenicity assays, Mutagenesis, Vol. 7/5, pp. 313-319.
 - (11) OECD (2000), "Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, N°19, OECD Publishing, Paris.
 - (12) Pacchierotti, F., V. Stocchi (2013), Analysis of chromosome aberrations in somatic and germ cells of the mouse, Methods in Molecular Biology, Vol. 1044, pp. 147-163.
 - (13) Lovell, D.P. *et al.* (1989), "Statistical Analysis of *in vivo* Cytogenetic Assays", in Statistical Evaluation of Mutagenicity Test Data. UKEMS SubCommittee on Guidelines for Mutagenicity Testing, Report, Part III, Kirkland, D.J. (ed.), Cambridge University Press, Cambridge, pp. 184-232.
-

Appendix 1

DEFINITIONS

Aneuploidy: Any deviation from the normal diploid (or haploid) number of chromosomes by one or more chromosomes, but not by multiples of entire set(s) of chromosomes (*cf.* 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.

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.

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

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

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

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

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

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

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

*Appendix 2***THE FACTORIAL DESIGN FOR IDENTIFYING SEX DIFFERENCES IN THE *IN VIVO* CHROMOSOMAL ABERRATION ASSAY****The factorial design and its analysis**

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

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

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

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

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

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

References

There are many statistical textbooks which discuss the theory, design, methodology, analysis and interpretation of factorial designs ranging from the simplest two factor analyses to the more complex forms used in Design of Experiment methodology. The following is a non-exhaustive list. Some books provide worked examples of comparable designs, in some cases with code for running the analyses using various software packages.

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

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

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

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

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

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

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

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

- (5) In Part B, Chapter B.12 is replaced by the following:

'B.12 Mammalian Erythrocyte Micronucleus Test

INTRODUCTION

This test method is equivalent to OECD test guideline 474 (2016). It is part of a series of test methods on genetic toxicology. An OECD document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to these Test Guidelines has been developed (1).

The mammalian *in vivo* micronucleus test is especially relevant for 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 responses. An *in vivo* assay is also useful for further investigation of genotoxicity detected by an *in vitro* system.

The mammalian *in vivo* micronucleus test is used for the detection of damage induced by the test chemical to the chromosomes or the mitotic apparatus of erythroblasts. The test evaluates micronucleus formation in erythrocytes sampled either in the bone marrow or peripheral blood cells of animals, usually rodents.

The purpose of the micronucleus test is to identify chemicals that cause cytogenetic damage which results in the formation of micronuclei containing either lagging chromosome fragments or whole chromosomes.

When a bone marrow erythroblast develops into an immature erythrocyte (sometimes also referred to as a polychromatic erythrocyte or reticulocyte), the main nucleus is extruded; any micronucleus that has been formed may remain behind in the cytoplasm. Visualisation or detection of micronuclei is facilitated in these cells because they lack a main nucleus. An increase in the frequency of micronucleated immature erythrocytes in treated animals is an indication of induced structural or numerical chromosomal aberrations.

Newly formed micronucleated erythrocytes are identified and quantitated by staining followed by either visual scoring using a microscope, or by automated analysis. Counting sufficient immature erythrocytes in the peripheral blood or bone marrow of adult animals is greatly facilitated by using an automated scoring platform. Such platforms are acceptable alternatives to manual evaluation (2). Comparative studies have shown that such methods, using appropriate calibration standards, can provide better inter- and intra-laboratory reproducibility and sensitivity than manual microscopic scoring (3) (4). Automated systems that can measure micronucleated erythrocyte frequencies include, but are not limited to, flow cytometers (5), image analysis platforms (6) (7), and laser scanning cytometers (8).

Although not normally done as part of the test, chromosome fragments can be distinguished from whole chromosomes by a number of criteria. These include identification of the presence or absence of a kinetochore or centromeric DNA, both of which are characteristic of intact chromosomes. The absence of kinetochore or centromeric DNA indicates that the micronucleus contains only fragments of chromosomes, while the presence is indicative of chromosome loss.

Definitions of terminology used are set out in Appendix 1.

INITIAL CONSIDERATIONS

The bone marrow of young adult rodents is the target tissue for genetic damage in this test since erythrocytes are produced in this tissue. The measurement of micronuclei in immature erythrocytes in peripheral blood is acceptable in other mammalian species for which adequate sensitivity to detect chemicals that cause structural or numerical chromosomal aberrations in these cells has been demonstrated (by induction of micronuclei in immature erythrocytes) and scientific justification is provided. The frequency of micronucleated immature erythrocytes is the principal endpoint. The frequency of mature erythrocytes that contain micronuclei in the peripheral blood also can be used as an endpoint in species without strong splenic selection against micronucleated cells and when animals are treated continuously for a period that exceeds the lifespan of the erythrocyte in the species used (e.g. 4 weeks or more in the mouse).

If there is evidence that the test chemical(s), or its metabolite(s), will not reach the target tissue, it may not be appropriate to use this test.

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 METHOD

Animals are exposed to the test chemical by an appropriate route. If bone marrow is used, the animals are humanely euthanised at an appropriate time(s) after treatment, the bone marrow is extracted, and preparations are made and stained (9) (10) (11) (12) (13) (14) (15). When peripheral blood is used, the blood is collected at an appropriate time(s) after treatment and preparations are made and stained (12) (16) (17) (18). When treatment is administered acutely, it is important to select bone marrow or blood harvest times at which the treatment-related induction of micronucleated immature erythrocytes can be detected. In the case of peripheral blood sampling, enough time must also have elapsed for these events to appear in circulating blood. Preparations are analysed for the presence of micronuclei, either by visualisation using a microscope, image analysis, flow cytometry, or laser scanning cytometry.

VERIFICATION OF LABORATORY PROFICIENCY

Proficiency Investigations

In order to establish sufficient experience with the conduct of the assay prior to using it for routine testing, the laboratory should have demonstrated the ability to reproduce expected results from published data (17) (19) (20) (21) (22) for micronucleus frequencies with a minimum of two positive control chemicals (including weak responses induced by low doses of positive controls), such as those listed in Table 1 and with compatible vehicle/solvent controls (see paragraph 26). These experiments should use doses that give reproducible and dose-related increases and demonstrate the sensitivity and dynamic range of the test system in the tissue of interest (bone marrow or peripheral blood) and using the scoring method to be employed within the laboratory. This requirement is not applicable to laboratories that have experience, i.e. that have a historical database available as defined in paragraphs 14-18.

Historical Control Data

During the course of the proficiency investigations, the laboratory should establish:

- A historical positive control range and distribution, and
- A historical negative control range and distribution.

When first acquiring data for a historical negative control distribution, concurrent negative controls should be consistent with published control data, where they exist. As more experimental data are added to the historical control distribution, concurrent negative controls should ideally be within the 95 % control limits of that distribution. The laboratory's historical negative control database should be statistically robust to ensure the ability of the laboratory to assess the distribution of their negative control data. The literature suggests that a minimum of 10 experiments may be necessary 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 (23)), to identify how variable their data are, and to show that the methodology is 'under control' in their laboratory. Further recommendations on how to build and use the historical data (i.e. criteria for inclusion and exclusion of data in historical data and the acceptability criteria for a given experiment) can be found in the literature (24).

Where the laboratory does not complete a sufficient number of experiments to establish a statistically robust negative control distribution (see paragraph 15) during the proficiency investigations (described in paragraph 13), it is acceptable that the distribution can be built during the first routine tests. This approach should follow the recommendations set out in the literature (24) and the negative control results obtained in these experiments should remain consistent with published negative control data.

Any changes to the experimental protocol should be considered in terms of their impact on the resulting data remaining consistent with the laboratory's existing historical control database. Only major inconsistencies should result in the establishment of a new historical control database where expert judgement determines that it differs from the previous distribution (see paragraph 15). During the re-establishment, a full negative control database may not be needed to permit the conduct of an actual test, provided that the laboratory can demonstrate that their concurrent negative control values remain either consistent with their previous database or with the corresponding published data.

Negative control data should consist of the incidence of micronucleated immature erythrocytes in each animal. Concurrent negative controls should ideally be within the 95 % control limits of the distribution of the laboratory's historical negative control database. Where concurrent negative control data fall outside the 95 % control limits, they may be acceptable for inclusion in the historical control distribution as long as these data are not extreme outliers and there is evidence that the test system is 'under control' (see paragraph 15) and no evidence of technical or human failure.

DESCRIPTION OF THE METHOD

Preparations

Selection of animal species

Commonly used laboratory strains of healthy young adult animals should be employed. Mice, rats, or another appropriate mammalian species may be used. When peripheral blood is used, it must be established that splenic removal of micronucleated cells from the circulation does not compromise the detection of induced micronuclei in the species selected. This has been clearly demonstrated for mouse and rat peripheral blood (2). The scientific justification for using species other than rats and mice should be provided in the report. If species other than rodents are used, it is recommended that the measurement of induced micronuclei be integrated into another appropriate toxicity test.

Animal housing and feeding conditions

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) of the same sex and treatment group if no aggressive behaviour is expected, preferably in solid floor cages with appropriate environmental enrichment. Animals may be housed individually only if scientifically justified.

Preparation of the animals

Healthy young adult animals (for rodents, ideally 6-10 weeks old at start of treatment, though slightly older animals are also acceptable) 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 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

Solid test chemicals should be dissolved or suspended in appropriate solvents or vehicles or admixed in diet or drinking water prior to dosing the animals. Liquid test chemicals may be dosed directly or diluted prior to dosing. For inhalation exposures, test chemicals can be administered as a 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

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 micronuclei 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.

Controls

Positive controls

A group of animals treated with a positive control chemical should normally be included with each test. This may be waived when the testing laboratory has demonstrated proficiency in the conduct of the test and has established a historical positive control range. When a concurrent positive control group is not included, scoring controls (fixed and unstained slides or cell suspension samples, as appropriate for the method of scoring) 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); for example, during proficiency testing and on a regular basis thereafter, where necessary.

Positive control chemicals should reliably produce a detectable increase in micronucleus frequency over the spontaneous level. When employing manual scoring by microscopy, 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. It is acceptable that the positive control be administered by a route different from the test chemical, using a different treatment schedule, and for sampling to occur only at a single time point. In addition, the use of chemical class-related positive control chemicals may be considered, when appropriate. Examples of positive control chemicals are included in Table 1.

Table 1

Examples of positive control chemicals

Chemicals and CASRN
Ethyl methanesulphonate [CASRN 62-50-0]
Methyl methanesulphonate [CASRN 66-27-3]
Ethyl nitrosourea [CASRN 759-73-9]
Mitomycin C [CASRN 50-07-7]
Cyclophosphamide (monohydrate) [CASRN 50-18-0 (CASRN 6055-19-2)]
Triethylenemelamine [CASRN 51-18-3]
Colchicine [CASRN 64-86-8] or Vinblastine [CASRN 865-21-4] — as aneugens

Negative controls

Negative control group animals should be included at every sampling time and otherwise handled in the same way as the treatment groups, except for not receiving treatment with the test chemical. If a solvent/vehicle is used in administering the test chemical, the control group should receive this solvent/vehicle. However, if consistent inter-animal variability and frequencies of cells with micronuclei are demonstrated by historical negative control data at each sampling time for the testing laboratory, only a single sampling for the negative control may be necessary. Where a single sampling is used for negative controls, it should be the first sampling time used in the study.

If peripheral blood is used, a pre-treatment sample is acceptable instead of a concurrent negative control for short-term studies when the resulting data are consistent with the historical control database for the testing laboratory. It has been shown for rats that pre-treatment sampling of small volumes (e.g. below 100 µl/day) has minimal impact on micronucleus background frequency (25).

PROCEDURE

Number and sex of animals

In general, the micronucleus response is similar between male and female animals and, therefore, most studies could be performed in either sex (26). Data demonstrating relevant differences between males and females (e.g. differences in systemic toxicity, metabolism, bioavailability, bone marrow toxicity, etc. including e.g. in a range-finding study) would encourage the use of both sexes. In this case, it may be appropriate to perform a study in both sexes, e.g. as part of a repeated dose toxicity study. It might be appropriate to use the factorial design in case both sexes are used. Details on how to analyse the data using this design are given in Appendix 2.

Group sizes at study initiation should be established with the aim of providing a minimum of 5 analysable animals of one sex, or of each sex if both are used, per group. Where human exposure to chemicals may be sex-specific, as for example with some pharmaceuticals, the test should be performed with the appropriate sex. As a guide to maximum typical animal requirements, a study in bone marrow conducted according to the parameters established in paragraph 37 with three dose groups and concurrent negative and positive controls (each group composed of five animals of a single sex) would require between 25 and 35 animals.

Dose levels

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 (27). 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, by inducing body weight depression or hematopoietic system cytotoxicity, but not death or evidence of pain, suffering or distress necessitating humane euthanasia (28)).

The highest dose may also be defined as a dose that produces toxicity in the bone marrow (e.g. a reduction in the proportion of immature erythrocytes among total erythrocytes in the bone marrow or peripheral blood of more than 50 %, but to not less than 20 % of the control value). However, when analysing CD71-positive cells in peripheral blood circulation (i.e., by flow cytometry), this very young fraction of immature erythrocytes responds to toxic challenges more quickly than the larger RNA-positive cohort of immature erythrocytes. Therefore, higher apparent toxicity may be evident with acute exposure designs examining the CD71-positive immature erythrocyte fraction as compared to those that identify immature erythrocytes based on RNA content. For this reason, when experiments utilise five or fewer days of treatment, the highest dose level for test chemicals causing toxicity may be defined as the dose that causes a statistically significant reduction in the proportion of CD71-positive immature erythrocytes among total erythrocytes but not to less than 5 % of the control value (29).

Chemicals that exhibit saturation of toxicokinetic properties, or induce detoxification processes that may lead to a decrease in exposure after long-term administration may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis.

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 an administration period of 14 days or more should be 1 000 mg/kg body weight/day, or for administration periods of less than 14 days, 2 000 mg/kg/body weight/day. 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 (bone marrow) 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.

Limit test

If dose range-finding experiments, or existing data from related animal strains, indicate that a treatment regime of at least the limit dose (described below) produces no observable toxic effects, (including no depression of bone marrow proliferation or other evidence of target tissue cytotoxicity), and if genotoxicity would not be expected based upon *in vitro* genotoxicity studies or data from structurally related chemicals, then a full study using three dose levels may not be considered necessary, provided it has been demonstrated that the test chemical(s) reach(es) the target tissue (bone marrow). In such cases, a single dose level, at the limit dose, may be sufficient. When administration occurs for 14 days or more, the limit dose is 1 000 mg/kg body weight/day. For administration periods of less than 14 days, the limit dose is 2 000 mg/kg/body weight/day.

Administration of doses

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, intratracheal, 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 generally not 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 sampling should be sufficient to allow detection of the effects (see paragraph 37). The maximum volume of liquid that can be administered by gavage or injection at one time depends on the size of the test animal. The volume should not normally exceed 1 ml/100 g body weight except in the case of aqueous solutions where a maximum of 2 ml/100 g may be used. The use of volumes greater than this should be justified. Except for irritating or corrosive test chemicals, which will normally produce exacerbated effects at higher concentrations, variability in test volume should be minimised by adjusting the concentration to ensure administration of a constant volume in relation to body weight at all dose levels.

Treatment schedule

Preferably, 2 or more treatments are performed, administered at 24-hour intervals, especially when integrating this test into other toxicity studies. In the alternative, single treatments can be administered, if scientifically justified (e.g. test chemicals known to block cell cycle). Test chemicals also may be administered as a split dose, i.e., two or more treatments on the same day separated by no more than 2-3 hours, to facilitate administering a large volume. Under these circumstances, or when administering the test chemical by inhalation, the sampling time should be scheduled based on the time of the last dosing or the end of exposure.

The test may be performed in mice or rats in one of three ways:

- (a) Animals are treated with the test chemical once. Samples of bone marrow are taken at least twice (from independent groups of animals), starting not earlier than 24 hours after treatment, but not extending beyond 48 hours after treatment with appropriate interval(s) between samples, unless a test chemical is known to have an exceptionally long half-life. The use of sampling times earlier than 24 hours after treatment should be justified. Samples of peripheral blood are taken at least twice (from the same group of animals), starting not earlier than 36 hours after treatment, with appropriate interval(s) following the first sample, but not extending beyond 72 hours. At the first sampling time, all dose groups should be treated and samples collected for analysis; however, at the later sampling time(s), only the highest dose needs to be administered. When a positive response is detected at one sampling time, additional sampling is not required unless quantitative dose-response information is needed. The described harvest times are a consequence of the kinetics of appearance and disappearance of the micronuclei in these 2 tissue compartments.
- (b) If 2 daily treatments are used (e.g. two treatments at 24 hour intervals), samples should be collected once between 18 and 24 hours following the final treatment for the bone marrow or once between 36 and 48 hours following the final treatment for peripheral blood (30). The described harvest times are a consequence of the kinetics of appearance and disappearance of the micronuclei in these 2 tissue compartments.
- (c) If three or more daily treatments are used (e.g. three or more treatments at approximately 24 hour intervals), bone marrow samples should be collected no later than 24 hours after the last treatment and peripheral blood should be collected no later than 40 hours after the last treatment (31). This treatment option accommodates combination of the comet assay (e.g. sampling 2-6 hours after the last treatment) with the micronucleus test, and integration of the micronucleus test with repeated-dose toxicity studies. Accumulated data suggested that micronucleus induction can be observed over these wider timeframes when 3 or more administrations have occurred (15).

Other dosing or sampling regimens may be used when relevant and scientifically justified, and to facilitate integration with other toxicity tests.

Observations

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 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 excessive toxicity should be humanely euthanised prior to completion of the test period (28). Under certain circumstances, animal body temperature could be monitored, since treatment-induced hyper- and hypothermia have been implicated in producing spurious results (32) (33) (34).

Target tissue exposure

A blood sample should be taken at appropriate time(s) in order to permit investigation of the plasma levels of the test chemicals for the purposes of demonstrating that exposure of the bone marrow occurred, where warranted and where other exposure data do not exist (see paragraph 48).

Bone marrow / blood preparation

Bone marrow cells are usually obtained from the femurs or tibias of the animals immediately following humane euthanasia. Commonly, cells are removed, prepared and stained using established methods. Small volumes of peripheral blood can be obtained, according to adequate animal welfare standards, either using a method that permits survival of the test animal, such as bleeding from the tail vein or other appropriate blood vessel, or by cardiac puncture or sampling from a large vessel at animal euthanasia. For both bone marrow or peripheral blood-derived erythrocytes, depending on the method of analysis, cells may be immediately stained supravitaly (16) (17) (18), smear preparations are made and then stained for microscopy, or fixed and stained appropriately for flow cytometric analysis. The use of a DNA specific stain [e.g. acridine orange (35) or Hoechst 33258 plus pyronin-Y (36)] can eliminate some of the artifacts associated with using a non-DNA specific stain. This advantage does not preclude the use of conventional stains (e.g. Giemsa for microscopic analysis). Additional systems [e.g. cellulose columns to remove nucleated cells (37) (38)] also can be used provided that these systems have been demonstrated to be compatible with sample preparation in the laboratory.

Where these methods are applicable, anti-kinetochore antibodies (39), FISH with pancentromeric DNA probes (40), or primed *in situ* labelling with pancentromere-specific primers, together with appropriate DNA counterstaining (41), can be used to identify the nature of the micronuclei (chromosome/chromosomal fragment) in order to determine whether the mechanism of micronucleus induction is due to clastogenic and/or aneugenic activity. Other methods for differentiation between clastogens and aneugens may be used if they have been shown to be effective.

Analysis (manual and automated)

All slides or samples for analysis, including those of positive and negative controls, should be independently coded before any type of analysis and should be randomised so the manual scorer is unaware of the treatment condition; such coding is not necessary when using automated scoring systems which do not rely on visual inspection and cannot be affected by operator bias. The proportion of immature among total (immature + mature) erythrocytes is determined for each animal by counting a total of at least 500 erythrocytes for bone marrow and 2 000 erythrocytes for peripheral blood (42). At least 4 000 immature erythrocytes per animal should be scored for the incidence of micronucleated immature erythrocytes (43). If the historical negative control database indicates the mean background micronucleated immature erythrocyte frequency is < 0,1 % in the testing laboratory, consideration should be given to scoring additional cells. When analysing samples, the proportion of immature erythrocytes to total erythrocytes in treated animals should not be less than 20 % of the vehicle/solvent control proportion when scoring by microscopy and not less than approximately 5 % of the vehicle/solvent control proportion when scoring CD71+ immature erythrocytes by cytometric methods (see paragraph 31) (29). For example, for a bone marrow assay scored by microscopy, if the control proportion of immature erythrocytes in the bone marrow is 50 %, the upper limit of toxicity would be 10 % immature erythrocytes.

Because the rat spleen sequesters and destroys micronucleated erythrocytes, to maintain high assay sensitivity when analysing rat peripheral blood, it is preferable to restrict the analysis of micronucleated immature erythrocytes to the youngest fraction. When using automated analysis methods, these most immature erythrocytes can be identified based on their high RNA content, or the high level of transferrin receptors (CD71+) expressed on their surface (31). However, direct comparison of different staining methods has shown that satisfactory results can be obtained with various methods, including conventional acridine orange staining (3) (4).

DATA AND REPORTING

Treatment of Results

Individual animal data should be presented in tabular form. The number of immature erythrocytes scored, the number of micronucleated immature erythrocytes, and the proportion of immature among total erythrocytes should be listed separately for each animal analysed. When mice are treated continuously for 4 weeks or more, the data on the number and proportion of micronucleated mature erythrocytes also should be given if collected. Data on animal toxicity and clinical signs should also be reported.

Acceptability Criteria

The following criteria determine the acceptability of the test:

- (a) The concurrent negative control data are considered acceptable for addition to the laboratory historical control database (see paragraphs 15-18).
- (b) The concurrent positive controls or scoring controls should induce responses that are compatible with those generated in the historical positive control database and produce a statistically significant increase compared with the concurrent negative control (see paragraphs 24-25).
- (c) The appropriate number of doses and cells has been analysed.
- (d) The criteria for the selection of highest dose are consistent with those described in paragraphs 30-33.

Evaluation and Interpretation of Results

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

- (a) At least one of the treatment groups exhibits a statistically significant increase in the frequency of micronucleated immature erythrocytes compared with the concurrent negative control,
- (b) This increase is dose-related at least at one sampling time when evaluated with an appropriate trend test, and
- (c) Any of these results are outside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits).

If only the highest dose is examined at a particular sampling time, a test chemical is considered clearly positive if there is a statistically significant increase compared with the concurrent negative control and the results are outside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits). Recommendations for the most appropriate statistical methods can be found in the literature (44) (45) (46) (47). When conducting a dose-response analysis, at least three treated dose groups should be analysed. Statistical tests should use the animal as the experimental unit. Positive results in the micronucleus test indicate that a test chemical induces micronuclei, which are the result of chromosomal damage or damage to the mitotic apparatus in the erythroblasts of the test species. In the case where a test was performed to detect centromeres within micronuclei, a test chemical that produces centromere-containing micronuclei (centromeric DNA or kinetochore, indicative of whole chromosome loss) is evidence that the test chemical is an aneugen.

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

- (a) None of the treatment groups exhibits a statistically significant increase in the frequency of micronucleated immature erythrocytes compared with the concurrent negative control,

- (b) There is no dose-related increase at any sampling time when evaluated by an appropriate trend test,
- (c) All results are inside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits), and
- (d) Bone marrow exposure to the test chemical(s) occurred.

Recommendations for the most appropriate statistical methods can be found in the literature (44) (45) (46) (47). Evidence of exposure of the bone marrow to a test chemical may include a depression of the immature to mature erythrocyte ratio or measurement of the plasma or blood levels of the test chemical. In case of intravenous administration, evidence of exposure is not needed. Alternatively, ADME data, obtained in an independent study using the same route and same species can be used to demonstrate bone marrow exposure. Negative results indicate that, under the test conditions, the test chemical does not produce micronuclei in the immature erythrocytes of the test species.

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

In cases where 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 judgement and/or further investigations of the existing experiments completed. In some cases, analysing more cells or performing a repeat experiment using modified experimental conditions could be useful.

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

Test Report

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, 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.

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 use;
- number, age and sex 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;
- details of test chemical preparation;
- details of the administration of the test chemical;
- rationale for route and duration of administration;
- methods for verifying that the test chemical(s) reached the general circulation or target tissue;
- 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;
- method of euthanasia;
- method of analgesia (where used);
- detailed description of treatment and sampling schedules and justifications for the choices;
- methods of slide preparation;
- procedures for isolating and preserving samples;
- methods for measurement of toxicity;
- criteria for scoring micronucleated immature erythrocytes;
- number of cells analysed per animal in determining the frequency of micronucleated immature erythrocytes and for determining the proportion of immature to mature erythrocytes;
- criteria for acceptability of the study;
- methods, such as use of anti-kinetochore antibodies or centromere-specific DNA probes, to characterise whether micronuclei contain whole or fragmented chromosomes, if applicable.

Results:

- animal condition prior to and throughout the test period, including signs of toxicity;
- proportion of immature erythrocytes among total erythrocytes;
- number of micronucleated immature erythrocytes, given separately for each animal;
- mean \pm standard deviation of micronucleated immature erythrocytes per group;
- dose-response relationship, where possible;
- statistical analyses and methods applied;
- concurrent negative and positive control data with ranges, means and standard deviations;
- historical negative and positive control data with ranges, means, standard deviations and 95 % control limits for the distribution, as well as the time period covered and the number of data points;
- data supporting that exposure of the bone marrow occurred;
- characterisation data indicating whether micronuclei contain whole or fragmented chromosomes, if applicable;
- criteria for a positive or negative response that are met.

*Discussion of the results.**Conclusion.**References.*

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) Hayashi, M. *et al.* (2007), *in vivo* erythrocyte micronucleus assay III. Validation and regulatory acceptance of automated scoring and the use of rat peripheral blood reticulocytes, with discussion of non-hematopoietic target cells and a single dose-level limit test, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 627/1, pp. 10-30.
- (3) MacGregor, J.T. *et al.* (2006), Flow cytometric analysis of micronuclei in peripheral blood reticulocytes: II. An efficient method of monitoring chromosomal damage in the rat, *Toxicology Sciences*, Vol. 94/1, pp. 92-107.
- (4) Dertinger, S.D. *et al.* (2006), Flow cytometric analysis of micronuclei in peripheral blood reticulocytes: I. Intra- and interlaboratory comparison with microscopic scoring, *Toxicological Sciences*, Vol. 94/1, pp. 83-91.
- (5) Dertinger, S.D. *et al.* (2011), Flow cytometric scoring of micronucleated erythrocytes: an efficient platform for assessing *in vivo* cytogenetic damage, *Mutagenesis*, Vol. 26/1, pp. 139-145.
- (6) Parton, J.W., W.P. Hoffman, M.L. Garriott (1996), Validation of an automated image analysis micronucleus scoring system, *Mutation Research*, Vol. 370/1, pp. 65-73.
- (7) Asano, N. *et al.* (1998), An automated new technique for scoring the rodent micronucleus assay: computerized image analysis of acridine orange supravital stained peripheral blood cells, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, Vol. 404/1-2, pp. 149-154.

- (8) Styles, J.A. *et al.* (2001), Automation of mouse micronucleus genotoxicity assay by laser scanning cytometry, *Cytometry*, Vol. 44/2, pp. 153-155.
- (9) Heddle, J.A. (1973), A rapid *in vivo* test for chromosomal damage, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, Vol. 18/2, pp. 187-190.
- (10) Schmid, W. (1975), The micronucleus test, *Mutation Research*, Vol. 31/1, pp. 9-15.
- (11) Heddle, J.A. *et al.* (1983), The induction of micronuclei as a measure of genotoxicity. A report of the U.S. Environmental Protection Agency Gene-Tox Program, *Mutation Research/Reviews in Genetic Toxicology*, Vol. 123/1, pp. 61-118.
- (12) Mavournin, K.H. *et al.* (1990), The *in vivo* micronucleus assay in mammalian bone marrow and peripheral blood. A report of the U.S. Environmental Protection Agency Gene-Tox Program, *Mutation Research/Reviews in Genetic Toxicology*, Vol. 239/1, pp. 29-80.
- (13) MacGregor, J.T. *et al.* (1983), Micronuclei in circulating erythrocytes: a rapid screen for chromosomal damage during routine toxicity testing in mice, *Developments in Toxicology Environmental Science*, Vol. 11, pp. 555-558.
- (14) MacGregor, J.T. *et al.* (1987), Guidelines for the conduct of micronucleus assays in mammalian bone marrow erythrocytes, *Mutation Research/Genetic Toxicology*, Vol. 189/2, pp. 103-112.
- (15) MacGregor, J.T. *et al.* (1990), The *in vivo* erythrocyte micronucleus test: measurement at steady state increases assay efficiency and permits integration with toxicity studies, *Fundamental and Applied Toxicology*, Vol. 14/3, pp. 513-522.
- (16) Hayashi, M. *et al.* (1990), The micronucleus assay with mouse peripheral blood reticulocytes using acridine orange-coated slides, *Mutation Research/Genetic Toxicology*, Vol. 245/4, pp. 245-249.
- (17) CSGMT/JEMS.MMS — The Collaborative Study Group for the Micronucleus Test (1992), Micronucleus test with mouse peripheral blood erythrocytes by acridine orange supravital staining: the summary report of the 5th collaborative study, *Mutation Research/Genetic Toxicology*, Vol. 278/2-3, pp. 83-98.
- (18) CSGMT/JEMS.MMS — The Mammalian Mutagenesis Study Group of the Environmental Mutagen Society of Japan (1995), Protocol recommended by the CSGMT/JEMS.MMS for the short-term mouse peripheral blood micronucleus test. The Collaborative Study Group for the Micronucleus Test (CSGMT) (CSGMT/JEMS.MMS, The Mammalian Mutagenesis Study Group of the Environmental Mutagen Society of Japan), *Mutagenesis*, Vol. 10/3, pp. 153-159.
- (19) Salamone, M.F., K.H. Mavournin (1994), Bone marrow micronucleus assay: a review of the mouse stocks used and their published mean spontaneous micronucleus frequencies, *Environmental and Molecular Mutagenesis*, Vol. 23/4, pp. 239-273.
- (20) Krishna, G., G. Urda, J. Paulissen (2000), Historical vehicle and positive control micronucleus data in mice and rats, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, Vol. 453/1, pp. 45-50.
- (21) Hayes, J. *et al.* (2009), The rat bone marrow micronucleus test—study design and statistical power, *Mutagenesis*, Vol. 24/5, pp. 419-424.
- (22) Wakata, A. *et al.* (1998), Evaluation of the rat micronucleus test with bone marrow and peripheral blood: summary of the 9th collaborative study by CSGMT/JEMS. MMS. Collaborative Study Group for the Micronucleus Test. Environmental Mutagen Society of Japan. Mammalian Mutagenicity Study Group, *Environmental and Molecular Mutagenesis*, Vol. 32/1, pp. 84-100.
- (23) Ryan, T.P. (2000), *Statistical Methods for Quality Improvement*, 2nd ed., John Wiley and Sons, New York.

- (24) Hayashi, M. *et al.* (2011), Compilation and use of genetic toxicity historical control data, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, Vol. 723/2, pp. 87-90.
- (25) Rothfuss, A. *et al.* (2011), Improvement of *in vivo* genotoxicity assessment: combination of acute tests and integration into standard toxicity testing, Mutation Research/Genetic Toxicology and Environmental Mutagenesis, Vol. 723/2, pp. 108-120.
- (26) Hayashi, M. *et al.* (1994), *in vivo* rodent erythrocyte micronucleus assay, Mutation Research/Environmental Mutagenesis and Related Subjects, Vol. 312/3, pp. 293-304.
- (27) Fielder, R.J. *et al.* (1992), Report of British Toxicology Society/UK Environmental Mutagen Society Working Group. Dose setting in *in vivo* mutagenicity assays, Mutagenesis, Vol. 7/5, pp. 313-319.
- (28) OECD (2000), "Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 19, OECD Publishing, Paris.
- (29) LeBaron, M.J. *et al.* (2013), Influence of counting methodology on erythrocyte ratios in the mouse micronucleus test, Environmental and Molecular Mutagenesis, Vol. 54/3, pp. 222-228.
- (30) Higashikuni, N., S. Sutou (1995), An optimal, generalized sampling time of 30 +/- 6 h after double dosing in the mouse peripheral blood micronucleus test, Mutagenesis, Vol. 10/4, pp. 313-319.
- (31) Hayashi, M. *et al.* (2000), *in vivo* rodent erythrocyte micronucleus assay. II. Some aspects of protocol design including repeated treatments, integration with toxicity testing, and automated scoring, Environmental and Molecular Mutagenesis, Vol. 35/3, pp. 234-252.
- (32) Asanami, S., K. Shimono (1997), High body temperature induces micronuclei in mouse bone marrow, Mutation Research/Genetic Toxicology and Environmental Mutagenesis, Vol. 390/1-2, pp. 79-83.
- (33) Asanami, S., K. Shimono, S. Kaneda (1998), Transient hypothermia induces micronuclei in mice, Mutation Research/Genetic Toxicology and Environmental Mutagenesis, Vol. 413/1, pp. 7-14.
- (34) Spencer, P.J. *et al.* (2007), Induction of micronuclei by phenol in the mouse bone marrow: I. Association with chemically induced hypothermia, Toxicological Sciences, Vol. 97/1, pp. 120-127.
- (35) Hayashi, M., T. Sofuni, M. Jr. Ishidate (1983), An application of Acridine Orange fluorescent staining to the micronucleus test, Mutation Research Letters, Vol. 120/4, pp. 241-247.
- (36) MacGregor, J.T., C.M. Wehr, R.G. Langlois (1983), A simple fluorescent staining procedure for micronuclei and RNA in erythrocytes using Hoechst 33258 and pyronin Y, Mutation Research, Vol. 120/4, pp. 269-275.
- (37) Romagna, F., C.D. Staniforth (1989), The automated bone marrow micronucleus test, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, Vol. 213/1, pp. 91-104.
- (38) Sun, J.T., M.J. Armstrong, S.M. Galloway (1999), Rapid method for improving slide quality in the bone marrow micronucleus assay; an adapted cellulose column procedure, Mutation Research, Vol. 439/1, pp. 121-126.
- (39) Miller, B.M., I.D. Adler (1990), Application of antikinetochore antibody staining (CREST staining) to micronuclei in erythrocytes induced *in vivo*, Mutagenesis, Vol. 5/4, pp. 411-415.
- (40) Miller, B.M. *et al.* (1991), Classification of micronuclei in murine erythrocytes: immunofluorescent staining using CREST antibodies compared to *in situ* hybridization with biotinylated gamma satellite DNA, Mutagenesis, Vol. 6/4, pp. 297-302.

- (41) Russo, A. (2002), PRINS tandem labeling of satellite DNA in the study of chromosome damage, *American Journal of Medical Genetics*, Vol. 107/2, pp. 99-104.
 - (42) Gollapudi, B.B., L.G. McFadden (1995), Sample size for the estimation of polychromatic to normochromatic erythrocyte ratio in the bone marrow micronucleus test, *Mutation Research*, Vol. 347/2, pp. 97-99.
 - (43) OECD (2014), "Statistical analysis supporting the revision of the genotoxicity Test Guidelines", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 198, OECD Publishing, Paris.
 - (44) Richold, M. *et al.* (1990), "In Vivo Cytogenetics Assays", in *Basic Mutagenicity Tests, UKEMS Recommended Procedures. UKEMS Subcommittee on Guidelines for Mutagenicity Testing. Report. Part I revised*, Kirkland, D. J. (ed.), Cambridge University Press, Cambridge, pp. 115-141.
 - (45) Lovell, D.P. *et al.* (1989), "Statistical Analysis of *in vivo* Cytogenetic Assays", in *Statistical Evaluation of Mutagenicity Test Data. UKEMS SubCommittee on Guidelines for Mutagenicity Testing, Report, Part III*, Kirkland, D.J. (ed.), Cambridge University Press, Cambridge, pp. 184-232.
 - (46) Hayashi, M. *et al.* (1994), Statistical analysis of data in mutagenicity assays: rodent micronucleus assay, *Environmental Health Perspectives*, Vol. 102/Suppl 1, pp. 49-52.
 - (47) Kim, B.S., M. Cho, H.J. Kim (2000), Statistical analysis of *in vivo* rodent micronucleus assay, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 469/2, pp. 233-241.
-

Appendix 1

DEFINITIONS

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.

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

Kinetochore: The protein structure that forms on the centromere of eukaryotic cells, which links the chromosome to microtubule polymers from the mitotic spindle during mitosis and meiosis and functions during cell division to pull sister chromatids apart.

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

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

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

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

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

*Appendix 2***THE FACTORIAL DESIGN FOR IDENTIFYING SEX DIFFERENCES IN THE IN VIVO MICRONUCLEUS ASSAY****The factorial design and its analysis**

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

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

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

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

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

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

References

There are many statistical textbooks which discuss the theory, design, methodology, analysis and interpretation of factorial designs ranging from the simplest two factor analyses to the more complex forms used in Design of Experiment methodology. The following is a non-exhaustive list. Some books provide worked examples of comparable designs, in some cases with code for running the analyses using various software packages.

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

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

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

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

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

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

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

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

- (6) In Part B, Chapter B.15. is deleted.
- (7) In Part B, Chapter B.16. is deleted.
- (8) In Part B, Chapter B.18. is deleted.
- (9) In Part B, Chapter B.19. is deleted.
- (10) In Part B, Chapter B.20. is deleted.
- (11) In Part B, Chapter B.24. is deleted.
- (12) In Part B, Chapter B.47. is replaced by the following:

'B.47 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

INTRODUCTION

This test method is equivalent to OECD test guideline (TG) 437 (2013). The Bovine Corneal Opacity and Permeability (BCOP) test method was evaluated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), in conjunction with the European Centre for the Validation of Alternative Methods (ECVAM) and the Japanese Center for the Validation of Alternative Methods (JaCVAM), in 2006 and 2010 (1)(2). In the first evaluation, the BCOP test method was evaluated for its usefulness to identify chemicals (substances and mixtures) inducing serious eye damage (1). In the second evaluation, the BCOP test method was evaluated for its usefulness to identify chemicals (substances and mixtures) not classified for eye irritation or serious eye damage (2). The BCOP validation database contained 113 substances and 100 mixtures in total (2)(3). From these evaluations and their peer review it was concluded that the test method can correctly identify chemicals (both substances and mixtures) inducing serious eye damage (Category 1) as well as those not requiring classification for eye irritation or serious eye damage, as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (4) and Regulation (EC) No 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) ⁽¹⁾ and it was therefore endorsed as scientifically valid for both purposes. Serious eye damage is 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. Test chemicals inducing serious eye damage are classified as UN GHS Category 1. 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 Category 1 or 2 (2A or 2B), i.e. they are referred to as UN GHS No Category. This test method includes the recommended use and limitations of the BCOP test method based on its evaluations. The main differences between the original 2009 version and the updated 2013 version of the OECD test guideline concern, but are not limited to: the use of the BCOP test method to identify chemicals not requiring classification according to UN GHS (paragraphs 2 and 7); clarifications on the applicability of the BCOP test method to the testing of alcohols, ketones and solids (paragraphs 6 and 7) and of substances and mixtures (paragraph 8); clarifications on how surfactant substances and surfactant-containing mixtures should be tested (paragraph 28); updates and clarifications regarding the positive controls (paragraphs 39 and 40); an update of the BCOP test method decision criteria (paragraph 47); an update of the study acceptance criteria (paragraph 48); an update to the test report elements (paragraph 49); an update of Appendix 1 on definitions; the addition of Appendix 2 for the predictive capacity of the BCOP test method under various classification systems; an update of Appendix 3 on the list of proficiency chemicals; and an update of Appendix 4 on the BCOP corneal holder (paragraph 1) and on the opacitometer (paragraphs 2 and 3).

⁽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, 31.12.2008, p. 1).

It is currently generally accepted that, in the foreseeable future, no single *in vitro* eye irritation test will be able to replace the *in vivo* Draize eye test to predict across the full range of irritation for different chemical classes. However, strategic combinations of several alternative test methods within a (tiered) testing strategy may be able to replace the Draize eye test (5). The Top-Down approach (5) is designed to be used when, based on existing information, a chemical is expected to have high irritancy potential, while the Bottom-Up approach (5) is designed to be used when, based on existing information, a chemical is expected not to cause sufficient eye irritation to require a classification. The BCOP test method is an *in vitro* test method that can be used under certain circumstances and with specific limitations for eye hazard classification and labeling of chemicals. While it is not considered valid as a stand-alone replacement for the *in vivo* rabbit eye test, the BCOP test method is recommended as an initial step within a testing strategy such as the Top-Down approach suggested by Scott *et al.* (5) to identify chemicals inducing serious eye damage, i.e. chemicals to be classified as UN GHS Category 1, without further testing (4). The BCOP test method is also recommended to identify chemicals that do not require classification for eye irritation or serious eye damage, as defined by the UN GHS (UN GHS No Category) (4) within a testing strategy such as the Bottom-up approach (5). However, a chemical that is not predicted as causing serious eye damage or as not classified for eye irritation/serious eye damage with the BCOP test method would require additional testing (*in vitro* and/or *in vivo*) to establish a definitive classification.

The purpose of this test method is to describe the procedures used to evaluate the eye hazard potential of a test chemical as measured by its ability to induce opacity and increased permeability in an isolated bovine cornea. Toxic effects to the cornea are measured by: (i) decreased light transmission (opacity), and (ii) increased passage of sodium fluorescein dye (permeability). The opacity and permeability assessments of the cornea following exposure to a test chemical are combined to derive an *In Vitro* Irritancy Score (IVIS), which is used to classify the irritancy level of the test chemical.

Definitions are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

This test method is based on the ICCVAM BCOP test method protocol (6)(7), which was originally developed from information obtained from the Institute for *in vitro* Sciences (IIVS) protocol and INVITTOX Protocol 124 (8). The latter represents the protocol used for the European Community-sponsored prevalidation study conducted in 1997-1998. Both of these protocols were based on the BCOP test method first reported by Gautheron *et al.* (9).

The BCOP test method can be used to identify chemicals inducing serious eye damage as defined by UN GHS, i.e. chemicals to be classified as UN GHS Category 1 (4). When used for this purpose, the BCOP test method has an overall accuracy of 79 % (150/191), a false positive rate of 25 % (32/126), and a false negative rate of 14 % (9/65), when compared to *in vivo* rabbit eye test method data classified according to the UN GHS classification system (3)

(see Appendix 2, Table 1). When test chemicals within certain chemical (i.e., alcohols, ketones) or physical (i.e., solids) classes are excluded from the database, the BCOP test method has an overall accuracy of 85 % (111/131), a false positive rate of 20 % (16/81), and a false negative rate of 8 % (4/50) for the UN GHS classification system (3). The potential shortcomings of the BCOP test method when used to identify chemicals inducing serious eye damage (UN GHS Category 1) are based on the high false positive rates for alcohols and ketones and the high false negative rate for solids observed in the validation database (1)(2)(3). However, since not all alcohols and ketones are over-predicted by the BCOP test method and some are correctly predicted as UN GHS Category 1, these two organic functional groups are not considered to be out of the applicability domain of the test method. It is up to the user of this test method to decide if a possible over-prediction of an alcohol or ketone can be accepted or if further testing should be performed in a weight-of-evidence approach. Regarding the false negative rates for solids, it should be noted that solids may lead to variable and extreme exposure conditions in the *in vivo* Draize eye irritation test, which may result in irrelevant predictions of their true irritation potential (10). It should also be noted that none of the false negatives identified in the ICCVAM validation database (2)(3), in the context of identifying chemicals inducing serious eye damage (UN GHS Category 1), resulted in $IVIS \leq 3$, which is the criterion used to identify a test chemical as a UN GHS No Category. Moreover, BCOP false negatives in this context are not critical since all test chemicals that produce an $3 < IVIS \leq 55$ would be subsequently tested with other adequately validated *in vitro* tests, or as a last option in rabbits, depending on regulatory requirements, using a sequential testing strategy in a weight-of-evidence approach. Given the fact that some solid chemicals are correctly predicted by the BCOP test method as UN GHS Category 1, this physical state is also not considered to be out of the applicability domain of the test method. Investigators could consider using this test method for all types of chemicals, whereby an $IVIS > 55$ should be accepted as indicative of a response inducing serious eye damage that should be classified as UN GHS Category 1 without further testing. However, as already mentioned, positive results obtained with alcohols or ketones should be interpreted cautiously due to potential over-prediction.

The BCOP test method can also be used to identify chemicals that do not require classification for eye irritation or serious eye damage under the UN GHS classification system (4). When used for this purpose, the BCOP test method has an overall accuracy of 69 % (135/196), a false positive rate of 69 % (61/89), and a false negative rate of 0 % (0/107), when compared to *in vivo* rabbit eye test method data classified according to the UN GHS classification system (3) (see Appendix 2, Table 2). The false positive rate obtained (*in vivo* UN GHS No Category chemicals producing an $IVIS > 3$, see paragraph 47) is considerably high, but not critical in this context since all test chemicals that produce an $3 < IVIS \leq 55$ would be subsequently tested with other adequately validated *in vitro* tests, or as a last option in rabbits, depending on regulatory requirements, using a sequential testing strategy in a weight-of-evidence approach. The BCOP test method shows no specific shortcomings for the testing of alcohols, ketones and solids when the purpose is to identify chemicals that do not require classification for eye irritation or serious eye damage (UN GHS No Category) (3). Investigators could consider using this test method for all types of chemicals, whereby a negative result ($IVIS \leq 3$) should be accepted as indicative that no classification is required (UN GHS No Category). Since the BCOP test method can only identify correctly 31 % of the chemicals that do not require classification for eye irritation or serious eye damage, this test method should not be the first choice to initiate a Bottom-Up approach (5), if other validated and accepted *in vitro* methods with similar high sensitivity but higher specificity are available.

The BCOP validation database contained 113 substances and 100 mixtures in total (2)(3). The BCOP test method is therefore considered applicable to the testing of both substances and mixtures.

The BCOP test method is not recommended for the identification of test chemicals that should be classified as irritating to eyes (UN GHS Category 2 or Category 2A) or test chemicals that should be classified as mildly irritating to eyes (UN GHS Category 2B) due to the considerable number of UN GHS Category 1 chemicals underclassified as UN GHS Category 2, 2A or 2B and UN GHS No Category chemicals overclassified as UN GHS Category 2, 2A or 2B (2)(3). For this purpose, further testing with another suitable method may be required.

All procedures with bovine eyes and bovine corneas should follow the testing facility's applicable regulations and procedures for handling animal-derived materials, which include, but are not limited to, tissues and tissue fluids. Universal laboratory precautions are recommended (11).

Whilst the BCOP test method does not consider conjunctival and iridal injuries, it addresses corneal effects, which are the major driver of classification *in vivo* when considering the UN GHS classification. The reversibility of corneal lesions cannot be evaluated *per se* in the BCOP test method. It has been proposed, based on rabbit eye studies, that an assessment of the initial depth of corneal injury may be used to identify some types of irreversible effects (12). However, further scientific knowledge is required to understand how irreversible effects not linked with initial high level injury occur. Finally, the BCOP test method does not allow for an assessment of the potential for systemic toxicity associated with ocular exposure.

This test method will be updated periodically as new information and data are considered. For example, histopathology may be potentially useful when a more complete characterisation of corneal damage is needed. As outlined in OECD Guidance Document No. 160 (13), users are encouraged to preserve corneas and prepare histopathology specimens that can be used to develop a database and decision criteria that may further improve the accuracy of this test method.

For any laboratory initially establishing this test method, the proficiency chemicals provided in Appendix 3 should be used. A laboratory can use these chemicals to demonstrate their technical competence in performing the BCOP test method prior to submitting BCOP test method data for regulatory hazard classification purposes.

PRINCIPLE OF THE TEST

The BCOP test method is an organotypic model that provides short-term maintenance of normal physiological and biochemical function of the bovine cornea *in vitro*. In this test method, damage by the test chemical is assessed by quantitative measurements of changes in corneal opacity and permeability with an opacitometer and a visible light spectrophotometer, respectively. Both measurements are used to calculate an IVIS, which is used to assign an *in vitro* irritancy hazard classification category for prediction of the *in vivo* ocular irritation potential of a test chemical (see Decision Criteria in paragraph 48).

The BCOP test method uses isolated corneas from the eyes of freshly slaughtered cattle. Corneal opacity is measured quantitatively as the amount of light transmission through the cornea. Permeability is measured quantitatively as the amount of sodium fluorescein dye that passes across the full thickness of the cornea, as detected in the medium in the posterior chamber. Test chemicals are applied to the epithelial surface of the cornea by addition to the anterior chamber of the corneal holder. Appendix 4 provides a description and a diagram of a corneal holder used in the BCOP test method. Corneal holders can be obtained commercially from different sources or can be constructed.

Source and Age of Bovine Eyes and Selection of Animal Species

Cattle sent to slaughterhouses are typically killed either for human consumption or for other commercial uses. Only healthy animals considered suitable for entry into the human food chain are used as a source of corneas for use in the BCOP test method. Because cattle have a wide range of weights, depending on breed, age, and sex, there is no recommended weight for the animal at the time of slaughter.

Variations in corneal dimensions can result when using eyes from animals of different ages. Corneas with a horizontal diameter $> 30,5$ mm and central corneal thickness (CCT) values $\geq 1\ 100$ μm are generally obtained from cattle older than eight years, while those with a horizontal diameter $< 28,5$ mm and CCT < 900 μm are generally obtained from cattle less than five years old (14). For this reason, eyes from cattle greater than 60 months old are not typically used. Eyes from cattle less than 12 months of age have not traditionally been used since the eyes are still developing and the corneal thickness and corneal diameter are considerably smaller than that reported for eyes from adult cattle. However, the use of corneas from young animals (*i.e.*, 6 to 12 months old) is permissible since there are some advantages, such as increased availability, a narrow age range, and decreased hazards related to potential worker exposure to Bovine Spongiform Encephalopathy (15). As further evaluation of the effect of corneal size or thickness on responsiveness to corrosive and irritant chemicals would be useful, users are encouraged to report the estimated age and/or weight of the animals providing the corneas used in a study.

Collection and Transport of Eyes to the Laboratory

Eyes are collected by slaughterhouse employees. To minimise mechanical and other types of damage to the eyes, the eyes should be enucleated as soon as possible after death and cooled immediately after enucleation and during transport. To prevent exposure of the eyes to potentially irritant chemicals, the slaughterhouse employees should not use detergent when rinsing the head of the animal.

Eyes should be immersed completely in cooled Hanks' Balanced Salt Solution (HBSS) in a suitably sized container, and transported to the laboratory in such a manner as to minimise deterioration and/or bacterial contamination. Because the eyes are collected during the slaughter process, they might be exposed to blood and other biological materials, including bacteria and other microorganisms. Therefore, it is important to ensure that the risk of contamination is minimised (*e.g.*, by keeping the container containing the eyes on wet ice during collection and transportation and by adding antibiotics to the HBSS used to store the eyes during transport [*e.g.* penicillin at 100 IU/ml and streptomycin at 100 $\mu\text{g/ml}$]).

The time interval between collection of the eyes and use of corneas in the BCOP test method should be minimised (typically collected and used on the same day) and should be demonstrated to not compromise the assay results. These results are based on the selection criteria for the eyes, as well as the positive and negative control responses. All eyes used in the assay should be from the same group of eyes collected on a specific day.

Selection Criteria for Eyes Used in the BCOP Test Method

The eyes, once they arrive at the laboratory, are carefully examined for defects including increased opacity, scratches, and neovascularisation. Only corneas from eyes free of such defects are to be used.

The quality of each cornea is also evaluated at later steps in the assay. Corneas that have opacity greater than seven opacity units or equivalent for the opacitometer and cornea holders used after an initial one hour equilibration period are to be discarded (NOTE: the opacitometer should be calibrated with opacity standards that are used to establish the opacity units, see Appendix 4).

Each treatment group (test chemical, concurrent negative and positive controls) consists of a minimum of three eyes. Three corneas should be used for the negative control corneas in the BCOP test method. Since all corneas are excised from the whole globe, and mounted in the corneal chambers, there is potential for artifacts from handling upon individual corneal opacity and permeability values (including negative control). Furthermore, the opacity and permeability values from the negative control corneas are used to correct the test chemical-treated and positive control-treated corneal opacity and permeability values in the IVIS calculations.

PROCEDURE

Preparation of the Eyes

Corneas, free of defects, are dissected with a 2 to 3 mm rim of sclera remaining to assist in subsequent handling, with care taken to avoid damage to the corneal epithelium and endothelium. Isolated corneas are mounted in specially designed corneal holders that consist of anterior and posterior compartments, which interface with the epithelial and endothelial sides of the cornea, respectively. Both chambers are filled to excess with pre-warmed phenol red free Eagle's Minimum Essential Medium (EMEM) (posterior chamber first), ensuring that no bubbles are formed. The device is then equilibrated at 32 ± 1 °C for at least one hour to allow the corneas to equilibrate with the medium and to achieve normal metabolic activity, to the extent possible (the approximate temperature of the corneal surface *in vivo* is 32 °C).

Following the equilibration period, fresh pre-warmed phenol red free EMEM is added to both chambers and baseline opacity readings are taken for each cornea. Any corneas that show macroscopic tissue damage (e.g. scratches, pigmentation, neovascularisation) or an opacity greater than seven opacity units or equivalent for the opacitometer and cornea holders used are discarded. A minimum of three corneas are selected as negative (or solvent) control corneas. The remaining corneas are then distributed into treatment and positive control groups.

Because the heat capacity of water is higher than that of air, water provides more stable temperature conditions for incubation. Therefore, the use a water bath for maintaining the corneal holder and its contents at 32 ± 1 °C is recommended. However, air incubators might also be used, assuming precaution to maintain temperature stability (e.g. by pre-warming of holders and media).

Application of the Test Chemical

Two different treatment protocols are used, one for liquids and surfactants (solids or liquids), and one for non-surfactant solids.

Liquids are tested undiluted. Semi-solids, creams, and waxes are typically tested as liquids. Neat surfactant substances are tested at a concentration of 10 % w/v in a 0,9 % sodium chloride solution, distilled water, or other solvent that has been demonstrated to have no adverse effects on the test system. Appropriate justification should be provided for alternative dilution concentrations. Mixtures containing surfactants may be tested undiluted or diluted to an appropriate concentration depending on the relevant exposure scenario *in vivo*. Appropriate justification should be provided for the concentration tested. Corneas are exposed to liquids and surfactants for 10 minutes. Use of other exposure times should be accompanied by adequate scientific rationale. Please see Appendix 1 for a definition of surfactant and surfactant-containing mixture.

Non-surfactant solids are typically tested as solutions or suspensions at 20 % w/v concentration in a 0,9 % sodium chloride solution, distilled water, or other solvent that has been demonstrated to have no adverse effects on the test system. In certain circumstances and with proper scientific justification, solids may also be tested neat by direct application onto the corneal surface using the open chamber method (see paragraph 32). Corneas are exposed to solids for four hours, but as with liquids and surfactants, alternative exposure times may be used with appropriate scientific rationale.

Different treatment methods can be used, depending on the physical nature and chemical characteristics (e.g. solids, liquids, viscous vs. non-viscous liquids) of the test chemical. The critical factor is ensuring that the test chemical adequately covers the epithelial surface and that it is adequately removed during the rinsing steps. A closed-chamber method is typically used for non-viscous to slightly viscous liquid test chemicals, while an open-chamber method is typically used for semi-viscous and viscous liquid test chemicals and for neat solids.

In the closed-chamber method, sufficient test chemical (750 µl) to cover the epithelial side of the cornea is introduced into the anterior chamber through the dosing holes on the top surface of the chamber, and the holes are subsequently sealed with the chamber plugs during the exposure. It is important to ensure that each cornea is exposed to a test chemical for the appropriate time interval.

In the open-chamber method, the window-locking ring and glass window from the anterior chamber are removed prior to treatment. The control or test chemical (750 µl, or enough test chemical to completely cover the cornea) is applied directly to the epithelial surface of the cornea using a micro-pipet. If a test chemical is difficult to pipet, the test chemical can be pressure-loaded into a positive displacement pipet to aid in dosing. The pipet tip of the positive displacement pipet is inserted into the dispensing tip of the syringe so that the material can be loaded into the displacement tip under pressure. Simultaneously, the syringe plunger is depressed as the pipet piston is drawn upwards. If air bubbles appear in the pipet tip, the test chemical is removed (expelled) and the process repeated until the tip is filled without air bubbles. If necessary, a normal syringe (without a needle) can be used since it permits measuring an accurate volume of test chemical and an easier application to the epithelial surface of the cornea. After dosing, the glass window is replaced on the anterior chamber to recreate a closed system.

Post-Exposure Incubation

After the exposure period, the test chemical, the negative control, or the positive control chemical is removed from the anterior chamber and the epithelium washed at least three times (or until no visual evidence of test chemical can be observed) with EMEM (containing phenol red). Phenol red-containing medium is used for rinsing since a colour change in the phenol red may be monitored to determine the effectiveness of rinsing acidic or alkaline test chemicals. The corneas are washed more than three times if the phenol red is still discoloured (yellow or purple), or the test chemical is still visible. Once the medium is free of test chemical, the corneas are given a final rinse with EMEM (without phenol red). The EMEM (without phenol red) is used as a final rinse to ensure removal of the phenol red from the anterior chamber prior to the opacity measurement. The anterior chamber is then refilled with fresh EMEM without phenol red.

For liquids or surfactants, after rinsing, the corneas are incubated for an additional two hours at 32 ± 1 °C. Longer post-exposure time may be useful in certain circumstances and could be considered on a case-by-case basis. Corneas treated with solids are rinsed thoroughly at the end of the four-hour exposure period, but do not require further incubation.

At the end of the post-exposure incubation period for liquids and surfactants and at the end of the four-hour exposure period for non-surfactant solids, the opacity and permeability of each cornea are recorded. Also, each cornea is observed visually and pertinent observations recorded (e.g., tissue peeling, residual test chemical, non-uniform opacity patterns). These observations could be important as they may be reflected by variations in the opacimeter readings.

Control Chemicals

Concurrent negative or solvent/vehicle controls and positive controls are included in each experiment.

When testing a liquid substance at 100 %, a concurrent negative control (e.g. 0,9 % sodium chloride solution or distilled water) is included in the BCOP test method so that nonspecific changes in the test system can be detected and to provide a baseline for the assay endpoints. It also ensures that the assay conditions do not inappropriately result in an irritant response.

When testing a diluted liquid, surfactant, or solid, a concurrent solvent/vehicle control group is included in the BCOP test method so that nonspecific changes in the test system can be detected and to provide a baseline for the assay endpoints. Only a solvent/vehicle that has been demonstrated to have no adverse effects on the test system can be used.

A chemical known to induce a positive response is included as a concurrent positive control in each experiment to verify the integrity of the test system and its correct conduct. However, to ensure that variability in the positive control response across time can be assessed, the magnitude of irritant response should not be excessive.

Examples of positive controls for liquid test chemicals are 100 % ethanol or 100 % dimethylformamide. An example of a positive control for solid test chemicals is 20 % w/v imidazole in 0,9 % sodium chloride solution.

Benchmark chemicals 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.

Endpoints Measured

Opacity is determined by the amount of light transmission through the cornea. Corneal opacity is measured quantitatively with the aid of an opacitometer, resulting in opacity values measured on a continuous scale.

Permeability is determined by the amount of sodium fluorescein dye that penetrates all corneal cell layers (*i.e.*, the epithelium on the outer cornea surface through the endothelium on the inner cornea surface). One ml sodium fluorescein solution (4 or 5 mg/ml when testing liquids and surfactants or non-surfactant solids, respectively) is added to the anterior chamber of the corneal holder, which interfaces with the epithelial side of the cornea, while the posterior chamber, which interfaces with the endothelial side of the cornea, is filled with fresh EMEM. The holder is then incubated in a horizontal position for 90 ± 5 min at 32 ± 1 °C. The amount of sodium fluorescein that crosses into the posterior chamber is quantitatively measured with the aid of UV/VIS spectrophotometry. Spectrophotometric measurements evaluated at 490 nm are recorded as optical density (OD_{490}) or absorbance values, which are measured on a continuous scale. The fluorescein permeability values are determined using OD_{490} values based upon a visible light spectrophotometer using a standard 1 cm path length.

Alternatively, a 96-well microtiter plate reader may be used provided that; (i) the linear range of the plate reader for determining fluorescein OD_{490} values can be established; and (ii), the correct volume of fluorescein samples are used in the 96-well plate to result in OD_{490} values equivalent to the standard 1 cm path length (this could require a completely full well [usually 360 µl]).

DATA AND REPORTING

Data Evaluation

Once the opacity and mean permeability (OD_{490}) values have been corrected for background opacity and the negative control permeability OD_{490} values, the mean opacity and permeability OD_{490} values for each treatment group should be combined in an empirically-derived formula to calculate an *in vitro* irritancy score (IVIS) for each treatment group as follows:

$$\text{IVIS} = \text{mean opacity value} + (15 \times \text{mean permeability } OD_{490} \text{ value})$$

Sina *et al.* (16) reported that this formula was derived during in-house and inter-laboratory studies. The data generated for a series of 36 compounds in a multi-laboratory study were subjected to a multivariate analysis to determine the equation of best fit between *in vivo* and *in vitro* data. Scientists at two separate companies performed this analysis and derived nearly identical equations.

The opacity and permeability values should also be evaluated independently to determine whether a test chemical induced corrosivity or severe irritation through only one of the two endpoints (see Decision Criteria).

Decision Criteria

The IVIS cut-off values for identifying test chemicals as inducing serious eye damage (UN GHS Category 1) and test chemicals not requiring classification for eye irritation or serious eye damage (UN GHS No Category) are given hereafter:

IVIS	UN GHS
≤ 3	No Category
$> 3; \leq 55$	No prediction can be made
> 55	Category 1

Study Acceptance Criteria

A test is considered acceptable if the positive control gives an IVIS that falls within two standard deviations of the current historical mean, which is to be updated at least 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). The negative or solvent/vehicle control responses should result in opacity and permeability values that are less than the established upper limits for background opacity and permeability values for bovine corneas treated with the respective negative or solvent/vehicle control. A single testing run composed of at least three corneas should be sufficient for a test chemical when the resulting classification is unequivocal. However, in cases of borderline results in the first testing run, a second testing run should be considered (but not necessarily required), as well as a third one in case of discordant mean IVIS results between the first two testing runs. In this context, a result in the first testing run is considered borderline if the predictions from the 3 corneas were non-concordant, such that:

- 2 of the 3 corneas gave discordant predictions from the mean of all 3 corneas, OR,
- 1 of the 3 corneas gave a discordant prediction from the mean of all 3 corneas, AND the discordant result was > 10 IVIS units from the cut-off threshold of 55.
- If the repeat testing run corroborates the prediction of the initial testing run (based upon the mean IVIS value), then a final decision can be taken without further testing. If the repeat testing run results in a non-concordant prediction from the initial testing run (based upon the mean IVIS value), then a third and final testing run should be conducted to resolve equivocal predictions, and to classify the test chemical. It may be permissible to waive further testing for classification and labeling in the event any testing run results in a UN GHS Category 1 prediction.

Test Report

The test report should include the following information, if relevant to the conduct of the study:

Test and Control Chemicals

- Chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known; the CAS Registry Number (RN), if known;

- Purity and composition of the test/control chemical (in percentage(s) by weight), to the extent this information is available;
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study;
- Treatment of the test/control chemicals prior to testing, if applicable (e.g. warming, grinding);
- Stability, if known.

Information Concerning the Sponsor and the Test Facility

- Name and address of the sponsor, test facility and study director.

Test Method Conditions

- Opacitometer used (e.g. model and specifications) and instrument settings;
- Calibration information for devices used for measuring opacity and permeability (e.g. opacitometer and spectrophotometer) to ensure linearity of measurements;
- Type of corneal holders used (e.g. model and specifications);
- Description of other equipment used;
- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time (e.g. periodic testing of proficiency chemicals).

Criteria for an Acceptable Test

- Acceptable concurrent positive and negative control ranges based on historical data;
- If applicable, acceptable concurrent benchmark control ranges based on historical data.

Eyes Collection and Preparation

- Identification of the source of the eyes (i.e., the facility from which they were collected);
- Corneal diameter as a measure of age of the source animal and suitability for the assay;
- Storage and transport conditions of eyes (e.g. date and time of eye collection, time interval prior to initiating testing, transport media and temperature conditions, any antibiotics used);
- Preparation & mounting of the bovine corneas including statements regarding their quality, temperature of corneal holders, and criteria for selection of corneas used for testing.

Test Procedure

- Number of replicates used;
- Identity of the negative and positive controls used (if applicable, also the solvent and benchmark controls);
- Test chemical concentration(s), application, exposure time and post-exposure incubation time used;
- Description of evaluation and decision criteria used;
- Description of study acceptance criteria used;
- Description of any modifications of the test procedure;
- Description of decision criteria used.

Results

- Tabulation of data from individual test samples (e.g. opacity and OD490 values and calculated IVIS for the test chemical and the positive, negative, and benchmark controls [if included], reported in tabular form, including data from replicate repeat experiments as appropriate, and means \pm the standard deviation for each experiment);
- Description of other effects observed;
- The derived *in vitro* UN GHS classification, if applicable.

Discussion of the Results

Conclusion

LITERATURE:

- (1) ICCVAM (2006). Test Method Evaluation Report — *In Vitro* Ocular Toxicity Test Methods for Identifying Ocular Severe Irritants and Corrosives. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No.: 07-4517. Available: http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_tmter.htm.
- (2) ICCVAM (2010). ICCVAM Test Method Evaluation Report: Current Validation Status of *In Vitro* Test Methods Proposed for Identifying Eye Injury Hazard Potential of Chemicals and Products. NIH Publication No.10-7553. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available: <http://iccvam.niehs.nih.gov/methods/ocutox/MildMod-TMER.htm>.
- (3) OECD (2013). Streamlined Summary Document supporting the Test Guideline 437 for eye irritation/corrosion. Series on Testing and Assessment, No.189, OECD, Paris.
- (4) UN (2011). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), ST/SG/AC.10/30 Rev 4, New York and Geneva: United Nations. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev04/04files_e.html.
- (5) 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., and 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.
- (6) ICCVAM (2006). ICCVAM Recommended BCOP Test Method Protocol. In: ICCVAM Test Method Evaluation Report — *in vitro* Ocular Toxicity Test Methods for Identifying Ocular Severe Irritants and Corrosives. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No.: 07-4517. Available: http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_tmter.htm.
- (7) ICCVAM (2010). ICCVAM Recommended BCOP Test Method Protocol. In: ICCVAM Test Method Evaluation Report — Current Validation Status of *In Vitro* Test Methods Proposed for Identifying Eye Injury Hazard Potential of Chemicals and Products. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No.: 10-7553A. Available: <http://iccvam.niehs.nih.gov/methods/ocutox/MildMod-TMER.htm>.
- (8) INVITTOX (1999). Protocol 124: Bovine Corneal Opacity and Permeability Assay — SOP of Microbiological Associates Ltd. Ispra, Italy: European Centre for the Validation of Alternative Methods (ECVAM).
- (9) Gautheron, P., Dukic, M., Alix, D. and Sina, J.F. (1992). Bovine corneal opacity and permeability test: An *in vitro* assay of ocular irritancy. *Fundam. Appl. Toxicol.* 18:442-449.

- (10) Prinsen, M.K. (2006). The Draize Eye Test and *in vitro* alternatives; a left-handed marriage? *Toxicol. in Vitro* 20:78-81.
- (11) Siegel, J.D., Rhinehart, E., Jackson, M., Chiarello, L., and the Healthcare Infection Control Practices Advisory Committee (2007). Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Healthcare Settings. Available: [<http://www.cdc.gov/ncidod/dhqp/pdf>].
- (12) Maurer, J.K., Parker, R.D. and 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.
- (13) OECD (2011). Guidance Document on The Bovine Corneal Opacity and Permeability (BCOP) and Isolated Chicken Eye (ICE) Test Methods: Collection of Tissues for Histological Evaluation and Collection of Data on Non-severe Irritants. Series on Testing and Assessment, No. 160. Adopted October 25, 2011. Paris: Organisation for Economic Co-operation and Development.
- (14) Doughty, M.J., Petrou, S. and Macmillan, H. (1995). Anatomy and morphology of the cornea of bovine eyes from a slaughterhouse. *Can. J. Zool.* 73:2159-2165.
- (15) Collee, J. and Bradley, R. (1997). BSE: A decade on — Part I. *The Lancet* 349: 636-641.
- (16) Sina, J.F., Galer, D.M., Sussman, R.S., Gautheron, P.D., Sargent, E.V., Leong, B., Shah, P.V., Curren, R.D., and Miller, K. (1995). A collaborative evaluation of seven alternatives to the Draize eye irritation test using pharmaceutical intermediates. *Fundam. Appl. Toxicol.* 26:20-31.
- (17) Chapter B.5 of this Annex, Acute eye irritation/corrosion.
- (18) ICCVAM (2006). Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: Bovine Corneal Opacity and Permeability Test Method. NIH Publication No.: 06-4512. Research Triangle Park: National Toxicology Program. Available: [http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_brd_bcop.html].
- (19) OECD (1998). Series on Good Laboratory Practice and Compliance Monitoring. No. 1: OECD Principles on Good Laboratory Practice (revised in 1997).

Available at: http://www.oecd.org/document/63/0,3343,en_2649_34381_2346175_1_1_1_1,00.html

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.

Benchmark chemical: A chemical used as a standard for comparison to a test chemical. A benchmark chemical should have the following properties; (i) a consistent and reliable source(s); (ii) structural and functional similarity to the class of chemicals 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: step-wise approach used for a 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 a mixture.

Cornea: The transparent part of the front of the eyeball that covers the iris and pupil and admits light to the interior.

Corneal opacity: Measurement of the extent of opaqueness of the cornea following exposure to a test chemical. Increased corneal opacity is indicative of damage to the cornea. Opacity can be evaluated subjectively as done in the Draize rabbit eye test, or objectively with an instrument such as an “opacitometer”.

Corneal permeability: Quantitative measurement of damage to the corneal epithelium by a determination of the amount of sodium fluorescein dye that passes through all corneal cell layers.

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 Category 2” (4).

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.

In Vitro Irritancy Score (IVIS): An empirically-derived formula used in the BCOP test method whereby the mean opacity and mean permeability values for each treatment group are combined into a single *in vitro* score for each treatment group. The IVIS = mean opacity value + (15 × mean permeability value).

Irreversible effects on the eye: See “Serious eye damage”.

Mixture: A mixture or a solution composed of two or more substances in which they do not react (4)

Negative 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.

Not Classified: Chemicals that are not classified for Eye irritation (UN GHS Category 2, 2A, or 2B) or Serious eye damage (UN GHS Category 1). Interchangeable with “UN GHS No Category”.

Opacitometer: An instrument used to measure “corneal opacity” by quantitatively evaluating light transmission through the cornea. The typical instrument has two compartments, each with its own light source and photocell. One compartment is used for the treated cornea, while the other is used to calibrate and zero the instrument. Light from a halogen lamp is sent through a control compartment (empty chamber without windows or liquid) to a photocell and compared to the light sent through the experimental compartment, which houses the chamber containing the cornea, to a photocell. The difference in light transmission from the photocells is compared and a numeric opacity value is presented on a digital display.

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.

Reversible effects on the eye: See “Eye irritation”.

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.

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 Category 1” (4).

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 samples 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 negative control, this sample also demonstrates whether the solvent or vehicle interacts with the test system.

Substance: Chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the 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 (4).

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.

Surfactant-containing mixture: In the context of this test method, it is a mixture containing one or more surfactants at a final concentration of > 5 %.

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 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.

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 (4).

UN GHS Category 1: See “Serious eye damage”.

UN GHS Category 2: See “Eye irritation”.

UN GHS No Category: Chemicals that do not meet the requirements for classification as UN GHS Category 1 or 2 (2A or 2B). Interchangeable with “Not Classified”.

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.

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 chemical.

Appendix 2

PREDICTIVE CAPACITY OF THE BCOP TEST METHOD

Table 1

Predictive Capacity of BCOP for identifying chemicals inducing serious eye damage [UN GHS/EU CLP Cat 1 vs Not Cat 1 (Cat 2 + No Cat); US EPA Cat I vs Not Cat I (Cat II + Cat III + Cat IV)]

Classifi- cation System	No.	Accuracy		Sensitivity		False Negatives		Specificity		False Positives	
		%	No.	%	No.	%	No.	%	No.	%	No.
UN GHS EU CLP	191	78,53	150/191	86,15	56/65	13,85	9/65	74,60	94/126	25,40	32/126
US EPA	190	78,95	150/190	85,71	54/63	14,29	9/63	75,59	96/127	24,41	31/127

Table 2

Predictive Capacity of BCOP for identifying chemicals not requiring classification for eye irritation or serious eye damage ("non-irritants") [UN GHS/EU CLP No Cat vs Not No Cat (Cat 1 + Cat 2); US EPA Cat IV vs Not Cat IV (Cat I + Cat II + Cat III)]

Classifi- cation System	No.	Accuracy		Sensitivity		False Negatives		Specificity		False Positives	
		%	No.	%	No.	%	No.	%	No.	%	No.
UN GHS EU CLP	196	68,88	135/196	100	107/107	0	0/107	31,46	28/89	68,54	61/89
US EPA	190	82,11	156/190	93,15	136/146	6,85	10/146	45,45	20/44	54,55	24/44

Appendix 3

PROFICIENCY CHEMICALS FOR THE BCOP TEST METHOD

Prior to routine use of this test method, laboratories should demonstrate technical proficiency by correctly identifying the eye hazard classification of the 13 chemicals recommended in Table 1. These chemicals were selected to represent the range of responses for eye hazards based on results in the *in vivo* rabbit eye test (TG 405) (17) and the UN GHS classification system (*i.e.*, Categories 1, 2A, 2B, or Not Classified) (4). Other selection criteria were that chemicals are commercially available, that there are high quality *in vivo* reference data available, and that there are high quality *in vitro* data available from the BCOP test method. Reference data are available in the Streamlined Summary Document (3) and in the ICCVAM Background Review Document for the BCOP test method (2)(18).

Table 1

Recommended chemicals for demonstrating technical proficiency with the BCOP test method

Chemical	CASRN	Chemical Class ⁽¹⁾	Physical Form	<i>In Vivo</i> Classification ⁽²⁾	BCOP Classification
Benzalkonium chloride (5 %)	8001-54-5	Onium compound	Liquid	Category 1	Category 1
Chlorhexidine	55-56-1	Amine, Amidine	Solid	Category 1	Category 1
Dibenzoyl-L- tartaric acid	2743-38-6	Carboxylic acid, Ester	Solid	Category 1	Category 1
Imidazole	288-32-4	Heterocyclic	Solid	Category 1	Category 1
Trichloroacetic acid (30 %)	76-03-9	Carboxylic acid	Liquid	Category 1	Category 1
2,6-Dichlorobenzoyl chloride	4659-45-4	Acyl halide	Liquid	Category 2A	No accurate/reliable prediction can be made
Ethyl-2-methylacetate	609-14-3	Ketone, Ester	Liquid	Category 2B	No accurate/reliable prediction can be made
Ammonium nitrate	6484-52-2	Inorganic salt	Solid	Category 2 ⁽³⁾	No accurate/reliable prediction can be made
EDTA, di-potassium salt	25102-12-9	Amine, Carboxylic acid (salt)	Solid	Not Classified	Not Classified
Tween 20	9005-64-5	Ester, Polyether	Liquid	Not Classified	Not Classified

Chemical	CASRN	Chemical Class ⁽¹⁾	Physical Form	<i>In Vivo</i> Classification ⁽²⁾	BCOP Classification
2-Mercaptopyrimidine	1450-85-7	Acyl halide	Solid	Not Classified	Not Classified
Phenylbutazone	50-33-9	Heterocyclic	Solid	Not Classified	Not Classified
Polyoxyethylene 23 lauryl ether (BRIJ-35) (10 %)	9002-92-0	Alcohol	Liquid	Not Classified	Not Classified

Abbreviations: CASRN = Chemical Abstracts Service Registry Number.

⁽¹⁾ Chemical classes were assigned to each test chemical using a standard classification scheme, based on the National Library of Medicine Medical Subject Headings (MeSH) classification system (available at <http://www.nlm.nih.gov/mesh>).

⁽²⁾ Based on results from the *in vivo* rabbit eye test (OECD TG 405) (17) and using the UN GHS (4).

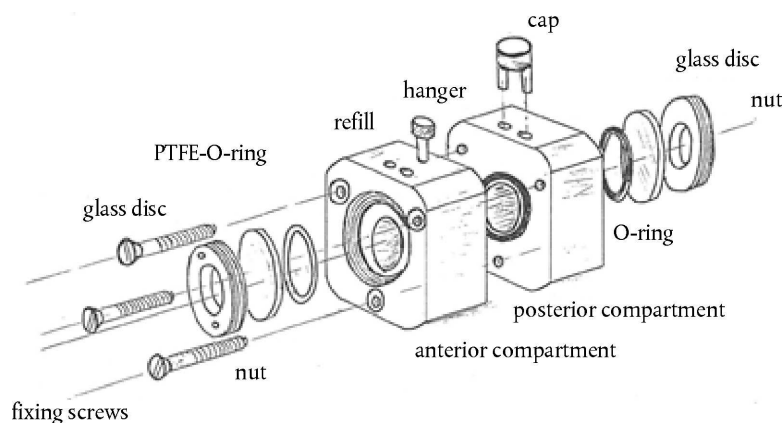
⁽³⁾ Classification as 2A or 2B depends on the interpretation of the UN GHS criterion for distinguishing between these two categories, i.e. 1 out of 3 vs. 2 out of 3 animals with effects at day 7 necessary to generate a Category 2A classification. The *in vivo* study included 3 animals. All endpoints apart from conjunctiva 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 conjunctiva redness score of 1 (at day 7) that fully recovered at day 10.

Appendix 4

THE BCOP CORNEAL HOLDER

The BCOP corneal holders are made of an inert material (e.g. polypropylene). The holders are comprised of two halves (an anterior and posterior chamber), and have two similar cylindrical internal chambers. Each chamber is designed to hold a volume of about 5 ml and terminates in a glass window, through which opacity measurements are recorded. Each of the inner chambers is 1,7 cm in diameter and 2,2 cm in depth ⁽¹⁾. An o-ring located on the posterior chamber is used to prevent leaks. The corneas are placed endothelial side down on the o-ring of the posterior chambers and the anterior chambers are placed on the epithelial side of the corneas. The chambers are maintained in place by three stainless steel screws located on the outer edges of the chamber. The end of each chamber houses a glass window, which can be removed for easy access to the cornea. An o-ring is also located between the glass window and the chamber to prevent leaks. Two holes on the top of each chamber permit introduction and removal of medium and test chemicals. They are closed with rubber caps during the treatment and incubation periods. The light transmission through corneal holders can potentially change as the effects of wear and tear or accumulation of specific chemical residues on the internal chamber bores or on the glass windows may affect light scatter or reflectance. The consequence could be increases or decreases in baseline light transmission (and conversely the baseline opacity readings) through the corneal holders, and may be evident as notable changes in the expected baseline initial corneal opacity measurements in individual chambers (i.e., the initial corneal opacity values in specific individual corneal holders may routinely differ by more than 2 or 3 opacity units from the expected baseline values). Each laboratory should consider establishing a program for evaluating for changes in the light transmission through the corneal holders, depending upon the nature of the chemistries tested and the frequency of use of the chambers. To establish baseline values, corneal holders may be checked before routine use by measuring the baseline opacity values (or light transmission) of chambers filled with complete medium, without corneas. The corneal holders are then periodically checked for changes in light transmission during periods of use. Each laboratory can establish the frequency for checking the corneal holders, based upon the chemicals tested, the frequency of use, and observations of changes in the baseline corneal opacity values. If notable changes in the light transmission through the corneal holders are observed, appropriate cleaning and/or polishing procedures of the interior surface of the cornea holders or replacement have to be considered.

Corneal holder: exploded diagramme



⁽¹⁾ The dimensions provided are based on a corneal holder that is used for cows ranging in age from 12 to 60 months old. In the event that animals 6 to 12 months are being used, the holder would instead need to be designed such that each chamber holds a volume of 4 mL, and each of the inner chambers is 1,5 cm in diameter and 2,2 cm in depth. With any newly designed corneal holder, it is very important that the ratio of exposed corneal surface area to posterior chamber volume should be the same as the ratio in the traditional corneal holder. This is necessary to assure that permeability values are correctly determined for the calculation of the IVIS by the proposed formula

Appendix 5

THE OPACITOMETER

The opacitometer is a light transmission measuring device. For example, for the OP-KIT equipment from Electro Design (Riom, France) used in the validation of the BCOP test method, light from a halogen lamp is sent through a control compartment (empty chamber without windows or liquid) to a photocell and compared to the light sent through the experimental compartment, which houses the chamber containing the cornea, to a photocell. The difference in light transmission from the photocells is compared and a numeric opacity value is presented on a digital display. The opacity units are established. Other types of opacitometers with a different setup (e.g., not requiring the parallel measurements of the control and experimental compartments) may be used if proven to give similar results to the validated equipment.

The opacitometer should provide a linear response through a range of opacity readings covering the cut-offs used for the different classifications described by the Prediction Model (i.e., up to the cut-off determining corrosiveness/severe irritancy). To ensure linear and accurate readings up to 75-80 opacity units, it is necessary to calibrate the opacitometer using a series of calibrators. Calibrators are placed into the calibration chamber (a corneal chamber designed to hold the calibrators) and read on the opacitometer. The calibration chamber is designed to hold the calibrators at approximately the same distance between the light and photocell that the corneas would be placed during the opacity measurements. Reference values and initial set point depend on the type of equipment used. Linearity of opacity measurements should be ensured by appropriate (instrument specific) procedures. For example, for the OP-KIT equipment from Electro Design (Riom, France), the opacitometer is first calibrated to 0 opacity units using the calibration chamber without a calibrator. Three different calibrators are then placed into the calibration chamber one by one and the opacities are measured. Calibrators 1, 2 and 3 should result in opacity readings equal to their set values of 75, 150, and 225 opacity units, respectively, $\pm 5\%$.

(13) In Part B, Chapter B.48 is replaced by the following:

'B.48 Isolated Chicken Eye Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage

INTRODUCTION

This test method is equivalent to OECD test guideline (TG) 438 (2013). The Isolated Chicken Eye (ICE) test method was evaluated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), in conjunction with the European Centre for the Validation of Alternative Methods (ECVAM) and the Japanese Centre for the Validation of Alternative Methods (JaCVAM), in 2006 and 2010 (1) (2) (3). In the first evaluation, the ICE was endorsed as a scientifically valid test method for use as a screening test to identify chemicals (substances and mixtures) inducing serious eye damage (Category 1) as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (1) (2) (4) and Regulation (EC) No 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) ⁽¹⁾. In the second evaluation, the ICE test method was evaluated for use as a screening test to identify chemicals not classified for eye irritation or serious eye damage as defined by UN GHS (3) (4). The results from the validation study and the peer review panel recommendations maintained the original recommendation for using the ICE for classification of chemicals inducing serious eye damage (UN GHS Category 1), as the available database remained unchanged since the original ICCVAM validation. At that stage, no further recommendations for an expansion of the ICE applicability domain to also include other categories were suggested. A re-evaluation of the *in vitro* and *in vivo* dataset used in the validation

⁽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, 31.12.2008, p. 1).

study was made with the focus of evaluating the usefulness of the ICE to identify chemicals not requiring classification for eye irritation or serious eye damage (5). This re-evaluation concluded that the ICE test method can also be used to identify chemicals not requiring classification for eye irritation and serious eye damage as defined by the UN GHS (4) (5). This test method includes the recommended uses and limitations of the ICE test method based on these evaluations. The main differences between the original 2009 version and the updated 2013 version of the OECD test guideline include, but are not limited to, the use of the ICE test method to identify chemicals not requiring classification according to the UN GHS Classification System, an update to the test report elements, an update of Appendix 1 on definitions, and an update to Appendix 2 on the proficiency chemicals.

It is currently generally accepted that, in the foreseeable future, no single *in vitro* eye irritation test will be able to replace the *in vivo* Draize eye test to predict across the full range of irritation for different chemical classes. However, strategic combinations of several alternative test methods within a (tiered) testing strategy may be able to replace the Draize eye test (6). The Top-Down approach (7) is designed to be used when, based on existing information, a chemical is expected to have high irritancy potential, while the Bottom-Up approach (7) is designed to be used when, based on existing information, a chemical is expected not to cause sufficient eye irritation to require a classification. The ICE test method is an *in vitro* test method that can be used, under certain circumstances and with specific limitations as described in paragraphs 8 to 10 for eye hazard classification and labelling of chemicals. While it is not considered valid as a stand-alone replacement for the *in vivo* rabbit eye test, the ICE test method is recommended as an initial step within a testing strategy such as the Top-Down approach suggested by Scott *et al.* (7) to identify chemicals inducing serious eye damage, i.e., chemicals to be classified as UN GHS Category 1 without further testing (4). The ICE test method is also recommended to identify chemicals that do not require classification for eye irritation or serious eye damage as defined by the UN GHS (No Category, NC) (4), and may therefore be used as an initial step within a Bottom-Up testing strategy approach (7). However, a chemical that is not predicted as causing serious eye damage or as not classified for eye irritation/serious eye damage with the ICE test method would require additional testing (*in vitro* and/or *in vivo*) to establish a definitive classification. Furthermore, the appropriate regulatory authorities should be consulted before using the ICE in a bottom up approach under other classification schemes than the UN GHS.

The purpose of this test method is to describe the procedures used to evaluate the eye hazard potential of a test chemical as measured by its ability to induce or not toxicity in an enucleated chicken eye. Toxic effects to the cornea are measured by (i) a qualitative assessment of opacity, (ii) a qualitative assessment of damage to epithelium based on application of fluorescein to the eye (fluorescein retention), (iii) a quantitative measurement of increased thickness (swelling), and (iv) a qualitative evaluation of macroscopic morphological damage to the surface. The corneal opacity, swelling, and damage assessments following exposure to a test chemical are assessed individually and then combined to derive an Eye Irritancy Classification.

Definitions are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

This test method is based on the protocol suggested in the OECD Guidance Document 160 (8), which was developed following the ICCVAM international validation study (1) (3) (9), with contributions from the European Centre for the Validation of Alternative Methods, the Japanese Center for the Validation of Alternative Methods, and TNO Quality of Life Department of Toxicology and Applied Pharmacology (Netherlands). The protocol is based on information obtained from published protocols, as well as the current protocol used by TNO (10) (11) (12) (13) (14).

A wide range of chemicals has been tested in the validation underlying this test method and the empirical database of the validation study amounted to 152 chemicals including 72 substances and 80 mixtures (5). The test method is applicable to solids, liquids, emulsions and gels. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Gases and aerosols have not been assessed yet in a validation study.

The ICE test method can be used to identify chemicals inducing serious eye damage, i.e., chemicals to be classified as UN GHS Category 1 (4). When used for this purpose, the identified limitations for the ICE test method are based on the high false positive rates for alcohols and the high false negative rates for solids and surfactants (1) (3) (9). However, false negative rates in this context (UN GHS Category 1 identified as not being UN GHS Category 1) are not critical since all test chemicals that come out negative would be subsequently tested with other adequately validated *in vitro* test(s), or as a last option in rabbits, depending on regulatory requirements, using a sequential testing strategy in a weight-of-evidence approach. It should be noted that solids may lead to variable and extreme exposure conditions in the *in vivo* Draize eye irritation test, which may result in irrelevant predictions of their true irritation potential (15). Investigators could consider using this test method for all types of chemicals, whereby a positive result should be accepted as indicative of serious eye damage, i.e., UN GHS Category 1 classification without further testing. However, positive results obtained with alcohols should be interpreted cautiously due to risk of over-prediction.

When used to identify chemicals inducing serious eye damage (UN GHS Category 1), the ICE test method has an overall accuracy of 86 % (120/140), a false positive rate of 6 % (7/113) and a false negative rate of 48 % (13/27) when compared to *in vivo* rabbit eye test method data classified according to the UN GHS classification system (4) (5).

The ICE test method can also be used to identify chemicals that do not require classification for eye irritation or serious eye damage under the UN GHS classification system (4). The appropriate regulatory authorities should be consulted before using the ICE in a bottom up approach under other classification schemes. This test method can be used for all types of chemicals, whereby a negative result could be accepted for not classifying a chemical for eye irritation and serious eye damage. However, on the basis of one result from the validation database, anti-fouling organic solvent-containing paints may be under-predicted (5).

When used to identify chemicals that do not require classification for eye irritation and serious eye damage, the ICE test method has an overall accuracy of 82 % (125/152), a false positive rate of 33 % (26/79), and a false negative rate of 1 % (1/73), when compared to *in vivo* rabbit eye test method data classified according to the UN GHS (4) (5). When test chemicals within certain classes (i.e., anti-fouling organic solvent containing paints) are excluded from the database, the accuracy of the ICE test method is 83 % (123/149), the false positive rate 33 % (26/78), and the false negative rate of 0 % (0/71) for the UN GHS classification system (4) (5).

The ICE test method is not recommended for the identification of test chemicals that should be classified as irritating to eyes (i.e., UN GHS Category 2 or Category 2A) or test chemicals that should be classified as mildly irritating to eyes (UN GHS Category 2B) due to the considerable number of UN GHS Category 1 chemicals underclassified as UN GHS Category 2, 2A or 2B and UN GHS No Category chemicals overclassified as UN GHS Category 2, 2A or 2B. For this purpose, further testing with another suitable method may be required.

All procedures with chicken eyes should follow the test facility's applicable regulations and procedures for handling of human or animal-derived materials, which include, but are not limited to, tissues and tissue fluids. Universal laboratory precautions are recommended (16).

Whilst the ICE test method does not consider conjunctival and iridal injuries as evaluated in the rabbit ocular irritancy test method, it addresses corneal effects which are the major driver of classification *in vivo* when considering the UN GHS Classification. Also, although the reversibility of corneal lesions cannot be evaluated *per se* in the ICE test method, it has been proposed, based on rabbit eye studies, that an assessment of the initial depth of corneal injury may be used to identify some types of irreversible effects (17). In particular, further scientific knowledge is required to understand how irreversible effects not linked with initial high level injury occur. Finally, the ICE test method does not allow for an assessment of the potential for systemic toxicity associated with ocular exposure.

This test method will be updated periodically as new information and data are considered. For example, histopathology may be potentially useful when a more complete characterisation of corneal damage is needed. To evaluate this possibility, users are encouraged to preserve eyes and prepare histopathology specimens that can be used to develop a database and decision criteria that may further improve the accuracy of this test method. The OECD has developed a Guidance Document on the use of *in vitro* ocular toxicity test methods, which includes detailed procedures on the collection of histopathology specimens and information on where to submit specimens and/or histopathology data (8).

For any laboratory initially establishing this assay, the proficiency chemicals provided in Appendix 2 should be used. A laboratory can use these chemicals to demonstrate their technical competence in performing the ICE test method prior to submitting ICE data for regulatory hazard classification purposes.

PRINCIPLE OF THE TEST

The ICE test method is an organotypic model that provides short-term maintenance of the chicken eye *in vitro*. In this test method, damage by the test chemical is assessed by determination of corneal swelling, opacity, and fluorescein retention. While the latter two parameters involve a qualitative assessment, analysis of corneal swelling provides for a quantitative assessment. Each measurement is either converted into a quantitative score used to calculate an overall Irritation Index, or assigned a qualitative categorisation that is used to assign an *in vitro* ocular hazard classification, either as UN GHS Category 1 or as UN GHS non-classified. Either of these outcomes can then be used to predict the potential *in vivo* serious eye damage or no requirement for eye hazard classification of a test chemical (see Decision Criteria). However, no classification can be given for chemicals not predicted as causing serious eye damage or as not classified with the ICE test method (see paragraph 11).

Source and Age of Chicken Eyes

Historically, eyes collected from chickens obtained from a slaughterhouse where they are killed for human consumption have been used for this assay, eliminating the need for laboratory animals. Only the eyes of healthy animals considered suitable for entry into the human food chain are used.

Although a controlled study to evaluate the optimum chicken age has not been conducted, the age and weight of the chickens used historically in this test method are that of spring chickens traditionally processed by a poultry slaughterhouse (*i.e.*, approximately 7 weeks old, 1,5 - 2,5 kg).

Collection and Transport of Eyes to the Laboratory

Heads should be removed immediately after sedation of the chickens, usually by electric shock, and incision of the neck for bleeding. A local source of chickens close to the laboratory should be located so that their heads can be transferred from the slaughterhouse to the laboratory quickly enough to minimise deterioration and/or bacterial contamination. The time interval between collection of the chicken heads and placing the eyes in the superfusion chamber following enucleation should be minimised (typically within two hours) to assure meeting assay acceptance criteria. All eyes used in the assay should be from the same group of eyes collected on a specific day.

Because eyes are dissected in the laboratory, the intact heads are transported from the slaughterhouse at ambient temperature (typically between 18 °C and 25 °C) in plastic boxes humidified with tissues moistened with isotonic saline.

Selection Criteria and Number of Eyes Used in the ICE

Eyes that have high baseline fluorescein staining (*i.e.*, > 0,5) or corneal opacity score (*i.e.*, > 0,5) after they are enucleated are rejected.

Each treatment group and concurrent positive control consists of at least three eyes. The negative control group or the solvent control (if using a solvent other than saline) consists of at least one eye.

In the case of solid materials leading to a GHS NC outcome, a second run of three eyes is recommended to confirm or discard the negative outcome.

PROCEDURE

Preparation of the Eyes

The eyelids are carefully excised, taking care not to damage the cornea. Corneal integrity is quickly assessed with a drop of 2 % (w/v) sodium fluorescein applied to the corneal surface for a few seconds, and then rinsed with isotonic saline. Fluorescein-treated eyes are then examined with a slit-lamp microscope to ensure that the cornea is undamaged (*i.e.*, fluorescein retention and corneal opacity scores $\leq 0,5$).

If undamaged, the eye is further dissected from the skull, taking care not to damage the cornea. The eyeball is pulled from the orbit by holding the nictitating membrane firmly with surgical forceps, and the eye muscles are cut with a bent, blunt-tipped scissor. It is important to avoid causing corneal damage due to excessive pressure (*i.e.*, compression artifacts).

When the eye is removed from the orbit, a visible portion of the optic nerve should be left attached. Once removed from the orbit, the eye is placed on an absorbent pad and the nictitating membrane and other connective tissue are cut away.

The enucleated eye is mounted in a stainless steel clamp with the cornea positioned vertically. The clamp is then transferred to a chamber of the superfusion apparatus (18). The clamps should be positioned in the superfusion apparatus such that the entire cornea is supplied with the isotonic saline drip (3-4 drops per minute or 0,1 to 0,15 ml/min). The chambers of the superfusion apparatus should be temperature controlled at $32 \pm 1,5$ °C. Appendix 3 provides a diagram of a typical superfusion apparatus and the eye clamps, which can be obtained commercially or constructed. The apparatus can be modified to meet the needs of an individual laboratory (e.g. to accommodate a different number of eyes).

After being placed in the superfusion apparatus, the eyes are again examined with a slit-lamp microscope to ensure that they have not been damaged during the dissection procedure. Corneal thickness should also be measured at this time at the corneal apex using the depth measuring device on the slit-lamp microscope. Eyes with; (i), a fluorescein retention score of > 0,5; (ii) corneal opacity > 0,5; or, (iii), any additional signs of damage should be replaced. For eyes that are not rejected based on any of these criteria, individual eyes with a corneal thickness deviating more than 10 % from the mean value for all eyes are to be rejected. Users should be aware that slit-lamp microscopes could yield different corneal thickness measurements if the slit-width setting is different. The slit-width should be set at 0,095 mm.

Once all eyes have been examined and approved, the eyes are incubated for approximately 45 to 60 minutes to equilibrate them to the test system prior to dosing. Following the equilibration period, a zero reference measurement is recorded for corneal thickness and opacity to serve as a baseline (*i.e.*, time = 0). The fluorescein score determined at dissection is used as the baseline measurement for that endpoint.

Application of the Test Chemical

Immediately following the zero reference measurements, the eye (in its holder) is removed from the superfusion apparatus, placed in a horizontal position, and the test chemical is applied to the cornea.

Liquid test chemicals are typically tested undiluted, but may be diluted if deemed necessary (e.g. as part of the study design). The preferred solvent for diluted test chemicals is physiological saline. However, alternative solvents may also be used under controlled conditions, but the appropriateness of solvents other than physiological saline should be demonstrated.

Liquid test chemicals are applied to the cornea such that the entire surface of the cornea is evenly covered with the test chemical; the standard volume is 0,03 ml.

If possible, solid test chemicals should be ground as finely as possible in a mortar and pestle, or comparable grinding tool. The powder is applied to the cornea such that the surface is uniformly covered with the test chemical; the standard amount is 0,03 g.

The test chemical (liquid or solid) is applied for 10 seconds and then rinsed from the eye with isotonic saline (approximately 20 ml) at ambient temperature. The eye (in its holder) is subsequently returned to the superfusion apparatus in the original upright position. In case of need, additional rinsing may be used after the 10-sec application and at subsequent time points (e.g. upon discovery of residues of test chemical on the cornea). In general the amount of saline additionally used for rinsing is not critical, but the observation of adherence of chemical to the cornea is important.

Control Chemicals

Concurrent negative or solvent/vehicle controls and positive controls should be included in each experiment.

When testing liquids at 100 % or solids, physiological saline is used as the concurrent negative control in the ICE test method to detect non-specific changes in the test system, and to ensure that the assay conditions do not inappropriately result in an irritant response.

When testing diluted liquids, a concurrent solvent/vehicle control group is included in the test method to detect non-specific changes in the test system, and to ensure that the assay conditions do not inappropriately result in an irritant response. As stated in paragraph 31, only a solvent/vehicle that has been demonstrated to have no adverse effects on the test system can be used.

A known ocular irritant is included as a concurrent positive control in each experiment to verify that an appropriate response is induced. As the ICE assay is being used in this test method to identify corrosive or severe irritants, the positive control should be a reference chemical that induces a severe response in this test method. However, to ensure that variability in the positive control response across time can be assessed, the magnitude of the severe response should not be excessive. Sufficient *in vitro* data for the positive control should be generated such that a statistically defined acceptable range for the positive control can be calculated. If adequate historical ICE test method data are not available for a particular positive control, studies may need to be conducted to provide this information.

Examples of positive controls for liquid test chemicals are 10 % acetic acid or 5 % benzalkonium chloride, while examples of positive controls for solid test chemicals are sodium hydroxide or imidazole.

Benchmark chemicals 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.

Endpoints Measured

Treated corneas are evaluated prior to treatment and at 30, 75, 120, 180, and 240 minutes (± 5 minutes) after the post-treatment rinse. These time points provide an adequate number of measurements over the four-hour treatment period, while leaving sufficient time between measurements for the requisite observations to be made for all eyes.

The endpoints evaluated are corneal opacity, swelling, fluorescein retention, and morphological effects (e.g. pitting or loosening of the epithelium). All of the endpoints, with the exception of fluorescein retention (which is determined only prior to treatment and 30 minutes after test chemical exposure) are determined at each of the above time points.

Photographs are advisable to document corneal opacity, fluorescein retention, morphological effects and, if conducted, histopathology.

After the final examination at four hours, users are encouraged to preserve eyes in an appropriate fixative (e.g. neutral buffered formalin) for possible histopathological examination (see paragraph 14 and reference (8) for details).

Corneal swelling is determined from corneal thickness measurements made with an optical pachymeter on a slit-lamp microscope. It is expressed as a percentage and is calculated from corneal thickness measurements according to the following formula:

$$\left(\frac{\text{corneal thickness at time } t - \text{corneal thickness at time } = 0}{\text{corneal thickness at time } = 0} \right) \times 100$$

The mean percentage of corneal swelling for all test eyes is calculated for all observation time points. Based on the highest mean score for corneal swelling, as observed at any time point, an overall category score is then given for each test chemical (see paragraph 51).

Corneal opacity is evaluated by using the area of the cornea that is most densely opacified for scoring as shown in Table 1. The mean corneal opacity value for all test eyes is calculated for all observation time points. Based on the highest mean score for corneal opacity, as observed at any time point, an overall category score is then given for each test chemical (see paragraph 51).

Table 1

Corneal opacity scores

Score	Observation
0	No opacity
0,5	Very faint opacity
1	Scattered or diffuse areas; details of the iris are clearly visible
2	Easily discernible translucent area; details of the iris are slightly obscured
3	Severe corneal opacity; no specific details of the iris are visible; size of the pupil is barely discernible
4	Complete corneal opacity; iris invisible

Fluorescein retention is evaluated at the 30 minute observation time point only as shown in Table 2. The mean fluorescein retention value of all test eyes is then calculated for the 30-minute observation time point, and used for the overall category score given for each test chemical (see paragraph 51).

Table 2

Fluorescein retention scores

Score	Observation
0	No fluorescein retention
0,5	Very minor single cell staining
1	Single cell staining scattered throughout the treated area of the cornea
2	Focal or confluent dense single cell staining
3	Confluent large areas of the cornea retaining fluorescein

Morphological effects include “pitting” of corneal epithelial cells, “loosening” of epithelium, “roughening” of the corneal surface and “sticking” of the test chemical to the cornea. These findings can vary in severity and may occur simultaneously. The classification of these findings is subjective according to the interpretation of the investigator.

DATA AND REPORTING

Data Evaluation

Results from corneal opacity, swelling, and fluorescein retention should be evaluated separately to generate an ICE class for each endpoint. The ICE classes for each endpoint are then combined to generate an Irritancy Classification for each test chemical.

Decision Criteria

Once each endpoint has been evaluated, ICE classes can be assigned based on a predetermined range. Interpretation of corneal swelling (Table 3), opacity (Table 4), and fluorescein retention (Table 5) using four ICE classes is done according to the scales shown below. It is important to note that the corneal swelling scores shown in Table 3 are only applicable if thickness is measured with a slit-lamp microscope (for example Haag-Streit BP900) with depth-measuring device no. 1 and slit-width setting at 9½, equalling 0,095 mm. Users should be aware that slit-lamp microscopes could yield different corneal thickness measurements if the slit-width setting is different.

Table 3

ICE classification criteria for corneal swelling

Mean Corneal Swelling (%) (*)	ICE Class
0 to 5	I
> 5 to 12	II

Mean Corneal Swelling (%) (*)	ICE Class
> 12 to 18 (> 75 min after treatment)	II
> 12 to 18 (\leq 75 min after treatment)	III
> 18 to 26	III
> 26 to 32 (> 75 min after treatment)	III
> 26 to 32 (\leq 75 min after treatment)	IV
> 32	IV

(*) Highest mean score observed at any time point.

Table 4

ICE classification criteria for opacity

Maximum Mean Opacity Score (*)	ICE Class
0,0-0,5	I
0,6-1,5	II
1,6-2,5	III
2,6-4,0	IV

(*) Maximum mean score observed at any time point (based on opacity scores as defined in Table 1).

Table 5

ICE classification criteria for mean fluorescein retention

Mean Fluorescein Retention Score at 30 minutes post-treatment (*)	ICE Class
0,0-0,5	I
0,6-1,5	II
1,6-2,5	III
2,6-3,0	IV

(*) Based on scores as defined in Table 2.

The *in vitro* classification for a test chemical is assessed by reading the GHS classification that corresponds to the combination of categories obtained for corneal swelling, corneal opacity, and fluorescein retention as described in Table 6.

Table 6

Overall *in vitro* classifications

UN GHS Classification	Combinations of the 3 Endpoints
No Category	3 × I 2 × I, 1 × II
No prediction can be made	Other combinations
Category 1	3 × IV 2 × IV, 1 × III 2 × IV, 1 × II (*) 2 × IV, 1 × I (*) Corneal opacity ≥ 3 at 30 min (in at least 2 eyes) Corneal opacity = 4 at any time point (in at least 2 eyes) Severe loosening of the epithelium (in at least 1 eye)
(*) Combinations less likely to occur.	

Study Acceptance Criteria

A test is considered acceptable if the concurrent negative or vehicle/solvent controls and the concurrent positive controls are identified as GHS Non-Classified and GHS Category 1, respectively.

Test Report

The test report should include the following information, if relevant to the conduct of the study:

Test Chemical and Control Chemicals

- Chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known;
- The CAS Registry Number (RN), if known;
- Purity and composition of the test /control chemicals (in percentage(s) by weight), to the extent this information is available;
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class water solubility relevant to the conduct of the study;
- Treatment of the test /control chemicals prior to testing, if applicable (e.g. warming, grinding);
- Stability, if known;

Information Concerning the Sponsor and the Test Facility

- Name and address of the sponsor, test facility and study director;
- Identification on the source of the eyes (e.g. the facility from which they were collected);

Test Method Conditions

- Description of test system used;

- Slit-lamp microscope used (e.g. model) and instrument settings for the slit-lamp microscope used;
- Reference to historical negative and positive control results and, if applicable, historical data demonstrating acceptable concurrent benchmark control ranges;
- The procedure used to ensure the integrity (*i.e.*, accuracy and reliability) of the test method over time (e.g. periodic testing of proficiency chemicals)).

Eyes Collection and Preparation

- Age and weight of the donor animal and if available, other specific characteristics of the animals from which the eyes were collected (e.g. sex, strain);
- Storage and transport conditions of eyes (e.g. date and time of eye collection, time interval between collection of chicken heads and placing the enucleated eyes in superfusion chamber);
- Preparation & mounting of the eyes including statements regarding their quality, temperature of eye chambers, and criteria for selection of eyes used for testing.

Test Procedure

- Number of replicates used;
- Identity of the negative and positive controls used (if applicable, also the solvent and benchmark controls);
- Test chemical dose, application and exposure time used;
- Observation time points (pre- and post- treatment);
- Description of evaluation and decision criteria used;
- Description of study acceptance criteria used;
- Description of any modifications of the test procedure.

Results

- Tabulation of corneal swelling, opacity and fluorescein retention scores obtained for each individual eye and at each observation time point, including the mean scores at each observation time of all tested eyes;
- The highest mean corneal swelling, opacity and fluorescein retention scores observed (from any time point), and its relating ICE class.
- Description of any other effects observed;
- The derived *in vitro* GHS classification;
- If appropriate, photographs of the eye;

Discussion of the Results

Conclusion

LITERATURE:

- (1) ICCVAM (2007). Test Method Evaluation Report — *In Vitro Ocular Toxicity* Test Methods for Identifying Ocular Severe Irritants and Corrosives. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No.: 07-4517. Available: http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_tmer.htm.
- (2) ESAC (2007). Statement on the conclusion of the ICCVAM retrospective study on organotypic *in vitro* assays as screening tests to identify potential ocular corrosives and severe eye irritants. Available: <http://ecvam.jrc.it/index.htm>.

- (3) ICCVAM (2010). ICCVAM Test Method Evaluation Report — Current Status of *in vitro* Test Methods for Identifying Mild/Moderate Ocular Irritants: The Isolated Chicken Eye (ICE) Test Method. Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No.: 10-7553A. Available: <http://iccvam.niehs.nih.gov/methods/ocutox/MildMod-TMER.htm>.
 - (4) United Nations (UN) (2011). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Fourth revised edition, UN New York and Geneva, 2011. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev04/04files_e.html.
 - (5) Streamlined Summary Document Supporting OECD Test Guideline 438 on the Isolated Chicken Eye for Eye Irritation/Corrosion. Series on Testing and Assessment no. 188 (Part 1 and Part 2), OECD, Paris.
 - (6) Chapter B.5 of this Annex, Acute eye irritation/corrosion.
 - (7) Scott L, Eskes C, Hoffman S, Adriaens E, Alepee N, Bufo M, Clothier R, Facchini D, Faller C, Guest R, Hamernik K, Harbell J, Hartung T, Kamp H, Le Varlet B, Meloni M, Mcnamee P, Osborn R, Pape W, Pfannenbecker U, Prinsen M, Seaman C, Spielmann H, Stokes W, Trouba K, Vassallo M, Van den Berghe C, Van Goethem F, 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. *Toxicology In Vitro* 24, 1-9.
 - (8) OECD (2011) Guidance Document on “The Bovine Corneal Opacity and Permeability (BCOP) and Isolated Chicken Eye (ICE) Test Methods: Collection of Tissues for Histological Evaluation and Collection of Data on Non-Severe Irritants”. Series on Testing and Assessment no. 160, OECD, Paris.
 - (9) ICCVAM. (2006). Background review document: Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: Isolated Chicken Eye Test Method. NIH Publication No.: 06-4513. Research Triangle Park: National Toxicology Program. Available at: http://iccvam.niehs.nih.gov/methods/ocutox/ivocutox/ocu_brd_ice.htm.
 - (10) Prinsen, M.K. and Koëter, B.W.M. (1993). Justification of the enucleated eye test with eyes of slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. *Fd. Chem. Toxicol.* 31:69-76.
 - (11) DB-ALM (INVITTOX) (2009). Protocol 80: Chicken enucleated eye test (CEET) / Isolated Chicken Eye Test, 13pp. Available: <http://ecvam-dbal.m.jrc.ec.europa.eu/>.
 - (12) Balls, M., Botham, P.A., Bruner, L.H. and Spielmann H. (1995). The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicol. In Vitro* 9:871-929.
 - (13) Prinsen, M.K. (1996). The chicken enucleated eye test (CEET): A practical (pre)screen for the assessment of eye irritation/corrosion potential of test materials. *Food Chem. Toxicol.* 34:291-296.
 - (14) Chamberlain, M., Gad, S.C., Gautheron, P. and Prinsen, M.K. (1997). IRAG Working Group I: Organotypic models for the assessment/prediction of ocular irritation. *Food Chem. Toxicol.* 35:23-37.
 - (15) Prinsen, M.K. (2006). The Draize Eye Test and *in vitro* alternatives; a left-handed marriage? *Toxicology in Vitro* 20,78-81.
 - (16) Siegel, J.D., Rhinehart, E., Jackson, M., Chiarello, L., and the Healthcare Infection Control Practices Advisory Committee (2007). Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Healthcare Settings. Available: <http://www.cdc.gov/ncidod/dhqp/pdf/isolation2007.pdf>.
 - (17) Maurer, J.K., Parker, R.D. and 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.
 - (18) Burton, A.B.G., M. York and R.S. Lawrence (1981). The *in vitro* assessment of severe irritants. *Fd. Cosmet.-Toxicol.* 19, 471-480.
-

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.

Benchmark chemical: A chemical used as a standard for comparison to a test chemical. A benchmark chemical should have the following properties; (i), a consistent and reliable source(s); (ii), structural and functional similarity to the class of chemicals 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: step-wise approach used for a 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 a mixture.

Cornea: The transparent part of the front of the eyeball that covers the iris and pupil and admits light to the interior.

Corneal opacity: Measurement of the extent of opaqueness of the cornea following exposure to a test chemical. Increased corneal opacity is indicative of damage to the cornea.

Corneal swelling: An objective measurement in the ICE test of the extent of distension of the cornea following exposure to a test chemical. It is expressed as a percentage and is calculated from baseline (pre-dose) corneal thickness measurements and the thickness recorded at regular intervals after exposure to the test chemical in the ICE test. The degree of corneal swelling is indicative of damage to the cornea.

Eye Irritation: Production of changes in the eye following the application of 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 Category 2” (4).

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.

Fluorescein retention: A subjective measurement in the ICE test of the extent of fluorescein sodium that is retained by epithelial cells in the cornea following exposure to a test substance. The degree of fluorescein retention is indicative of damage to the corneal epithelium.

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.

Irreversible effects on the eye: see “Serious eye damage” and “UN GHS Category 1”.

Mixture: A mixture or a solution composed of two or more substances in which they do not react (4)

Negative 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.

Not Classified: Substances that are not classified for eye irritation (UN GHS Category 2) or serious damage to eye (UN GHS Category 1). Interchangeable with “UN GHS No Category”.

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 severe response should not be excessive.

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.

Reversible effects on the Eye: see “Eye Irritation” and “UN GHS Category 2”.

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 Category 1” (4).

Slit-lamp microscope: An instrument used to directly examine the eye under the magnification of a binocular microscope by creating a stereoscopic, erect image. In the ICE test method, this instrument is used to view the anterior structures of the chicken eye as well as to objectively measure corneal thickness with a depth-measuring device attachment.

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 samples 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 negative control, this sample also demonstrates whether the solvent or vehicle interacts with the test system.

Substance: Chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the 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 (4).

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.

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 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.

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 (4).

UN GHS Category 1: see “Serious damage to eyes” and/or “Irreversible effects on the eye”.

UN GHS Category 2: see “Eye Irritation” and/or “Reversible effects to the eye”.

UN GHS No Category: Substances that do not meet the requirements for classification as UN GHS Category 1 or 2 (2A or 2B). Interchangeable with “Not classified”.

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.

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 chemical.

Appendix 2

PROFICIENCY CHEMICALS FOR THE ICE TEST METHOD

Prior to routine use of a test method that adheres to this test method, laboratories should demonstrate technical proficiency by correctly identifying the eye hazard classification of the 13 chemicals recommended in Table 1. These chemicals were selected to represent the range of responses for eye hazards based on results from the *in vivo* rabbit eye test (TG 405) and the UN GHS classification system (i.e., UN GHS Categories 1, 2A, 2B, or No Category) (4)(6). Other selection criteria were that chemicals are commercially available, there are high quality *in vivo* reference data available, and there are high quality data from the ICE *in vitro* method. Reference data are available in the SSD (5) and in the ICCVAM Background Review Documents for the ICE test method (9).

Table 1

Recommended chemicals for demonstrating technical proficiency with ICE

Chemical	CASRN	Chemical Class ⁽¹⁾	Physical Form	<i>In Vivo</i> Classification ⁽²⁾	<i>In Vitro</i> Classification ⁽³⁾
Benzalkonium chloride (5 %)	8001-54-5	Onium compound	Liquid	Category 1	Category 1
Chlorhexidine	55-56-1	Amine, Amidine	Solid	Category 1	Category 1
Dibenzoyl-L-tartaric acid	2743-38-6	Carboxylic acid, Ester	Solid	Category 1	Category 1
Imidazole	288-32-4	Heterocyclic	Solid	Category 1	Category 1
Trichloroacetic acid (30 %)	76-03-9	Carboxylic Acid	Liquid	Category 1	Category 1
2,6-Dichlorobenzoyl chloride	4659-45-4	Acyl halide	Liquid	Category 2A	No predictions can be made ⁽⁴⁾
Ammonium nitrate	6484-52-2	Inorganic salt	Solid	Category 2A ⁽⁵⁾	No predictions can be made ⁽⁴⁾
Ethyl-2-methylacetoacetate	609-14-3	Ketone, Ester	Liquid	Category 2B	No predictions can be made ⁽⁴⁾
Dimethyl sulfoxide	67-68-5	Organic sulphur compound	Liquid	No Category	No Category
Glycerol	56-81-5	Alcohol	Liquid	No Category	No Category (borderline)

Chemical	CASRN	Chemical Class ⁽¹⁾	Physical Form	<i>In Vivo</i> Classification ⁽²⁾	<i>In Vitro</i> Classification ⁽³⁾
Methylcyclopentane	96-37-7	Hydrocarbon (cyclic)	Liquid	No Category	No Category
n-Hexane	110-54-3	Hydrocarbon (acyclic)	Liquid	No Category	No Category
Triacetin	102-76-1	Lipid	Liquid	Not classified	No Category

Abbreviations: CASRN = Chemical Abstracts Service Registry Number.

⁽¹⁾ Chemical classes were assigned to each test chemical using a standard classification scheme, based on the National Library of Medicine Medical Subject Headings (MeSH) classification system (available at <http://www.nlm.nih.gov/mesh>)

⁽²⁾ Based on results from the *in vivo* rabbit eye test (OECD TG 405) and using the UN GHS (4)(6).

⁽³⁾ Based on results in ICE as described in table 6.

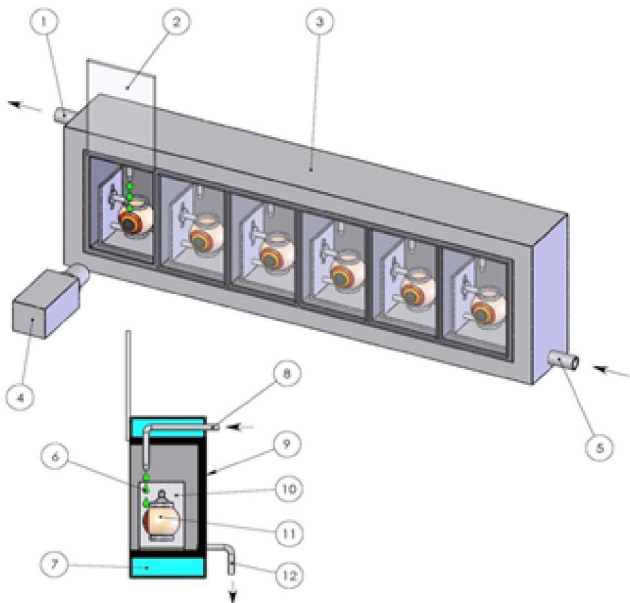
⁽⁴⁾ Combination of ICE scores other than the ones described in table 6 for the identification of GHS no-category and GHS Category 1 (see table 6)

⁽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 conjunctiva 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 conjunctiva redness score of 1 (at day 7) that fully recovered at day 10.

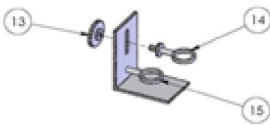
Appendix 3

DIAGRAMS OF THE ICE SUPERFUSION APPARATUS AND EYE CLAMPS

(See Burton et al. (18) for additional generic descriptions of the superfusion apparatus and eye clamp)



CROSS SECTION COMPARTMENT



EYE HOLDER

Item No.	Description	Item No.	Description
1	Outlet warm water	9	Compartment
2	Sliding door	10	Eye holder
3	Superfusion apparatus	11	Chicken eye
4	Optical measuring instrument	12	Outlet saline solution
5	Inlet warm water	13	Setscrew
6	Saline solution	14	Adjustable upper arm
7	Warm water	15	Fixed lower arm'
8	Inlet saline solution		

(14) In Part B, Chapter B.49 is replaced by the following:

'B.49 *In Vitro* Mammalian Cell Micronucleus Test

INTRODUCTION

This test method is equivalent to OECD test guideline 487 (2016). It is part of a series of test methods on genetic toxicology. An OECD document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to these Test Guidelines has been developed (1).

The *in vitro* micronucleus (MNvit) test is a genotoxicity test for the detection of micronuclei (MN) in the cytoplasm of interphase cells. Micronuclei may originate from acentric chromosome fragments (*i.e.* lacking a centromere), or whole chromosomes that are unable to migrate to the poles during the anaphase stage of cell division. Therefore the MNvit test is an *in vitro* method that provides a comprehensive basis for investigating chromosome damaging potential *in vitro* because both aneugens and clastogens can be detected (2) (3) in cells that have undergone cell division during or after exposure to the test chemical (see paragraph 13 for more details). Micronuclei represent damage that has been transmitted to daughter cells, whereas chromosome aberrations scored in metaphase cells may not be transmitted. In either case, the changes may not be compatible with cell survival.

This test method allows the use of protocols with and without the actin polymerisation inhibitor cytochalasin B (cytoB). The addition of cytoB prior to mitosis results in cells that are binucleate and therefore allows for the identification and analysis of micronuclei in only those cells that have completed one mitosis (4) (5). This test method also allows for the use of protocols without cytokinesis block, provided there is evidence that the cell population analysed has undergone mitosis.

In addition to using the MNvit test to identify chemicals that induce micronuclei, the use of immunochemical labelling of kinetochores, or hybridisation with centromeric/telomeric probes (fluorescence *in situ* hybridisation (FISH)), also can provide additional information on the mechanisms of chromosome damage and micronucleus formation (6) (7) (8) (9) (10) (11) (12) (13) (14) (15) (16) (17). Those labelling and hybridisation procedures can be used when there is an increase in micronucleus formation and the investigator wishes to determine if the increase was the result of clastogenic and/or aneugenic events.

Because micronuclei in interphase cells can be assessed relatively objectively, laboratory personnel need only determine the number of binucleate cells when cytoB is used and the incidence of micronucleate cells in all cases. As a result, the slides can be scored relatively quickly and analysis can be automated. This makes it practical to score thousands instead of hundreds of cells per treatment, increasing the power of the test. Finally, as micronuclei may arise from lagging chromosomes, there is the potential to detect aneuploidy-inducing agents that are difficult to study in conventional chromosomal aberration tests, e.g. Chapter B.10 of this annex (18). However, the MNvit test as described in this test method does not allow for the differentiation of chemicals inducing changes in chromosome number and/or ploidy from those inducing clastogenicity without special techniques such as FISH mentioned under paragraph 4.

The MNvit test is robust and can be conducted in a variety of cell types, and in the presence or absence of cytoB. There are extensive data to support the validity of the MNvit test using various cell types (cultures of cell lines or primary cell cultures) (19) (20) (21) (22) (23) (24) (25) (26) (27) (28) (29) (30) (31) (32) (33) (34) (35) (36). These include, in particular, the international validation studies co-ordinated by the Société Française de Toxicologie Génétique (SFTG) (19) (20) (21) (22) (23) and the reports of the International Workshop on Genotoxicity Testing (5) (17). The available data have also been re-evaluated in a weight-of-evidence retrospective validation study by the European Centre for the Validation of Alternative Methods (ECVAM) of the European Commission (EC), and the test method has been endorsed as scientifically valid by the ECVAM Scientific Advisory Committee (ESAC) (37) (38) (39).

The mammalian cell MNvit test may employ cultures of cell lines or primary cell cultures, of human or rodent origin. Because the background frequency of micronuclei will influence the sensitivity of the test, it is recommended that cell types with a stable and defined background frequency of micronucleus formation be used. The cells used are selected on the basis of their ability to grow well in culture, stability of their karyotype (including chromosome number) and spontaneous frequency of micronuclei (40). At the present time, the available data do not allow firm recommendations to be made but suggest it is important, when evaluating chemical hazards to consider the *p53* status, genetic (karyotype) stability, DNA repair capacity and origin (rodent versus human) of the cells chosen for testing. The users of this test method are thus encouraged to consider the influence of these and other cell characteristics on the performance of a cell line in detecting the induction of micronuclei, as knowledge evolves in this area.

Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

Tests conducted *in vitro* generally require the use of an exogenous source of metabolic activation unless the cells are metabolically competent with respect to the test chemicals. The exogenous metabolic activation system does not entirely mimic *in vivo* conditions. Care should be taken to avoid conditions that could lead to artifactual positive results which do not reflect the genotoxicity of the test chemicals. Such conditions include changes in pH (41) (42) (43) or osmolality, interaction with the cell culture medium (44) (45) or excessive levels of cytotoxicity (see paragraph 29).

To analyse the induction of micronuclei, it is essential that mitosis has occurred in both treated and untreated cultures. The most informative stage for scoring micronuclei is in cells that have completed one mitosis during or after treatment with the test chemical. For Manufactured Nanomaterials, specific adaptations of this test method are needed but they are not described in this test method.

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

Cell cultures of human or other mammalian origin are exposed to the test chemical both with and without an exogenous source of metabolic activation unless cells with an adequate metabolising capability are used (see paragraph 19).

During or after exposure to the test chemical, the cells are grown for a period sufficient to allow chromosome damage or other effects on cell cycle/cell division to lead to the formation of micronuclei in interphase cells. For induction of aneuploidy, the test chemical should ordinarily be present during mitosis. Harvested and stained interphase cells are analysed for the presence of micronuclei. Ideally, micronuclei should only be scored in those cells that have completed mitosis during exposure to the test chemical or during the post-treatment period, if one is used. In cultures that have been treated with a cytokinesis blocker, this is easily achieved by scoring only binucleate cells. In the absence of a cytokinesis blocker, it is important to demonstrate that the cells analysed are likely to have undergone cell division, based on an increase in the cell population, during or after exposure to the test chemical. For all protocols, it is important to demonstrate that cell proliferation has occurred in both the control and treated cultures, and the extent of test chemical-induced cytotoxicity or cytostasis should be assessed in all of the cultures that are scored for micronuclei.

DESCRIPTION OF THE METHOD

Cells

Cultured primary human or other mammalian peripheral blood lymphocytes (7) (20) (46) (47) and a number of rodent cell lines such as CHO, V79, CHL/IU, and L5178Y cells or human cell lines such as TK6 can be used (19) (20) (21) (22) (23) (26) (27) (28) (29) (31) (33) (34) (35) (36) (see paragraph 6). Other cell lines such as HT29 (48), Caco-2 (49), HepaRG (50) (51), HepG2 cells (52) (53), A549 and primary Syrian Hamster Embryo cells (54) have been used for micronucleus testing but at this time have not been extensively validated. Therefore the use of those cell lines and types should be justified based on their demonstrated performance in the test, as described in the Acceptability Criteria section. Cyto B was reported to potentially impact L5178Y cell growth and therefore is not recommended with this cell line (23). When primary cells are used, for animal welfare reasons, the use of cells from human origin should be considered where feasible and sampled in accordance with the human ethical principles and regulations.

Human peripheral blood lymphocytes should be obtained from young (approximately 18-35 years of age), non-smoking individuals with no known illness or recent exposures to genotoxic agents (e.g. chemicals, ionising radiation) at levels that would increase the background incidence of micronucleate cells. This would ensure the background incidence of micronucleate cells to be low and consistent. The baseline incidence of micronucleate cells increases with age and this trend is more marked in females than in males (55). If cells from more than one donor are pooled for use, the number of donors should be specified. It is necessary to demonstrate that the cells have divided from the beginning of treatment with the test chemical to cell sampling. Cell cultures are maintained in an exponential growth phase (cell lines) or stimulated to divide (primary cultures of lymphocytes) to expose the cells at different stages of the cell cycle, since the sensitivity of cell stages to the test chemicals may not be known. The primary cells that need to be stimulated with mitogenic agents in order to divide are generally no longer synchronised during exposure to the test chemical (e.g. human lymphocytes after a 48-hour mitogenic stimulation). The use of synchronised cells during treatment with the test chemical is not recommended, but can be acceptable if justified.

Media and culture conditions

Appropriate culture medium and incubation conditions (culture vessels, humidified atmosphere of 5 % CO₂ if appropriate, temperature of 37 °C) should be used for maintaining cultures. Cell lines should be checked routinely for the stability of the modal chromosome number and the absence of *Mycoplasma* contamination, and cells should not be used if contaminated or if the modal chromosome number has changed. The normal cell cycle time of cell lines or primary cultures used in the testing laboratory should be established and should be consistent with the published cell characteristics.

Preparation of cultures

Cell lines: cells 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 until harvest time (e.g. confluence should be avoided for cells growing in monolayers).

Lymphocytes: whole blood treated with an anti-coagulant (e.g. heparin), or separated lymphocytes, are cultured (e.g. for 48 hours for human lymphocytes) in the presence of a mitogen (e.g. phytohaemagglutinin (PHA) for human lymphocytes) in order to induce cell division prior to exposure to the test chemical and cytoB.

Metabolic activation

Exogenous metabolising systems should be used when employing cells with inadequate endogenous metabolic capacity. The most commonly used system that is recommended by default, unless another system is 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 (56) (57) or a combination of phenobarbital and b-naphthoflavone (58) (59) (60). The latter combination does not conflict with the Stockholm Convention on Persistent Organic Pollutants (61) and has been shown to be as effective as Aroclor 1254 for inducing mixed-function oxidases (58) (59) (60). 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 use of products that reduce the mitotic index, especially calcium complexing products (62), should be avoided during treatment. The choice of type and concentration of exogenous metabolic activation system or metabolic inducer employed may be influenced by the class of chemicals being tested.

Test chemical preparation

Solid test chemicals should be prepared in appropriate solvents and diluted, if appropriate, prior to treatment of the cells. 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 vessels (63) (64) (65). Preparations of the test chemical should be made just prior to treatment unless stability data demonstrate the acceptability of storage.

Test Conditions

Solvents

The solvent should be chosen to optimise the solubility of the test chemicals without adversely impacting the conduct of the assay, i.e. changing cell growth, affecting 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 (DMSO). Generally organic solvents should not exceed 1 % (v/v). If cytoB is dissolved in DMSO, the total amount of organic solvent used for both the test chemical and cytoB should not exceed 1 % (v/v); otherwise, untreated controls should be used to ensure that the percentage of organic solvent has no adverse effect. 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 chemical, the test system and their lack of genetic toxicity at the concentration used. In the absence of that supporting data, it is important to include untreated controls (see Appendix 1), as well as solvent controls to demonstrate that no deleterious or chromosomal effects (e.g. aneuploidy or clastogenicity) are induced by the chosen solvent.

Use of cytoB as a cytokinesis blocker

One of the most important considerations in the performance of the MNvit test is ensuring that the cells being scored have completed mitosis during the treatment or the post-treatment incubation period, if one is used. Micronucleus scoring, therefore, should be limited to cells that have gone through mitosis during or after treatment. CytoB is the agent that has been most widely used to block cytokinesis because it inhibits actin assembly, and thus prevents separation of daughter cells after mitosis, leading to the formation of binucleate cells (6) (66) (67). The effect of the test chemical on cell proliferation kinetics can be measured simultaneously, when cytoB is used. CytoB should be used as a cytokinesis blocker when human lymphocytes are used because cell cycle times will be variable among donors and because not all lymphocytes will respond to PHA stimulation. CytoB is not mandatory for other cell types if it can be established they have undergone division as described in paragraph 27. Moreover CytoB is not generally used when samples are evaluated for micronuclei using flow cytometric methods.

The appropriate concentration of cytoB should be determined by the laboratory for each cell type to achieve the optimal frequency of binucleate cells in the solvent control cultures and should be shown to produce a good yield of binucleate cells for scoring. The appropriate concentration of cytoB is usually between 3 and 6 µg/ml (19).

Measuring cell proliferation and cytotoxicity and choosing treatment concentrations

When determining the highest test chemical concentration, concentrations that have the capability of producing artifactual positive responses, such as those producing excessive cytotoxicity (see paragraph 29), precipitation in the culture medium (see paragraph 30), or marked changes in pH or osmolality (see paragraph 9), 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 artifactual positive results and to maintain appropriate culture conditions.

Measurements of cell proliferation are made to assure that sufficient treated cells have undergone mitosis during the test and that the treatments are conducted at appropriate levels of cytotoxicity (see paragraph 29). Cytotoxicity should be determined in the main experiment with and without metabolic activation using an appropriate indication of cell death and growth (see paragraphs 26 and 27). While the evaluation of cytotoxicity in an initial preliminary test may be useful to better define the concentrations to be used in the main experiment, an initial test is not mandatory. If performed, it should not replace the measurement of cytotoxicity in the main experiment.

Treatment of cultures with cytoB and measurement of the relative frequencies of mononucleate, binucleate, and multi-nucleate cells in the culture provides an accurate method of quantifying the effect on cell proliferation and the cytotoxic or cytostatic activity of a treatment (6), and ensures that only cells that divided during or after treatment are microscopically scored. The cytokinesis-block proliferation index (CBPI) (6) (27) (68) or the Replication Index (RI) from at least 500 cells per culture (see Appendix 2 for formulas) are recommended to estimate the cytotoxic and cytostatic activity of a treatment by comparing values in the treated and control cultures. Assessment of other indicators of cytotoxicity (e.g. cell integrity, apoptosis, necrosis, metaphase counting, cell cycle) could provide useful information, but should not be used in place of CBPI or RI.

In studies without cytoB, it is necessary to demonstrate that the cells in culture have divided, so that a substantial proportion of the cells scored have undergone division during or following treatment with the test chemical, otherwise false negative responses may be produced. The measurement of Relative Population Doubling (RPD) or Relative Increase in Cell Count (RICC) is recommended to estimate the cytotoxic and cytostatic activity of a treatment (17) (68) (69) (70) (71) (see Appendix 2 for formulas). At extended sampling times (e.g. treatment for 1,5-2 normal cell cycle lengths and harvest after an additional 1,5-2 normal cell cycle lengths, leading to sampling times longer than 3-4 normal cell cycle lengths in total as described in paragraphs 38 and 39), RPD might underestimate cytotoxicity (71). Under these circumstances RICC might be a better measure or the evaluation of cytotoxicity after a 1,5-2 normal cell cycle lengths would be a helpful estimate. Assessment of other markers for cytotoxicity or cytostasis (e.g. cell integrity, apoptosis, necrosis, metaphase counting, Proliferation index (PI), cell cycle, nucleoplasmic bridges or nuclear buds) could provide useful additional information, but should not be used in place of either the RPD or RICC.

At least three test concentrations (not including the solvent and positive controls) that meet the acceptability criteria (appropriate cytotoxicity, number of cells, etc) should be evaluated. Whatever the types of cells (cell lines or primary cultures of lymphocytes), either replicate or single treated cultures may be used at each concentration tested. While the use of duplicate cultures is advisable, single cultures are also acceptable provided that the same total number of cells are scored for either single or duplicate cultures. The use of single cultures is particularly relevant when more than 3 concentrations are assessed (see paragraphs 44-45). The results obtained from the independent replicate cultures at a given concentration can be pooled for the data analysis. 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 as described in paragraph 29 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 obtain data at low and moderate cytotoxicity or to study the dose response relationship in detail, it will be necessary to use more closely spaced concentrations and/or more than three concentrations (single cultures or replicates) in particular in situations where a repeat experiment is required (see paragraph 60).

If the maximum concentration is based on cytotoxicity, the highest concentration should aim to achieve $55 \pm 5\%$ cytotoxicity using the recommended cytotoxicity parameters (i.e. reduction in RICC and RPD for cell lines when cytoB is not used, and reduction in CBPI or RI when cytoB is used to $45 \pm 5\%$ of the concurrent negative control) (72). Care should be taken in interpreting positive results only found in the higher end of this $55 \pm 5\%$ cytotoxicity range (71).

For poorly soluble test chemicals that are not cytotoxic at concentrations lower than 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 inducing turbidity or with visible precipitate because artifactual 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 (e.g. staining or scoring). The determination of solubility in the culture medium prior to the experiment may be useful.

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 (73) (74) (75). When the test chemical is not of defined composition, e.g. a substance of unknown or variable composition, complex reaction products or biological materials (UVCB) (76), environmental extract, 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 (93).

Controls

Concurrent negative controls (see paragraph 21), consisting of solvent alone in the treatment medium and processed in the same way as the treatment cultures, should be included for every harvest time.

Concurrent positive controls are needed to demonstrate the ability of the laboratory to identify clastogens and aneugens 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 chemicals can be used, if justified.

At the present time, no aneugens are known that require metabolic activation for their genotoxic activity (17). Because *in vitro* mammalian cell tests for genetic toxicity are sufficiently standardised for the short-term treatments done concurrently with and without metabolic activation using the same treatment duration, the use of positive controls may be confined to a clastogen requiring metabolic activation. In this case a single clastogenic positive control response will demonstrate both the activity of the metabolic activation system and the responsiveness of the test system. However, long term treatment (without S9) should have its own positive control, as the treatment duration will differ from the test using metabolic activation. If a clastogen is selected as the single positive control for short-term treatment with and without metabolic activation, an aneugen should be selected for the long-term treatment without metabolic activation. Positive controls for both clastogenicity and aneugenicity should be used in metabolically competent cells that do not require S9.

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 (*i.e.* the effects are clear but do not immediately reveal the identity of the coded slides to the reader), and the response should not be compromised by cytotoxicity exceeding the limits specified in this test method.

Table 1

Reference chemicals recommended for assessing laboratory proficiency and for the selection of positive controls

Category	Chemical	CASRN
1. Clastogens active without metabolic activation		
	Methyl methanesulphonate	66-27-3
	Mitomycin C	50-07-7
	4-Nitroquinoline-N-Oxide	56-57-5
	Cytosine arabinoside	147-94-4
2. Clastogens requiring metabolic activation		
	Benzo(a)pyrene	50-32-8
	Cyclophosphamide	50-18-0

Category	Chemical	CASRN
3. Aneugens		
	Colchicine	64-86-8
	Vinblastine	143-67-9

PROCEDURE

Treatment Schedule

In order to maximise the probability of detecting an aneugen or clastogen acting at a specific stage in the cell cycle, it is important that sufficient numbers of cells representing all of the various stages of their cell cycles are treated with the test chemical. All treatments should commence and end while the cells are growing exponentially and the cells should continue to grow up to the time of sampling. The treatment schedule for cell lines and primary cell cultures may, therefore, differ somewhat from that for lymphocytes which require mitogenic stimulation to begin their cell cycle (17). For lymphocytes, the most efficient approach is to start the treatment with the test chemical at 44-48 hours after PHA stimulation, when cells will be dividing asynchronously (6).

Published data (19) indicate that most aneugens and clastogens will be detected by a short term treatment period of 3 to 6 hours in the presence and absence of S9, followed by removal of the test chemical and sampling at a time equivalent to about 1,5 - 2,0 normal cell cycle lengths after the beginning of treatment (7).

However, for thorough evaluation, which would be needed to conclude a negative outcome, all three following experimental conditions should be conducted using a short term treatment with and without metabolic activation and long term treatment without metabolic activation (see paragraphs 56, 57 and 58):

- Cells should be exposed to the test chemical without metabolic activation for 3-6 hours, and sampled at a time equivalent to about 1,5 - 2,0 normal cell cycle lengths after the beginning of treatment (19),
- Cells should be exposed to the test chemical with metabolic activation for 3-6 hours, and sampled at a time equivalent to about 1,5 - 2,0 normal cell cycle lengths after the beginning of treatment (19),
- Cells should be continuously exposed without metabolic activation until sampling at a time equivalent to about 1,5 - 2,0 normal cell cycle lengths.

In the event that any of the above experimental conditions lead to a positive response, it may not be necessary to investigate any of the other treatment regimens.

If it is known or suspected that the test chemical affects the cell cycling time (e.g. when testing nucleoside analogues), especially for p53 competent cells (35) (36) (77), sampling or recovery times may be extended by up to a further 1,5 - 2,0 normal cell cycle lengths (i.e. total 3,0 to 4,0 cell cycle lengths after the beginning of short-term and long-term treatments). These options address situations where there may be concern regarding possible interactions between the test chemical and cytoB. When using extended sampling times (i.e. total 3,0 to 4,0 cell cycle lengths culture time), care should be taken to ensure that the cells are still actively dividing. For example, for lymphocytes exponential growth may be declining at 96 hours following stimulation and monolayer cultures of cells may become confluent.

The suggested cell treatment schedules are summarised in Table 2. These general treatment schedules may be modified (and should be justified) depending on the stability or reactivity of the test chemical or the particular growth characteristics of the cells being used.

Table 2

Cell treatment and harvest times for the MNvit test

Lymphocytes, primary cells and cell lines treated <u>with</u> cytoB	+ S9 Short treatment	Treat for 3-6 hours in the presence of S9; remove the S9 and treatment medium; add fresh medium and cytoB; harvest 1,5 - 2,0 normal cell cycle lengths after the beginning of treatment.
	– S9 Short treatment	Treat for 3-6 hours; remove the treatment medium; add fresh medium and cytoB; harvest 1,5 - 2,0 normal cell cycle lengths after the beginning of treatment.
	– S9 Extended treatment	Treat for 1,5 - 2 normal cell cycle lengths in the presence of cytoB; harvest at the end of the treatment period.

Cell lines treated without cytoB

(Identical to the treatment schedules outlined above with the exception that no cytoB is added)

For monolayer cultures, mitotic cells (identifiable as being round and detaching from the surface) may be present at the end of the 3-6 hour treatment. Because these mitotic cells are easily detached, they can be lost when the medium containing the test chemical is removed. If there is evidence for a substantial increase in the number of mitotic cells compared with controls, indicating likely mitotic arrest, then the cells should be collected by centrifugation and added back to the culture, to avoid losing cells that are in mitosis, and at risk for micronuclei/chromosome aberration, at the time of harvest.

Cell harvest and slide preparation

Each culture should be harvested and processed separately. Cell preparation may involve hypotonic treatment, but this step is not necessary if adequate cell spreading is otherwise achieved. Different techniques can be used in slide preparation provided that high-quality cell preparations for scoring are obtained. Cells with intact cell membrane and intact cytoplasm should be retained to allow the detection of micronuclei and (in the cytokinesis-block method) reliable identification of binucleate cells.

The slides can be stained using various methods, such as Giemsa or fluorescent DNA specific dyes. The use of appropriate fluorescent stains (e.g. acridine orange (78) or Hoechst 33258 plus pyronin-Y (79)) can eliminate some of the artifacts associated with using a non-DNA specific stain. Anti-kinetochore antibodies, FISH with pancentromeric DNA probes, or primed *in situ* labelling with pancentromere-specific primers, together with appropriate DNA counterstaining, can be used to identify the contents (whole chromosomes will be stained while acentric chromosome fragments will not) of micronuclei if mechanistic information of their formation is of interest (16) (17). Other methods for differentiation between clastogens and aneugens may be used if they have been shown to be effective and validated. For example, for certain cell lines the measurements of sub-2N nuclei as hypodiploid events using techniques such as image analysis, laser scanning cytometry or flow cytometry could also provide useful information (80) (81) (82). Morphological observations of nuclei could also give indications of possible aneuploidy. Moreover, a test for metaphase chromosome aberrations, preferably in the same cell type and protocol with comparable sensitivity, could also be a useful way to determine whether micronuclei are due to chromosome breakage (knowing that chromosome loss would not be detected in the chromosome aberration test).

Analysis

All slides, including those of the solvent and the untreated (if used) and positive controls, should be independently coded before the microscopic analysis of micronucleus frequencies. Appropriate techniques should be used to control any bias or drift when using an automated scoring system, for instance, flow cytometry, laser scanning cytometry or image analysis. Regardless of the automated platform is used to enumerate micronuclei, CBPI, RI, RPD, or RICC should be assessed concurrently.

In cytoB-treated cultures, micronucleus frequencies should be analysed in at least 2 000 binucleate cells per concentration and control (83), equally divided among the replicates, if replicates are used. In the case of single cultures per dose (see paragraph 28), at least 2 000 binucleate cells per culture (83) should be scored in this single culture. If substantially fewer than 1 000 binucleate cells per culture (for duplicate cultures), or 2 000 (for single culture), are available for scoring at each concentration, and if a significant increase in micronuclei is not detected, the test should be repeated using more cells, or at less cytotoxic concentrations, whichever is appropriate. Care should be taken not to score binucleate cells with irregular shapes or where the two nuclei differ greatly in size. In addition, binucleate cells should not be confused with poorly spread multi-nucleate cells. Cells containing more than two main nuclei should not be analysed for micronuclei, as the baseline micronucleus frequency may be higher in these cells (84). Scoring of mononucleate cells is acceptable if the test chemical is shown to interfere with cytoB activity. A repeat test without CytoB might be useful in such cases. Scoring mononucleate cells in addition to binucleate cells could provide useful information (85) (86), but is not mandatory.

In cell lines tested without cytoB treatment, micronuclei should be scored in at least 2 000 cells per test concentration and control (83), equally divided among the replicates, if replicates are used. When single cultures per concentration are used (see paragraph 28), at least 2 000 cells per culture should be scored in this single culture. If substantially fewer than 1 000 cells per culture (for duplicate cultures), or 2 000 (for single culture), are available for scoring at each concentration, and if a significant increase in micronuclei is not detected, the test should be repeated using more cells, or at less cytotoxic concentrations, whichever is appropriate.

When cytoB is used, a CBPI or an RI should be determined to assess cell proliferation (see Appendix 2) using at least 500 cells per culture. When treatments are performed in the absence of cytoB, it is essential to provide evidence that the cells in culture have divided, as discussed in paragraphs 24-28.

Proficiency of the laboratory

In order to establish sufficient experience with the assay prior to using it for routine testing, the laboratory should have performed a series of experiments with reference positive chemicals acting via different mechanisms (at least one with and one without metabolic activation, and one acting via an aneugenic mechanism, and selected from the chemicals listed in Table 1) and various negative controls (including untreated cultures and various solvents/vehicle). 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 49 to 52.

A selection of positive control chemicals (see Table 1) should be investigated with short and 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 clastogenic and aneugenic chemicals, determine the effectiveness of the metabolic activation system and demonstrate the appropriateness of the scoring procedures (microscopic visual analysis, flow cytometry, laser scanning cytometry or image analysis). A range of concentrations of the selected chemicals 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

The laboratory should establish:

- A historical positive control range and distribution,
- A historical negative (untreated, solvent) control range and distribution.

When first acquiring data for an historical negative control distribution, concurrent negative controls should be consistent with published negative control data where they exist. As more experimental data are added to the control distribution, concurrent negative controls should ideally be within the 95 % control limits of that distribution (87) (88). 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 (88)), to identify how variable their positive and negative control data are, and to show that the methodology is 'under control' in their laboratory (83). 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 (87).

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.

Negative control data should consist of the incidence of micronucleated cells from a single culture or the sum of replicate cultures as described in paragraph 28. Concurrent negative controls should ideally be within the 95 % control limits of the distribution of the laboratory's historical negative control database (87) (88). Where concurrent negative control data fall outside the 95 % control limits, they may be acceptable for inclusion in the historical control distribution as long as these data are not extreme outliers and there is evidence that the test system is 'under control' (see paragraph 50) and there is evidence of absence of technical or human failure.

DATA AND REPORTING

Presentation of the results

If the cytokinesis-block technique is used, only the frequencies of binucleate cells with micronuclei (independent of the number of micronuclei per cell) are used in the evaluation of micronucleus induction. The scoring of the numbers of cells with one, two, or more micronuclei can be reported separately and could provide useful information, but is not mandatory.

Concurrent measures of cytotoxicity and/or cytostasis for all treated, negative and positive control cultures should be determined (16). The CBPI or the RI should be calculated for all treated and control cultures as measurements of cell cycle delay when the cytokinesis-block method is used. In the absence of cytoB, the RPD or the RICC should be used (see Appendix 2).

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

Acceptability Criteria

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 50.
- Concurrent positive controls (see paragraph 50) should induce responses that are compatible with those generated in the laboratory's historical positive control data base and produce a statistically significant increase compared with the concurrent negative control.
- Cell proliferation criteria in the solvent control should be fulfilled (paragraph 25-27).
- All experimental conditions were tested unless one resulted in positive results (paragraphs 36-40).
- Adequate number of cells and concentrations are analysable (paragraphs 28 and 44-46).
- The criteria for the selection of top concentration are consistent with those described in paragraphs 24-31.

Evaluation and interpretation of results

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 paragraphs 36-39):

- at least one of the test concentrations exhibits a statistically significant increase compared with the concurrent negative control (89)
- the increase is dose-related in at least one experimental condition when evaluated with an appropriate trend test (see paragraph 28)
- any of the results are outside the distribution of the historical negative control data (e.g. Poisson-based 95 % control limits; see paragraph 52).

When all of these criteria are met, the test chemical is then considered able to induce chromosome breaks and/or gain or loss in this test system. Recommendations for the most appropriate statistical methods can also be found in the literature (90) (91) (92).

Providing that all acceptability criteria are fulfilled, a test chemical is considered clearly negative if, in all experimental conditions examined (see paragraphs 36-39):

- 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 limits; see paragraph 52).

The test chemical is then considered unable to induce chromosome breaks and/or gain or loss in this test system. Recommendations for the most appropriate statistical methods can also be found in the literature (90) (91) (92).

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

In case 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. Scoring additional cells (where appropriate) or 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.

In rare cases, even after further investigations, the data set will not allow a conclusion of positive or negative, and will therefore be concluded as equivocal.

Test chemicals that induce micronuclei in the MNvit test may do so because they induce chromosome breakage, chromosome loss, or a combination of the two. Further analysis using anti-kinetochore antibodies, centromere specific *in situ* probes, or other methods may be used to determine whether the mechanism of micronucleus induction is due to clastogenic and/or aneugenic activity.

Test Report

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;
- reactivity of the test chemicals with the solvent/vehicle or cell culture media;
- 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:

- type and source of cells used;
- suitability of the cell type used;
- absence of mycoplasma, in case of cell lines;
- for cell lines, information on cell cycle length or proliferation index;
- where lymphocytes are used, sex of blood donors, age and any relevant information on the donor, whole blood or separated lymphocytes, mitogen used;
- normal (negative control) cell cycle time;
- number of passages, if available, for cell lines;
- methods for the maintenance of cell cultures, for cell lines;
- modal number of chromosomes, for cell lines;

Test Conditions:

- identity of the cytokinesis blocking substance (e.g. cytoB), if used, and its concentration and duration of cell exposure;
- concentration of the test chemical expressed as a final concentration in the culture medium (e.g. µg or mg/mL, or mM of culture medium);
- rationale for the selection of concentrations and the number of cultures, including cytotoxicity data and solubility limitations;
- composition of media, CO₂ concentration, if applicable, humidity level;
- concentration (and/or volume) of the solvent and test chemical added in the culture medium;
- incubation temperature and time;
- duration of treatment;
- harvest time after treatment;
- cell density at seeding, if applicable;
- 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 (e.g. enzymatic activity, sterility, metabolic capability);
- positive and negative control chemicals, final concentrations, conditions and durations of treatment and recovery periods;
- methods of slide preparation and the staining technique used;
- criteria for scoring micronucleate cells (selection of analysable cells and identification of micronucleus);
- numbers of cells analysed;

- methods for the measurements of cytotoxicity;
- any supplementary information relevant to cytotoxicity and method used;
- criteria for considering studies as positive, negative, or equivocal;
- method(s) of statistical analysis used;
- methods, such as use of anti-kinetochore antibody or pan-centromeric specific probes, to characterise whether micronuclei contain whole or fragmented chromosomes, if applicable;
- methods used to determine pH, osmolality and precipitation.

Results:

- definition of acceptable cells for analysis;
- in the absence of cyto B, the number of cells treated and the number of cells harvested for each culture in case of cell lines;
- measurement of cytotoxicity used, e.g. CBPI or RI in the case of cytokinesis-block method; RICC or RPD when cytokinesis-block methods are not used; other observations if any (e.g. cell confluency, apoptosis, necrosis, metaphase counting, frequency of binucleated cells);
- signs of precipitation and time of the determination;
- data on pH and osmolality of the treatment medium, if determined;
- distribution of mono-, bi-, and multi-nucleate cells if a cytokinesis block method is used;
- number of cells with micronuclei given separately for each treated and control culture, and defining whether from binucleate or mononucleate cells, where appropriate;
- 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 deviation and 95 % control limits for the distribution, as well as the number of data;
- statistical analysis; p-values if any.

Discussion of the results.

Conclusions.

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) Kirsch-Volders, M. (1997). Towards a validation of the micronucleus test. *Mutation Research*, Vol. 392/1-2, pp. 1-4.
- (3) Parry, J.M., A. Sors (1993). The detection and assessment of the aneugenic potential of environmental chemicals: the European Community aneuploidy project. *Mutation Research*, Vol. 287/1, pp. 3-15.
- (4) Fenech, M., A.A. Morley (1985). Solutions to the kinetic problem in the micronucleus assay. *Cytobios*, Vol. 43/172-173, pp. 233-246.

- (5) Kirsch-Volders, M. *et al.* (2000). Report from the *In Vitro* Micronucleus Assay Working Group. *Environmental and Molecular Mutagenesis*, Vol. 35/3, pp. 167-172.
- (6) Fenech, M. (2007). Cytokinesis-block micronucleus cytome assay. *Nature Protocols*, Vol. 2/5, pp. 1084-1104.
- (7) Fenech, M., A.A. Morley (1986). Cytokinesis-block micronucleus method in human lymphocytes: effect of *in-vivo* ageing and low dose X-irradiation. *Mutation Research*, Vol. 161/2, pp. 193-198.
- (8) Eastmond, D.A., J.D. Tucker (1989). Identification of aneuploidy-inducing agents using cytokinesis-blocked human lymphocytes and an antikinetochore antibody. *Environmental and Molecular Mutagenesis*, Vol. 13/1, pp. 34-43.
- (9) Eastmond, D.A., D. Pinkel (1990). Detection of aneuploidy and aneuploidy-inducing agents in human lymphocytes using fluorescence *in-situ* hybridisation with chromosome-specific DNA probe. *Mutation Research*, Vol. 234/5, pp. 9-20.
- (10) Miller, B.M. *et al.* (1991). Classification of micronuclei in murine erythrocytes: immunofluorescent staining using CREST antibodies compared to *in situ* hybridization with biotinylated gamma satellite DNA. *Mutagenesis*, Vol. 6/4, pp. 297-302.
- (11) Farooqi, Z., F. Darroudi, A. T. Natarajan (1993). The use of fluorescence *in-situ* hybridisation for the detection of aneugens in cytokinesis-blocked mouse splenocytes. *Mutagenesis*, Vol. 8/4, pp. 329-334.
- (12) Migliore, L. *et al.* (1993). Cytogenetic damage induced in human lymphocytes by four vanadium compounds and micronucleus analysis by fluorescence *in situ* hybridization with a centromeric probe. *Mutation Research*, Vol. 319/3, pp. 205-213.
- (13) Norppa, H., L. Renzi, C. Lindholm (1993). Detection of whole chromosomes in micronuclei of cytokinesis-blocked human lymphocytes by antikinetochore staining and *in situ* hybridization. *Mutagenesis*, Vol. 8/6, pp. 519-525.
- (14) Eastmond, D.A., D.S. Rupa, L.S. Hasegawa (1994). Detection of hyperdiploidy and chromosome breakage in interphase human lymphocytes following exposure to the benzene metabolite hydroquinone using multicolor fluorescence *in situ* hybridization with DNA probes. *Mutation Research*, Vol. 322/1, pp. 9-20.
- (15) Marshall, R.R. *et al.* (1996). Fluorescence *in situ* hybridisation (FISH) with chromosome-specific centromeric probes: a sensitive method to detect aneuploidy. *Mutation Research*, Vol. 372/2, pp. 233-245.
- (16) Zijno, P. *et al.* (1996). Analysis of chromosome segregation by means of fluorescence *in situ* hybridization: application to cytokinesis-blocked human lymphocytes. *Mutation Research*, Vol. 372/2, pp. 211-219.
- (17) Kirsch-Volders *et al.* (2003). Report from the *in vitro* micronucleus assay working group. *Mutation Research*, Vol. 540/2, pp. 153-163.
- (18) Chapter B.10 of this Annex: *In Vitro Mammalian Chromosome Aberration Test*,.
- (19) Lorge, E. *et al.* (2006). SFTG International collaborative Study on *in vitro* micronucleus test. I. General conditions and overall conclusions of the study. *Mutation Research*, Vol. 607/1, pp. 13-36.
- (20) Clare, G. *et al.* (2006). SFTG International collaborative study on the *in vitro* micronucleus test. II. Using human lymphocytes. *Mutation Research*, Vol. 607/1, pp. 37-60.
- (21) Aardema, M.J. *et al.* (2006). SFTG International collaborative study on the *in vitro* micronucleus test, III. Using CHO cells. *Mutation Research*, Vol. 607/1, pp. 61-87.
- (22) Wakata, A. *et al.* (2006). SFTG International collaborative study on the *in vitro* micronucleus test, IV. Using CHO/IU cells. *Mutation Research*, Vol. 607/1, pp. 88-124.

- (23) Oliver, J. *et al.* (2006). SFTG International collaborative study on the *in vitro* micronucleus test, V. Using L5178Y cells. *Mutation Research*, Vol. 607/1, pp. 125-152.
- (24) Albertini, S. *et al.* (1997). Detailed data on *in vitro* MNT and *in vitro* CA: industrial experience. *Mutation Research*, Vol. 392/1-2, pp. 187-208.
- (25) Miller, B. *et al.* (1997). Comparative evaluation of the *in vitro* micronucleus test and the *in vitro* chromosome aberration test: industrial experience. *Mutation Research*, Vol. 392/1-2, pp. 45-59.
- (26) Miller, B. *et al.* (1998). Evaluation of the *in vitro* micronucleus test as an alternative to the *in vitro* chromosomal aberration assay: position of the GUM Working Group on the *in vitro* micronucleus test. Gesellschaft für Umwelt-Mutations-forschung, *Mutation Research*, Vol. 410, pp. 81-116.
- (27) Kalweit, S. *et al.* (1999). Chemically induced micronucleus formation in V79 cells — comparison of three different test approaches. *Mutation Research*, Vol. 439/2, pp. 183-190.
- (28) Kersten, B. *et al.* (1999). The application of the micronucleus test in Chinese hamster V79 cells to detect drug-induced photogenotoxicity. *Mutation Research*, Vol. 445/1, pp. 55-71.
- (29) von der Hude, W. *et al.* (2000). *In vitro* micronucleus assay with Chinese hamster V79 cells — results of a collaborative study with *in situ* exposure to 26 chemical substances. *Mutation Research*, Vol. 468/2, pp. 137-163.
- (30) Garriott, M.L., J.B. Phelps, W.P. Hoffman (2002). A protocol for the *in vitro* micronucleus test, I. Contributions to the development of a protocol suitable for regulatory submissions from an examination of 16 chemicals with different mechanisms of action and different levels of activity. *Mutation Research*, Vol. 517/1-2, pp. 123-134.
- (31) Matsushima, T. *et al.* (1999). Validation study of the *in vitro* micronucleus test in a Chinese hamster lung cell line (CHL/IU). *Mutagenesis*, Vol. 14/6, pp. 569-580.
- (32) Elhajouji, A., E. Lorge (2006). Special Issue: SFTG International collaborative study on *in vitro* micronucleus test. *Mutation Research*, Vol. 607/1, pp. 1-152.
- (33) Kirkland, D. (2010). Evaluation of different cytotoxic and cytostatic measures for the *in vitro* micronucleus test (MNVit): Introduction to the collaborative trial. *Mutation Research*, Vol. 702/2, pp. 139-147.
- (34) Hashimoto K. *et al.* (2011). Comparison of four different treatment conditions of extended exposure in the *in vitro* micronucleus assay using TK6 lymphoblastoid cells. *Regulatory Toxicology and Pharmacology*, Vol. 59/1, pp. 28-36.
- (35) Honma, M., M. Hayashi (2011). Comparison of *in vitro* micronucleus and gene mutation assay results for p53-competent versus p53-deficient human lymphoblastoid cells. *Environmental and Molecular Mutagenesis*, Vol. 52/5, pp. 373-384.
- (36) Zhang, L.S. *et al.* (1995). A comparative study of TK6 human lymphoblastoid and L5178Y mouse lymphoma cell lines in the *in vitro* micronucleus test. *Mutation Research Letters*, Vol. 347/3-4, pp. 105-115.
- (37) ECVAM (2006). Statement by the European Centre for the Validation of Alternative Methods (ECVAM) Scientific Advisory Committee (ESAC) on the scientific validity of the *in vitro* micronucleus test as an alternative to the *in vitro* chromosome aberration assay for genotoxicity testing. ESAC 25th meeting, 16-17 November 2006, Available at: <http://ecvam.jrc.it/index.htm>.
- (38) ESAC (2006). ECVAM Scientific Advisory Committee (ESAC) Peer Review, Retrospective Validation of the *In Vitro* Micronucleus Test, Summary and Conclusions of the Peer Review Panel. Available at: <http://ecvam.jrc.it/index.htm>.
- (39) Corvi, R. *et al.* (2008). ECVAM Retrospective Validation of *in vitro* Micronucleus Test (MNT). *Mutagenesis*, Vol. 23/4, pp. 271-283.

- (40) ILSI paper (draft). Lorge, E., M.M. Moore, J. Clements, M. O'Donovan, M. Honma, A. Kohara, J. van Benthem, S. Galloway, M.J. Armstrong, A. Sutter, V. Thybaud, B. Gollapudi, M. Aardema, J. Young-Tannir. Standardized Cell Sources and Recommendations for Good Cell Culture Practices in Genotoxicity Testing. *Mutation Research*.
- (41) Scott, D. *et al.* (1991). International Commission for Protection Against Environmental Mutagens and Carcinogens, Genotoxicity under extreme culture conditions. A report from ICPEMC Task Group 9. *Mutation Research*, Vol. 257/2, pp. 147-205.
- (42) Morita, T. *et al.* (1992). Clastogenicity of low pH to various cultured mammalian cells. *Mutation Research*, Vol. 268/2, pp. 297-305.
- (43) Brusick, D. (1986). Genotoxic effects in cultured mammalian cells produced by low pH treatment conditions and increased ion concentrations. *Environmental Mutagenesis*, Vol. 8/6, pp. 789-886.
- (44) Long, L.H. *et al.* (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 Research*, Vol. 634/1-2, pp. 177-183.
- (45) Nesslany, F. *et al.* (2008). Characterization of the Genotoxicity of Nitrilotriacetic Acid. *Environmental and Molecular Mutation.*, Vol. 49, pp. 439-452.
- (46) Fenech, M., A.A. Morley (1985). Measurement of micronuclei in lymphocytes. *Mutation Research*, Vol. 147/1-2, pp. 29-36.
- (47) Fenech, M. (1997). The advantages and disadvantages of cytokinesis-blood micronucleus method. *Mutation Research*, Vol. 392, pp. 11-18.
- (48) Payne, C.M. *et al.* (2010). Hydrophobic bile acid-induced micronuclei formation, mitotic perturbations, and decreases in spindle checkpoint proteins: relevance to genomic instability in colon carcinogenesis. *Nutrition and Cancer*, Vol. 62/6, pp. 825-840.
- (49) Bazin, E. *et al.* (2010). Genotoxicity of a Freshwater Cyanotoxin, Cylindrospermopsin, in Two Human Cell Lines: Caco-2 and HepaRG. *Environmental and Molecular Mutagenesis*, Vol. 51/3, pp. 251-259.
- (50) Le Hegarat, L. *et al.* (2010). Assessment of the genotoxic potential of indirect chemical mutagens in HepaRG cells by the comet and the cytokinesis-block micronucleus assays. *Mutagenesis*, Vol. 25/6, pp. 555-560.
- (51) Josse, R. *et al.* (2012). An adaptation of the human HepaRG cells to the *in vitro* micronucleus assay. *Mutagenesis*, Vol. 27/3, pp. 295-304.
- (52) Ehrlich, V. *et al.* (2002). Fumonisin B₁ is genotoxic in human derived hepatoma (HepG2) cells. *Mutagenesis*, Vol. 17/3, pp. 257-260.
- (53) Knasmüller, S. *et al.* (2004). Use of human-derived liver cell lines for the detection of environmental and dietary genotoxins; current state of knowledge. *Toxicology*, Vol. 198/1-3, pp. 315-328.
- (54) Gibson, D.P. *et al.* (1997). Induction of micronuclei in Syrian hamster embryo cells: comparison to results in the SHE cell transformation assay for National Toxicology Program test chemicals. *Mutation Research*, Vol. 392/1-2, pp. 61-70.
- (55) Bonassi, S. *et al.* (2001). HUMAN MicroNucleus Project: international database comparison for results with the cytokinesis-block micronucleus assay in human lymphocytes, I. Effect of laboratory protocol, scoring criteria and host factors on the frequency of micronuclei. *Environmental and Molecular Mutagenesis*, Vol. 37/1, pp. 31-45.
- (56) Maron, D.M., B.N. Ames (1983). Revised methods for the Salmonella mutagenicity test. *Mutation Research*, Vol. 113/3-4, pp. 173-215.
- (57) 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. *Journal of Environmental Pathology and Toxicology*, Vol. 4/1, pp. 55-65.

- (58) Elliott, B.M. *et al.* (1992). Alternatives to Aroclor 1254-induced S9 in *in-vitro* genotoxicity assays. *Mutagenesis*, Vol. 7, pp. 175-177.
- (59) Matsushima, T. *et al.* (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.
- (60) Johnson, T.E., D. R. Umbenhauer, S.M. Galloway (1996). Human liver S-9 metabolic activation: proficiency in cytogenetic assays and comparison with phenobarbital/beta-naphthoflavone or Aroclor 1254 induced rat S-9. *Environmental and Molecular Mutagenesis*, Vol. 28, pp. 51-59.
- (61) UNEP (2001). Stockholm Convention on Persistent Organic Pollutants, United Nations Environment Programme (UNEP). Available at: <http://www.pops.int/>
- (62) Tucker, J.D., M.L. Christensen (1987). Effects of anticoagulants upon sister-chromatid exchanges, cell-cycle kinetics, and mitotic index in human peripheral lymphocytes. *Mutation Research*, Vol.190/3, pp. 225-8.
- (63) Krahn, D.F., F.C. Barsky, K.T. McCooey (1982). "CHO/HGPRT Mutation Assay: Evaluation of Gases and Volatile Liquids", in *Genotoxic Effects of Airborne Agents*. Tice, R.R., D.L. Costa, K.M. Schaich (eds), Plenum, New York, pp. 91-103.
- (64) Zamora, P.O. *et al.* (1983). Evaluation of an exposure system using cells grown on collagen gels for detecting highly volatile mutagens in the CHO/HGPRT mutation assay. *Environmental Mutagenesis*, Vol. 5/6, pp. 795-801.
- (65) Asakura, M. *et al.* (2008). An improved system for exposure of cultured mammalian cells to gaseous compounds in the chromosomal aberration assay. *Mutation Research*, Vol. 652/2, pp. 122-130.
- (66) Fenech, M. (1993). The cytokinesis-block micronucleus technique: a detailed description of the method and its application to genotoxicity studies in human populations. *Mutation Research*, Vol. 285/1, pp. 35-44.
- (67) Phelps, J.B., M.L. Garriott, W.P. Hoffman (2002). A protocol for the *in vitro* micronucleus test. II. Contributions to the validation of a protocol suitable for regulatory submissions from an examination of 10 chemicals with different mechanisms of action and different levels of activity. *Mutation Research*, Vol. 521/1-2, pp. 103-112.
- (68) Kirsch-Volders, M. *et al.* (2004). Corrigendum to "Report from the *in vitro* micronucleus assay working group". *Mutation Research*, 564, 97-100.
- (69) Lorge, E. *et al.* (2008). Comparison of different methods for an accurate assessment of cytotoxicity in the *in vitro* micronucleus test. I. Theoretical aspects. *Mutation Research*, Vol. 655/1-2, pp. 1-3.
- (70) Surrallés, J. *et al.* (1995). Induction of micronuclei by five pyrethroid insecticides in whole-blood and isolated human lymphocyte cultures. *Mutation Research*, Vol. 341/3, pp. 169-184.
- (71) Honma, M. (2011). Cytotoxicity measurement in *in vitro* chromosome aberration test and micronucleus test. *Mutation Research*, Vol. 724, pp. 86-87.
- (72) Pfuhler, S. *et al.* (2011). *In vitro* genotoxicity test approaches with better predictivity: Summary of an IWGT workshop. *Mutation Research*, Vol. 723/2, pp. 101-107.
- (73) 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). ENV/JM/TG(2014)17. Available upon request.
- (74) Morita T., M. Honma, K. Morikawa (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 Research*, Vol. 741, pp. 32-56.
- (75) Brookmire L., J.J. Chen, D.D. Levy (2013). Evaluation of the Highest Concentrations Used in the *in vitro* Chromosome Aberrations Assay. *Environmental and Molecular Mutagenesis*, Vol. 54/1, pp. 36-43.

- (76) EPA, Office of Chemical Safety and Pollution Prevention (2011). Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials: UVCB Substances. <http://www.epa.gov/opptintr/newchemicals/pubs/uvcb.txt>.
- (77) Sobol, Z. *et al.* (2012). Development and validation of an *in vitro* micronucleus assay platform in TK6 cells. *Mutation Research*, Vol. 746/1, pp. 29-34.
- (78) Hayashi, M., T. Sofuni, M. Jr. Ishidate (1983). An Application of Acridine Orange Fluorescent Staining to the Micronucleus Test. *Mutation Research*, Vol. 120/4, pp. 241-247.
- (79) MacGregor, J. T., C.M. Wehr, R.G. Langlois (1983). A Simple Fluorescent Staining Procedure for Micronuclei and RNA in Erythrocytes Using Hoechst 33258 and Pyronin Y. *Mutation Research*, Vol. 120/4, pp. 269-275.
- (80) Bryce, S.M. *et al.* (2011). Miniaturized flow cytometry-based CHO-K1 micronucleus assay discriminates aneugenic and clastogenic modes of action. *Environmental and Molecular Mutagenesis*, Vol. 52/4, pp. 280-286.
- (81) Nicolette, J. *et al.* (2011). *in vitro* micronucleus screening of pharmaceutical candidates by flow cytometry in Chinese hamster V79 cells. *Environmental and Molecular Mutagenesis*, Vol. 52/5, pp. 355-362.
- (82) Shi, J., R. Bezabhe, A. Szkudlinska (2010). Further evaluation of a flow cytometric *in vitro* micronucleus assay in CHO-K1 cells: a reliable platform that detects micronuclei and discriminates apoptotic bodies. *Mutagenesis*, Vol. 25/1, pp. 33-40.
- (83) OECD (2014). Statistical analysis supporting the revision of the genotoxicity Test Guidelines. OECD Environment, Health and Safety Publications (EHS), Series on testing and assessment, No. 198, OECD Publishing, Paris.
- (84) Fenech, M. *et al.* (2003). HUMN project: detailed description of the scoring criteria for the cytokinesis-block micronucleus assay using isolated human lymphocyte cultures. *Mutation Research*, Vol. 534/1-2, pp. 65-75.
- (85) Elhajoui, A., M. Cunha, M. Kirsch-Volders (1998). Spindle poisons can induce polyploidy by mitotic slippage and micronucleate mononucleates in the cytokinesis-block assay. *Mutagenesis*, Vol. 13/2, pp. 193-8.
- (86) Kirsch-Volders, M. *et al.* (2011). The *in vitro* MN assay in 2011: origin and fate, biological significance, protocols, high throughput methodologies and toxicological relevance. *Archives of Toxicology*, Vol. 85/8, pp. 873-99.
- (87) Hayashi, M. *et al.* (2010). Compilation and use of genetic toxicity historical control Data. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 723/2, pp. 87-90.
- (88) Ryan, T. P. (2000). *Statistical Methods for Quality Improvement*. 2nd ed., John Wiley and Sons, New York.
- (89) Hoffman, W.P., M.L. Garriott, C. Lee (2003). "In vitro micronucleus test", in *Encyclopedia of Biopharmaceutical Statistics*, 2nd ed. Chow, S. (ed.), Marcel Dekker, Inc. New York, pp. 463-467.
- (90) Fleiss, J. L., B. Levin, M.C. Paik (2003). *Statistical Methods for Rates and Proportions*, 3rd ed. John Wiley & Sons, New York.
- (91) Galloway, S.M. *et al.* (1987). Chromosome aberration and sister chromatid exchanges in Chinese hamster ovary cells: Evaluation of 108 chemicals. *Environmental and Molecular Mutagenesis*, Vol. 10/suppl. 10, pp. 1-175.
- (92) Richardson, C. *et al.* (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.
- (93) International Conference on Harmonisation (ICH) Guidance S2 (R1) on Genotoxicity Testing and Data Interpretation for Pharmaceuticals Intended For Human Use.
-

Appendix 1

DEFINITIONS:

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

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

Apoptosis: programmed cell death characterised by a series of steps leading to the disintegration of cells into membrane-bound particles that are then eliminated by phagocytosis or by shedding.

Cell proliferation: the increase in cell number as a result of mitotic cell division.

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

Chemical: a substance or a mixture.

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

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

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

Cytokinesis-Block Proliferation index (CBPI): the proportion of second-division cells in the treated population relative to the untreated control (see Appendix 2 for formula).

Cytostasis: inhibition of cell growth (see Appendix 2 for formula).

Cytotoxicity: For the assays covered in this test method performed in the presence of cytochalasin B, cytotoxicity is identified as a reduction in cytokinesis-block proliferation index (CBPI) or Replication Index (RI) of the treated cells as compared to the negative control (see paragraph 26 and Appendix 2)

For the assays covered in this test method performed in the absence of cytochalasin B, cytotoxicity is identified as a reduction in relative population doubling (RPD) or relative increase in cell count (RICC) of the treated cells as compared to the negative control (see paragraph 27 and Appendix 2).

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

Interphase cells: cells not in the mitotic stage.

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

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

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

Mitotic index: 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 cell proliferation of that population.

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

Non-disjunctio: failure of paired chromatids to disjoin and properly segregate to the developing daughter cells, resulting in daughter cells with abnormal numbers of chromosomes.

p53 status: p53 protein is involved in cell cycle regulation, apoptosis and DNA repair. Cells deficient in functional p53 protein, unable to arrest cell cycle or to eliminate damaged cells via apoptosis or other mechanisms (e.g. induction of DNA repair) related to p53 functions in response to DNA damage, should be theoretically more prone to gene mutations or chromosomal aberrations.

Polyploidy: numerical chromosome aberrations in cells or organisms involving entire set(s) of chromosomes, as opposed to an individual chromosome or chromosomes (aneuploidy).

Proliferation Index (PI): method for cytotoxicity measurement when cytoB is not used (see Appendix 2 for formula).

Relative Increase in Cell Count (RICC): method for cytotoxicity measurement when cytoB is not used (see Appendix 2 for formula).

Relative Population Doubling (RPD): method for cytotoxicity measurement when cytoB is not used (see Appendix 2 for formula).

Replication Index (RI): the proportion of cell division cycles completed in a treated culture, relative to the untreated control, during the exposure period and recovery (see Appendix 2 for formula).

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

S9 mix: mix of the S9 liver 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. no test chemical nor solvent) but are processed concurrently in the same way as the cultures receiving the test chemical.

Appendix 2

FORMULAS FOR CYTOTOXICITY ASSESSMENT

When cytoB is used, evaluation of cytotoxicity should be based on the **Cytokinesis-Block Proliferation Index (CBPI)** or **Replication Index (RI)** (17) (69). The CBPI indicates the average number of nuclei per cell, and may be used to calculate cell proliferation. The RI indicates the relative number of cell cycles per cell during the period of exposure to cytoB in treated cultures compared to control cultures and can be used to calculate the % cytostasis:

$$\% \text{ Cytostasis} = 100 - 100\{(\text{CBPI}_T - 1) \div (\text{CBPI}_C - 1)\}$$

and:

T = test chemical treatment culture

C = control culture

where:

$$\text{CBPI} = \frac{((\text{No. mononucleate cells}) + (2 \times \text{No. binucleate cells}) + (3 \times \text{No. multinucleate cells}))}{(\text{Total number of cells})}$$

Thus, a CBPI of 1 (all cells are mononucleate) is equivalent to 100 % cytostasis.

$$\text{Cytostasis} = 100 - \text{RI}$$

$$\text{RI} = \frac{((\text{No. binucleate cells}) + (2 \times \text{No. multinucleate cells})) / (\text{Total number of cells})_T}{((\text{No. binucleate cells}) + (2 \times \text{No. multinucleate cells})) / (\text{Total number of cells})_C} \times 100$$

T = treated cultures

C = control cultures

Thus, an RI of 53 % means that, compared to the numbers of cells that have divided to form binucleate and multinucleate cells in the control culture, only 53 % of this number divided in the treated culture, i.e. 47 % cytostasis.

When cytoB is not used, evaluation of cytotoxicity based on **Relative Increase in Cell Counts (RICC)** or on **Relative Population Doubling (RPD)** is recommended (69), as both take into account the proportion of the cell population which has divided.

$$\text{RICC}(\%) = \frac{(\text{Increase in number of cells in treated cultures}(\text{final-starting}))}{(\text{Increase in numbers of cells in control cultures}(\text{final-starting}))} \times 100$$

$$\text{RPD}(\%) = \frac{(\text{No. of Population doublings in treated cultures})}{(\text{No. of population doublings in control cultures})} \times 100$$

where:

$$\text{Population Doubling} = [\log (\text{Post-treatment cell number} \div \text{Initial cell number})] \div \log 2$$

Thus, a RICC, or a RPD of 53 % indicates 47 % cytotoxicity/cytostasis.

By using a **Proliferation Index (PI)**, cytotoxicity may be assessed via counting the number of clones consisting of 1 cell (cl1), 2 cells (cl2), 3 to 4 cells (cl4) and 5 to 8 cells (cl8).

$$\text{PI} = \frac{((1 \times \text{cl1}) + (2 \times \text{cl2}) + (3 \times \text{cl4}) + (4 \times \text{cl8}))}{(\text{cl1} + \text{cl2} + \text{cl4} + \text{cl8})}$$

The PI has been used as a valuable and reliable cytotoxicity parameter also for cell lines cultured *in vitro* in the absence of cytoB (35) (36) (37) (38) and can be seen as a useful additional parameter.

In any case, the number of cells before treatment should be the same for treated and negative control cultures.

While RCC (*i.e.* Number of cells in treated cultures/Number of cells in control cultures) had been used as cytotoxicity parameter in the past, is no longer recommended because it can underestimate cytotoxicity.

When using automated scoring systems, for instance, flow cytometry, laser scanning cytometry or image analysis, the number of cells in the formula can be substituted by the number of nuclei.

In the negative control cultures, population doubling or replication index should be compatible with the requirement to sample cells after treatment at a time equivalent to about 1,5 - 2,0 normal cell cycle.'

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

'B.59 *In Chemico* Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA)

INTRODUCTION

This test method (TM) is equivalent to the OECD test guideline (TG) 442C (2015). 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 Regulation (EC) No 1272/2008 of the European Parliament and Council on Classification, Labelling and Packaging of Substances and Mixtures (CLP) ⁽¹⁾. This test method provides an *in chemico* procedure (Direct Peptide Reactivity Assay — DPRA) to be used for supporting the discrimination between skin sensitizers and non-sensitizers in accordance with the UN GHS and CLP.

There is general agreement regarding the key biological events underlying skin sensitisation. The existing knowledge of the chemical and biological mechanisms associated with skin sensitisation has been summarised in the form of an Adverse Outcome Pathway (AOP) (2), from the molecular initiating event through the intermediate events to the adverse effect namely allergic contact dermatitis in humans or contact hypersensitivity in rodents. Within the skin sensitisation AOP, the molecular initiating event is the covalent binding of electrophilic substances to nucleophilic centres in skin proteins.

The assessment of skin sensitisation has typically involved the use of laboratory animals. The classical methods based on guinea-pigs, the Magnusson Kligman Guinea Pig Maximisation Test (GMPT) and the Buehler Test (TM B.6 (3)), study both the induction and elicitation phases of skin sensitisation. A murine test, the Local Lymph Node Assay (LLNA, TM B.42 (4)) and its two non-radioactive modifications, LLNA: DA (TM B.50 (5)) and LLNA: BrdU-ELISA (TM B.51 (6)), which all assess the induction response exclusively, have also gained acceptance since they provide an advantage over the guinea pig tests in terms of animal welfare and an objective measurement of the induction phase of skin sensitisation.

More recently, mechanistically based *in chemico* and *in vitro* test methods have been considered scientifically valid for the evaluation of the skin sensitisation hazard of chemicals. However, combinations of non-animal methods (*in silico*, *in chemico*, *in vitro*) within Integrated Approaches to Testing and Assessment (IATA) will be needed to be able to fully substitute for the animal tests currently in use given the restricted AOP mechanistic coverage of each of the currently available non-animal test methods (2) (7).

⁽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, 31.12.2008, p. 1).

The DPRA is proposed to address the molecular initiating event of the skin sensitisation AOP, namely protein reactivity, by quantifying the reactivity of test chemicals towards model synthetic peptides containing either lysine or cysteine (8). Cysteine and lysine percent peptide depletion values are then used to categorise a substance in one of four classes of reactivity for supporting the discrimination between skin sensitisers and non-sensitisers (9).

The DPRA has been evaluated in a European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)-lead validation study and subsequent independent peer review by the EURL ECVAM Scientific Advisory Committee (ESAC) and was considered scientifically valid (10) to be used as part of an IATA to support the discrimination between skin sensitisers and non-sensitisers for the purpose of hazard classification and labelling. Examples on the use of DPRA data in combination with other information are reported in the literature (11) (12) (13) (14).

Definitions are provided in Appendix I.

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

The correlation of protein reactivity with skin sensitisation potential is well established (15) (16) (17). Nevertheless, since protein binding represents only one key event, albeit the molecular initiating event of the skin sensitisation AOP, protein reactivity information generated with testing and non-testing methods may not be sufficient on its own to conclude on the absence of skin sensitisation potential of chemicals. Therefore, data generated with this test method should be considered in the context of integrated approaches such as IATA, combining them with other 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.

This test method can be used, in combination with other complementary information, to support the discrimination between skin sensitisers (i.e. UN GHS/CLP Category 1) and non-sensitisers in the context of IATA. This test method cannot be used on its own, neither to sub-categorise skin sensitisers into subcategories 1A and 1B as defined by UN GHS/CLP, nor to predict potency for safety assessment decisions. However, depending on the regulatory framework, a positive result with the DPRA may be used on its own to classify a chemical into UN GHS/CLP category 1.

The DPRA test method proved to be transferable to laboratories experienced in high-performance liquid chromatography (HPLC) analysis. The level of reproducibility in predictions that can be expected from the test method is in the order of 85 % within laboratories and 80 % between laboratories (10). Results generated in the validation study (18) and published studies (19) overall indicate that the accuracy of the DPRA in discriminating sensitisers (i.e. UN GHS/CLP Cat. 1) from non-sensitisers is 80 % (N=157) with a sensitivity of 80 % (88/109) and specificity of 77 % (37/48) when compared to LLNA results. The DPRA is more likely to under predict 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) (18) (19). However, the accuracy values given here for the DPRA as a stand-alone test method are only indicative since the test method should be considered in combination with other sources of information in the context of an IATA and in accordance with the provisions of paragraph 9 above. 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 the species of interest, i.e. humans. On the basis of the overall data available, the DPRA 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 physico-chemical properties (8) (9) (10) (19). Taken together, this information indicates the usefulness of the DPRA to contribute to the identification of skin sensitisation hazard.

The term “test chemical” is used in this test method to refer to what is being tested and is not related to the applicability of the DPRA to the testing of substances and/or mixtures. This test method is not applicable for the testing of metal compounds since they are known to react with proteins with mechanisms other than covalent binding. A test chemical should be soluble in an appropriate solvent at a final concentration of 100 mM (see paragraph 18). However, test chemicals that are not soluble at this concentration may still be tested at lower soluble concentrations. In such a case, a positive result could still be used to support the identification of the test chemical as a skin sensitiser but no firm conclusion on the lack of reactivity should be drawn from a negative result. Limited information is currently available on the applicability of the DPRA to mixtures of known composition (18) (19). The DPRA is nevertheless considered to be technically applicable to the testing of multi-constituent substances and mixtures of known composition (see paragraph 18). 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 current prediction model cannot be used for complex mixtures of unknown composition or for substances of unknown or variable composition, complex reaction products or biological materials (i.e. UVCB substances) due to the defined molar ratio of test chemical and peptide. For this purpose a new prediction model based on a gravimetric approach will need to be developed. In cases where evidence can be demonstrated on the non-applicability of the test method to other specific categories of chemicals, the test method should not be used for those specific categories of chemicals.

This test method is an *in chemico* method that does not encompass a metabolic system. Chemicals that require enzymatic bioactivation to exert their skin sensitisation potential (i.e. pro-haptens) cannot be detected by the test method. Chemicals that become sensitisers after abiotic transformation (i.e. pre-haptens) are reported to be in some cases correctly detected by the test method (18). In the light of the above, negative results obtained with the test method should be interpreted in the context of the stated limitations and in the connection with other information sources within the framework of an IATA. Test chemicals that do not covalently bind to the peptide but promote its oxidation (i.e. cysteine dimerisation) could lead to a potential over estimation of peptide depletion, resulting in possible false positive predictions and/or assignment to a higher reactivity class (see paragraphs 29 and 30).

As described, the DPRA supports the discrimination between skin sensitisers and non-sensitisers. However, it may also potentially contribute to the assessment of sensitising potency (11) when used in integrated approaches such as IATA. However further work, preferably based on human data, is required to determine how DPRA results may possibly inform potency assessment.

PRINCIPLE OF THE TEST

The DPRA is an *in chemico* method which quantifies the remaining concentration of cysteine- or lysine-containing peptide following 24 hours incubation with the test chemical at $25 \pm 2,5$ °C. The synthetic peptides contain phenylalanine to aid in the detection. Relative peptide concentration is measured by high-performance liquid chromatography (HPLC) with gradient elution and UV detection at 220 nm. Cysteine- and lysine peptide percent depletion values are then calculated and used in a prediction model (see paragraph 29) which allows assigning the test chemical to one of four reactivity classes used to support the discrimination between sensitisers and non-sensitisers.

Prior to routine use of the method described in this test method, laboratories should demonstrate technical proficiency, using the ten proficiency substances listed in Appendix 2.

PROCEDURE

This test method is based on the DPRA DB-ALM protocol n° 154 (20) 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 method in the laboratory. The following is a description of the main components and procedures for the DPRA. If an alternative HPLC set-up is used, its equivalence to the validated set-up described in the DB-ALM protocol should be demonstrated (e.g. by testing the proficiency substances in Appendix 2).

Preparation of the cysteine or lysine-containing peptides

Stock solutions of cysteine (Ac-RFAACAA-COOH) and lysine (Ac-RFAAKAA-COOH) containing synthetic peptides of purity higher than 85 % and preferably in the range of 90-95 %, should be freshly prepared just before their incubation with the test chemical. The final concentration of the cysteine peptide should be 0,667 mM in pH 7,5 phosphate buffer whereas the final concentration of the lysine peptide should be 0,667 mM in pH 10,2 ammonium acetate buffer. The HPLC run sequence should be set up in order to keep the HPLC analysis time less than 30 hours. For the HPLC set up used in the validation study and described in this test method, up to 26 analysis samples (which include the test chemical, the positive control and the appropriate number of solvent controls based on the number of individual solvents used in the test, each tested in triplicate), can be accommodated in a single HPLC run. All of the replicates analysed in the same run should use the identical cysteine and lysine peptide stock solutions. It is recommended to prove individual peptide batches for proper solubility prior to their use.

Preparation of the test chemical

Solubility of the test chemical in an appropriate solvent should be assessed before performing the assay following the solubilisation procedure described in the DPRA DB-ALM protocol (20). An appropriate solvent will dissolve the test chemical completely. Since in the DPRA the test chemical is incubated in large excess with either the cysteine or the lysine peptides, visual inspection of the forming of a clear solution is considered sufficient to ascertain that the test chemical (and all of its components in the case of testing a multi-constituent substance or a mixture) is dissolved. Suitable solvents are acetonitrile, water, 1:1 mixture water:acetonitrile, isopropanol, acetone or 1:1 mixture acetone:acetonitrile. Other solvents can be used as long as they do not impact on the stability of the peptide as monitored with reference controls C (i.e. samples constituted by the peptide alone dissolved in the appropriate solvent; see Appendix 3). As a last option if the test chemical is not soluble in any of these solvents attempts should be made to solubilise it in 300 µL of DMSO and dilute the resulting solution with 2 700 µL of acetonitrile and if the test chemical is not soluble in this mixture attempts should be made to solubilise the same amount of test chemical in 1 500 µL of DMSO and dilute the resulting solution with 1 500 µL of acetonitrile. The test chemical should be pre-weighed into glass vials and dissolved immediately before testing in an appropriate solvent to prepare a 100 mM solution. For mixtures and multi-constituent substances of known composition, a single purity should be determined by the sum of the proportion of its constituents (excluding water), and a single apparent molecular weight should be determined by considering the individual molecular weights of each component in the mixture (excluding water) and their individual proportions. The resulting purity and apparent molecular weight should then be used to calculate the weight of test chemical necessary to prepare a 100 mM solution. For polymers for which a predominant molecular weight cannot be determined, the molecular weight of the monomer (or the apparent molecular weight of the various monomers constituting the polymer) may be considered to prepare a 100 mM solution. However, when testing mixtures, multi-constituent substances or polymers of known composition, it should be considered to also test the neat chemical. For liquids, the neat chemical should be tested as such without any prior dilution by incubating it at 1:10 and 1:50 molar ratio with the cysteine and lysine peptides, respectively. For solids, the test chemical should be dissolved to its maximum soluble concentration in the same solvent used to prepare the apparent 100 mM solution. It should then be tested as such without any further dilution by incubating it at 1:10 and 1:50 ratio with the cysteine and lysine peptides, respectively. Concordant results (reactive or non-reactive) between the apparent 100 mM solution and the neat chemical should allow for a firm conclusion on the result.

Preparation of the positive control, reference controls and coelution controls

Cinnamic aldehyde (CAS 104-55-2; ≥ 95 % food-grade purity) should be used as positive control (PC) at a concentration of 100 mM in acetonitrile. Other suitable positive controls preferentially providing mid-range depletion values may be used if historical data are available to derive comparable run acceptance criteria. In addition, reference controls (i.e. samples containing only the peptide dissolved in the appropriate solvent) should also be included in the HPLC run sequence and these are used to verify the HPLC system suitability prior to the analysis (reference controls A), the stability of the reference controls over time (reference controls B) and to verify that the solvent used to dissolve the test chemical does not impact the percent peptide depletion (reference controls C) (see Appendix 3). The appropriate reference control for each chemical is used to calculate the percent peptide depletion for that chemical (see paragraph 26). In addition a co-elution control constituted by the test chemical alone for each of the test chemicals analysed should be included in the run sequence to detect possible co-elution of the test chemical with either the lysine or the cysteine peptide.

Incubation of the test chemical with the cysteine and lysine peptide solutions

Cysteine and lysine peptide solutions should be incubated in glass autosampler vials with the test chemical at 1:10 and 1:50 ratio respectively. If a precipitate is observed immediately upon addition of the test chemical solution to the peptide solution, due to low aqueous solubility of the test chemical, in this case one cannot be sure how much test chemical remained in the solution to react with the peptide. Therefore, in such a case, a positive result could still be used, but a negative result is uncertain and should be interpreted with due care (see also provisions in paragraph 11 for the testing of chemicals not soluble up to a concentration of 100 mM). The reaction solution should be left in the dark at $25 \pm 2,5$ °C for 24 ± 2 hours before running the HPLC analysis. Each test chemical should be analysed in triplicate for both peptides. Samples have to be visually inspected prior to HPLC analysis. If a precipitate or phase separation is observed, samples may be centrifuged at low speed (100-400 xg) to force precipitate to the bottom of the vial as a precaution since large amounts of precipitate may clog the HPLC tubing or columns. If a precipitation or phase separation is observed after the incubation period, peptide depletion may be underestimated and a conclusion on the lack of reactivity cannot be drawn with sufficient confidence in case of a negative result.

Preparation of the HPLC standard calibration curve

A standard calibration curve should be generated for both the cysteine and the lysine peptides. Peptide standards should be prepared in a solution of 20 % or 25 % acetonitrile:buffer using phosphate buffer (pH 7,5) for the cysteine peptide and ammonium acetate buffer (pH 10,2) for the lysine peptide. Using serial dilution standards of the peptide stock solution (0,667 mM), 6 calibration solutions should be prepared to cover the range from 0,534 to 0,0167 mM. A blank of the dilution buffer should also be included in the standard calibration curve. Suitable calibration curves should have an $r^2 > 0,99$.

HPLC preparation and analysis

The suitability of the HPLC system should be verified before conducting the analysis. Peptide depletion is monitored by HPLC coupled with an UV detector (photodiode array detector or fixed wavelength absorbance detector with 220 nm signal). The appropriate column is installed in the HPLC system. The HPLC set-up described in the validated protocol uses a Zorbax SB-C-18 2,1 mm \times 100 mm \times 3,5 micron as preferred column. With this reversed-phase HPLC column, the entire system should be equilibrated at 30 °C with 50 % phase A (0,1 % (v/v)

trifluoroacetic acid in water) and 50 % phase B (0,085 % (v/v) trifluoroacetic acid in acetonitrile) for at least 2 hours before running. The HPLC analysis should be performed using a flow rate of 0,35 ml/min and a linear gradient from 10 % to 25 % acetonitrile over 10 minutes, followed by a rapid increase to 90 % acetonitrile to remove other materials. Equal volumes of each standard, sample and control should be injected. The column should be re-equilibrated under initial conditions for 7 minutes between injections. If a different reversed-phase HPLC column is used, the set-up parameters described above may need to be adjusted to guarantee an appropriate elution and integration of the cysteine and lysine peptides, including the injection volume, which may vary according to the system used (typically in the range from 3-10 µl). Importantly, if an alternative HPLC set-up is used, its equivalence to the validated set-up described above should be demonstrated (e.g. by testing the proficiency substances in Appendix 2). Absorbance is monitored at 220 nm. If a photodiode array detector is used, absorbance at 258 nm should also be recorded. It should be noted that some supplies of acetonitrile could have a negative impact on peptide stability and this has to be assessed when a new batch of acetonitrile is used. The ratio of the 220 peak area and the 258 peak area can be used as an indicator of co-elution. For each sample a ratio in the range of 90 % < mean ⁽¹⁾ area ratio of control samples < 100 % would give a good indication that co-elution has not occurred.

There may be test chemicals which could promote the oxidation of the cysteine peptide. The peak of the dimerised cysteine peptide may be visually monitored. If dimerisation appears to have occurred, this should be noted as percent peptide depletion may be over-estimated leading to false positive predictions and/or assignment to a higher reactivity class (see paragraphs 29 and 30).

HPLC analysis for the cysteine and lysine peptides can be performed concurrently (if two HPLC systems are available) or on separate days. If analysis is conducted on separate days then all test chemical solutions should be freshly prepared for both assays on each day. The analysis should be timed to assure that the injection of the first sample starts 22 to 26 hours after the test chemical was mixed with the peptide solution. The HPLC run sequence should be set up in order to keep the HPLC analysis time less than 30 hours. For the HPLC set up used in the validation study and described in this test method, up to 26 analysis samples can be accommodated in a single HPLC run (see also paragraph 17). An example of HPLC analysis sequence is provided in Appendix 3.

DATA AND REPORTING

Data evaluation

The concentration of cysteine or lysine peptide is photometrically determined at 220 nm in each sample by measuring the peak area (area under the curve, AUC) of the appropriate peaks and by calculating the concentration of peptide using the linear calibration curve derived from the standards.

The percent peptide depletion is determined in each sample by measuring the peak area and dividing it by the mean peak area of the relevant reference controls C (see Appendix 3) according to the formula described below.

$$\text{Percent peptide depletion} = \left[1 - \left(\frac{\text{Peptide peak area in replicate injection}}{\text{mean peptide peak area in reference controls C}} \right) \right] \times 100$$

Acceptance criteria

The following criteria should be met for a run to be considered valid:

- (a) the standard calibration curve should have an $r^2 > 0,99$,

⁽¹⁾ For mean it is meant arithmetic mean throughout the document.

- (b) the mean percent peptide depletion value of the three replicates for the positive control cinnamic aldehyde should be between 60,8 % and 100 % for the cysteine peptide and between 40,2 % and 69,0 % for the lysine peptide and the maximum standard deviation (SD) for the positive control replicates should be < 14,9 % for the percent cysteine depletion and < 11,6 % for the percent lysine depletion, and
- (c) the mean peptide concentration of reference controls A should be $0,50 \pm 0,05$ mM and the coefficient of variation (CV) of peptide peak areas for the nine reference controls B and C in acetonitrile should be < 15,0 %.

If one or more of these criteria is not met the run should be repeated.

The following criteria should be met for a test chemical's results to be considered valid:

- (a) the maximum standard deviation for the test chemical replicates should be < 14,9 % for the percent cysteine depletion and < 11,6 % for the percent lysine depletion,
- (b) the mean peptide concentration of the three reference controls C in the appropriate solvent should be $0,50 \pm 0,05$ mM. If these criteria are not met the data should be rejected and the run should be repeated for that specific test chemical.

Prediction model

The mean percent cysteine and percent lysine depletion value is calculated for each test chemical. Negative depletion is considered as "0" when calculating the mean. By using the cysteine 1:10/lysine 1:50 prediction model shown in Table 1, the threshold of 6,38 % average peptide depletion should be used to support the discrimination between skin sensitisers and non-sensitisers in the framework of an IATA. Application of the prediction model for assigning a test chemical to a reactivity class (i.e. low, moderate and high reactivity) may perhaps prove useful to inform potency assessment within the framework of an IATA.

Table 1

Cysteine 1:10/lysine 1:50 prediction model ⁽¹⁾

Mean of cysteine and lysine % depletion	Reactivity Class	DPRA Prediction ⁽²⁾
$0 \% \leq \text{mean \% depletion} \leq 6,38 \%$	No or minimal reactivity	Negative
$6,38 \% < \text{mean \% depletion} \leq 22,62 \%$	Low reactivity	Positive
$22,62 \% < \text{mean \% depletion} \leq 42,47 \%$	Moderate reactivity	
$42,47 \% < \text{mean \% depletion} \leq 100 \%$	High reactivity	

⁽¹⁾ The numbers refer to statistically generated threshold values and are not related to the precision of the measurement.

⁽²⁾ A DPRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 9 and 12.

There might be cases where the test chemical (the substance or one or several of the components of a multi-constituent substance or a mixture) absorbs significantly at 220 nm and has the same retention time of the peptide (co-elution). Co-elution may be resolved by slightly adjusting the HPLC set-up in order to further separate the elution time of the test chemical and the peptide. If an alternative HPLC set-up is used to try to resolve co-elution, its equivalence to the validated set-up should be demonstrated (e.g. by testing the proficiency substances in Appendix 2). When co-elution occurs the peak of the peptide cannot be integrated and the calculation of the percent peptide depletion is not possible. If co-elution of such test chemicals occurs with both the cysteine and the lysine peptides then the analysis should be reported as “inconclusive”. In cases where co-elution occurs only with the lysine peptide, then the cysteine 1:10 prediction model reported in Table 2 can be used.

Table 2

Cysteine 1:10 prediction model ⁽¹⁾

Cysteine (Cys) % depletion	Reactivity class	DPRA prediction ⁽²⁾
0 % ≤ Cys % depletion ≤ 13,89 %	No or minimal reactivity	Negative
13,89 % < Cys % depletion ≤ 23,09 %	Low reactivity	Positive
23,09 % < Cys % depletion ≤ 98,24 %	Moderate reactivity	
98,24 % < Cys % depletion ≤ 100 %	High reactivity	

⁽¹⁾ The numbers refer to statistically generated threshold values and are not related to the precision of the measurement.

⁽²⁾ A DPRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 9 and 12.

There might be other cases where the overlap in retention time between the test chemical and either of the peptides is incomplete. In such cases percent peptide depletion values can be estimated and used in the cysteine 1:10/lysine 1:50 prediction model, however assignment of the test chemical to a reactivity class cannot be made with accuracy.

A single HPLC analysis for both the cysteine and the lysine peptide should be sufficient for a test chemical when the result is unequivocal. However, in cases of results close to the threshold used to discriminate between positive and negative results (i.e. borderline results), additional testing may be necessary. If situations where the mean percent depletion falls in the range of 3 % to 10 % for the cysteine 1:10/lysine 1:50 prediction model or the cysteine percent depletion falls in the range of 9 % to 17 % for the cysteine 1:10 prediction model, a second run should be considered, as well as a third one in case of discordant results between the first two runs.

Test report

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, 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);
 - Concentration(s) tested;
 - 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 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);
 - Concentration(s) tested;
 - Storage conditions and stability to the extent available.

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;
 - 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.
- Solvent/vehicle
- Solvent/vehicle used and ratio of its constituents, if applicable;
 - Chemical identification(s), 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 solvents / vehicles than those mentioned in the test method are used and to the extent available;
 - Storage conditions and stability to the extent available;

- Justification for choice of solvent for each test chemical;
- For acetonitrile, results of test of impact on peptide stability.

Preparation of peptides, positive control and test chemical

- Characterisation of peptide solutions (supplier, lot, exact weight of peptide, volume added for the stock solution);
- Characterisation of positive control solution (exact weight of positive control substance, volume added for the test solution);
- Characterisation of test chemical solutions (exact weight of test chemical, volume added for the test solution).

HPLC instrument setting and analysis

- Type of HPLC instrument, HPLC and guard columns, detector, autosampler;
- Parameters relevant for the HPLC analysis such as column temperature, injection volumes, flow rate and gradient.

System suitability

- Peptide peak area at 220 nm of each standard and reference control A replicate;
- Linear calibration curve graphically represented and the r^2 reported;
- Peptide concentration of each reference control A replicate;
- Mean peptide concentration (mM) of the three reference controls A, SD and CV;
- Peptide concentration of reference controls A and C.

Analysis sequence

- For reference controls:
 - Peptide peak area at 220 nm of each B and C replicate;
 - Mean peptide peak area at 220 nm of the nine reference controls B and C in acetonitrile, SD and CV (for stability of reference controls over analysis time);
 - For each solvent used, the mean peptide peak area at 220 nm of the three appropriate reference controls C (for the calculation of percent peptide depletion);
 - For each solvent used, the peptide concentration (mM) of the three appropriate reference controls C;
 - For each solvent used, the mean peptide concentration (mM) of the three appropriate reference controls C, SD and CV.
- For positive control:
 - Peptide peak area at 220 nm of each replicate;
 - Percent peptide depletion of each replicate;
 - Mean percent peptide depletion of the three replicates, SD and CV.
- For each test chemical:
 - Appearance of precipitate in the reaction mixture at the end of the incubation time, if observed. If precipitate was re-solubilised or centrifuged;

- Presence of co-elution;
- Description of any other relevant observations, if applicable;
- Peptide peak area at 220 nm of each replicate;
- Percent peptide depletion of each replicate;
- Mean of percent peptide depletion of the three replicate, SD and CV;
- Mean of percent cysteine and percent lysine depletion values;
- Prediction model used and DPRA prediction.

Proficiency testing

- If applicable, the procedure used to demonstrate proficiency of the laboratory in performing the test method (e.g. by testing of proficiency substances) or to demonstrate reproducible performance of the test method over time.

Discussion of the results

- Discussion of the results obtained with the DPRA test method;
- Discussion of the test method results in the context of an IATA if other relevant information is available.

Conclusion

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) 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, OECD, Paris.
- (3) Chapter B.6 of this Annex: Skin Sensitisation.
- (4) Chapter B.42 of this Annex: The Local Lymph Node Assay
- (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) Adler *et al.* (2011). Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Archives of Toxicology 85:367-485.
- (8) Gerberick *et al.* (2004). Development of a peptide reactivity assay for screening contact allergens. Toxicological Sciences 81:332-343.
- (9) Gerberick *et al.* (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. Toxicological Sciences 97:417-427.
- (10) EC EURL-ECVAM (2013). Recommendation on the Direct Peptide Reactivity Assay (DPRA) for skin sensitisation testing. Available at: <https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/eurl-ecvam-recommendation-on-the-direct-peptide-reactivity-assay-dpra>
- (11) Jaworska *et al.* (2013). Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. Journal of Applied Toxicology, published online, 14 May 2013, DOI: 10.1002/jat.2869.

-
- (12) Bauch *et al.* (2012). Putting the parts together: combining *in vitro* methods to test for skin sensitizing potential. *Regulatory Toxicology and Pharmacology* 63: 489-504.
- (13) Nukada *et al.* (2013). Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. *Toxicology in vitro* 27:609 618.
- (14) Ball *et al.* (2011). Evaluating the sensitization potential of surfactants: integrating data from the local lymph node assay, guinea pig maximization test, and *in vitro* methods in a weight-of-evidence approach. *Regulatory Toxicology and Pharmacology* 60:389-400.
- (15) Landsteiner and Jacobs (1936). Studies on the sensitization of animals with simple chemical compounds. *Journal of Experimental Medicine* 64:625-639.
- (16) Dupuis and Benezra (1982). Allergic contact dermatitis to simple chemicals: a molecular approach. New York & Basel: Marcel Dekker Inc.
- (17) Lepoittevin *et al.* (1998). Allergic contact dermatitis: the molecular basis. Springer, Berlin.
- (18) EC EURL ECVAM (2012). Direct Peptide Reactivity Assay (DPRA) Validation Study Report 74pp. Accessible at: http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations/eurl-ecvam-recommendation-on-the-direct-peptide-reactivity-assay-dpra
- (19) Natsch *et al.* (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *Journal of Applied Toxicology*, published online, 9 April 2013, DOI:10.1002/jat.2868.
- (20) DB-ALM (INVITTOX) Protocol 154: Direct Peptide Reactivity assay (DPRA) for skin sensitisation testing 17pp. Accessible at: <http://ecvam-dbalm.jrc.ec.europa.eu/>
- (21) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Series on Testing and Assessment, No. 34. Organisation for Economic Cooperation and Development, Paris, France.
- (22) FDA (Food and Drug Administration (2001). Guidance for Industry: Bioanalytical Method Validation 22pp. Accessible at: www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidance/ucm070107.pdf - 138
- (23) ECETOC (2003). Contact sensitization: Classification according to potency. European Centre for Ecotoxicology & Toxicology of Chemicals (Technical Report No. 87).
-

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 (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 (2).

Calibration curve: The relationship between the experimental response value and the analytical concentration (also called *standard curve*) of a known substance.

Chemical: A substance or a mixture.

Coefficient of variation: A measure of variability that is calculated for a group of replicate data by dividing the standard deviation by the mean. It can be multiplied by 100 for expression as a percentage.

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.

Molecular Initiating Event: Chemical-induced perturbation of a biological system at the molecular level identified to be the starting event in the adverse outcome pathway.

Mixture: A mixture or a solution composed of two or more substances in which they do not react (1).

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.

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.

Reference 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 negative control, this sample also demonstrates whether the solvent or vehicle interacts with the test system.

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 (21).

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 (21).

Reproducibility: The agreement among results obtained from testing the same chemical using the same test protocol (see reliability) (21).

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 (21).

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 (21).

Substance: Chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the 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 (1).

System suitability: Determination of instrument performance (e.g. sensitivity) by analysis of a reference standard prior to running the analytical batch (22).

Test chemical: The term “test chemical” is used to refer to what is being tested.

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).

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 (21).

Appendix 2

PROFICIENCY SUBSTANCES

***In Chemico* Skin Sensitisation: Direct Peptide Reactivity Assay**

Prior to routine use of this test method, laboratories should demonstrate technical proficiency by correctly obtaining the expected DPRA prediction for the 10 proficiency substances recommended in Table 1 and by obtaining cysteine and lysine depletion values that fall within the respective reference range for 8 out of the 10 proficiency substances for each peptide. These proficiency substances were selected to represent the range of responses for skin sensitisation hazards. Other selection criteria were that they are commercially available, that high quality *in vivo* reference data and high quality *in vitro* data generated with the DPRA are available, and that they were used in the EURL ECVAM-coordinated validation study to demonstrate successful implementation of the test method in the laboratories participating in the study.

Table 1

Recommended proficiency substances for demonstrating technical proficiency with the Direct Peptide Reactivity Assay

Proficiency substances	CASRN	Physical state	<i>In vivo</i> prediction ⁽¹⁾	DPRA prediction ⁽²⁾	Range ⁽³⁾ of % cysteine peptide depletion	Range ⁽³⁾ of % lysine peptide depletion
2,4-Dinitrochlorobenzene	97-00-7	Solid	Sensitiser (extreme)	Positive	90-100	15-45
Oxazolone	15646-46-5	Solid	Sensitiser (extreme)	Positive	60-80	10-55
Formaldehyde	50-00-0	Liquid	Sensitiser (strong)	Positive	30-60	0-24
Benzylideneacetone	122-57-6	Solid	Sensitiser (moderate)	Positive	80-100	0-7
Farnesal	19317-11-4	Liquid	Sensitiser (weak)	Positive	15-55	0-25
2,3-Butanedione	431-03-8	Liquid	Sensitiser (weak)	Positive	60-100	10-45
1-Butanol	71-36-3	Liquid	Non-sensitizer	Negative	0-7	0-5,5
6-Methylcoumarin	92-48-8	Solid	Non-sensitizer	Negative	0-7	0-5,5

Proficiency substances	CASRN	Physical state	<i>In vivo</i> prediction ⁽¹⁾	DPRA prediction ⁽²⁾	Range ⁽³⁾ of % cysteine peptide depletion	Range ⁽³⁾ of % lysine peptide depletion
Lactic Acid	50-21-5	Liquid	Non-sensitizer	Negative	0-7	0-5,5
4-Methoxyacetophenone	100-06-1	Solid	Non-sensitizer	Negative	0-7	0-5,5

⁽¹⁾ The *in vivo* hazard and (potency) predictions are based on LLNA data (19). The *in vivo* potency is derived using the criteria proposed by ECETOC (23).

⁽²⁾ A DPRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 9 and 11.

⁽³⁾ Ranges determined on the basis of at least 10 depletion values generated by 6 independent laboratories.

Appendix 3

EXAMPLES OF ANALYSIS SEQUENCE

Calibration standards and reference controls	STD1 STD2 STD3 STD4 STD5 STD6 Dilution buffer Reference control A, rep 1 Reference control A, rep 2 Reference control A, rep 3
Co-elution controls	Co-elution control 1 for test chemical 1 Co-elution control 2 for test chemical 2
Reference controls	Reference control B, rep 1 Reference control B, rep 2 Reference control B, rep 3
First set of replicates	Reference control C, rep 1 Cinnamic aldehyde, rep 1 Sample 1, rep 1 Sample 2, rep 1
Second set of replicates	Reference control C, rep 2 Cinnamic aldehyde, rep 2 Sample 1, rep 2 Sample 2, rep 2
Third set of replicates	Reference control C, rep 3 Cinnamic aldehyde, rep 3 Sample 1, rep 3 Sample 2, rep 3
Reference controls	Reference control B, rep 4 Reference control B, rep 5 Reference control B, rep 6

Three sets of reference controls (i.e. samples constituted only by the peptide dissolved in the appropriate solvent) should be included in the analysis sequence:

Reference control A: used to verify the suitability of the HPLC system.

Reference control B: included at the beginning and at the end of the analysis sequence to verify stability of reference controls over the analysis time.

Reference control C: included in the analysis sequence to verify that the solvent used to dissolve the test chemical does not impact the percent peptide depletion.

B.60 *In Vitro* Skin Sensitisation: ARE-Nrf2 Luciferase Test Method**INTRODUCTION**

This test method (TM) is equivalent to OECD test guideline (TG) 442D (2015). 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 Regulation (EC) No 1272/2008 of the European Parliament and of the Council on Classification, Labelling and Packaging of Substances and Mixtures (CLP) ⁽¹⁾. This test method provides an *in vitro* procedure (the ARE-Nrf2 luciferase assay) to be used for supporting the discrimination between skin sensitizers and non-sensitizers in accordance with the UN GHS (1) and CLP.

There is general agreement regarding the key biological events underlying skin sensitization. The existing knowledge of the chemical and biological mechanisms associated with skin sensitization has been summarised in the form of an Adverse Outcome Pathway (AOP) (2), going from the molecular initiating event through the intermediate events up to the adverse health effect, i.e. allergic contact dermatitis in humans or contact hypersensitivity in rodents (2) (3). The molecular initiating 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 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, typically assessed by expression of specific cell surface markers, chemokines and cytokines. The fourth key event is T-cell proliferation, which is indirectly assessed in the murine Local Lymph Node Assay (4).

The assessment of skin sensitization has typically involved the use of laboratory animals. The classical methods based on guinea-pigs, the Magnusson Kligman Guinea Pig Maximisation Test (GMPT) and the Buehler Test (TM B.6 (5)), study both the induction and elicitation phases of skin sensitization. A murine test, the Local Lymph Node Assay (LLNA) (TM B.42 (4)) and its two non-radioactive modifications, LLNA: DA (TM B.50 (6)) and LLNA: BrdU-ELISA (TM B.51 (7)), which all assess the induction response exclusively, have also gained acceptance since they provide advantages over the guinea pig tests in terms of both animal welfare and objective measurement of the induction phase of skin sensitization.

More recently, mechanistically-based *in chemico* and *in vitro* test methods have been considered scientifically valid for the evaluation of the skin sensitization hazard of chemicals. However, combinations of non-animal methods (*in silico*, *in chemico*, *in vitro*) within Integrated Approaches to Testing and Assessment (IATA) will be needed to be able to fully substitute for the animal tests currently in use given the restricted AOP mechanistic coverage of each of the currently available non-animal test methods (2) (3).

This test method (ARE-Nrf2 luciferase assay) is proposed to address the second key event as explained in paragraph 2. Skin sensitizers have been reported to induce genes that are regulated by the antioxidant response element (ARE) (8) (9). Small electrophilic substances such as skin sensitizers can act on the sensor protein Keap1 (Kelch-like ECH-associated protein 1), by e.g. covalent modification of its cysteine residue, resulting in its dissociation from the transcription factor Nrf2 (nuclear factor-erythroid 2-related factor 2). The dissociated Nrf2 can then activate ARE-dependent genes such as those coding for phase II detoxifying enzymes (8) (10) (11).

Currently, the only *in vitro* ARE-Nrf2 luciferase assay covered by this test method is the KeratinoSens™ assay for which validation studies have been completed (9) (12) (13) followed by an independent peer review conducted by the European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) (14). The KeratinoSens™ assay was considered scientifically valid to be used as part of an IATA, to support the discrimination between skin sensitizers and non-sensitizers for the purpose of hazard classification and labelling (14). Laboratories willing to implement the test method can obtain the recombinant cell line used in the KeratinoSens™ assay by establishing a licence agreement with the test method developer (15).

⁽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, 31.12.2008, p. 1).

Definitions are provided in Appendix 1.

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

Since activation of the Keap1-Nrf2-ARE pathway addresses only the second key event of the skin sensitisation AOP, information from test methods based on the activation of this pathway is unlikely to be sufficient when used on its own to conclude on the skin sensitisation potential of chemicals. Therefore, data generated with the present test method should be considered in the context of integrated approaches, such as IATA, combining them with other 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. Examples on how to use the ARE-Nrf2 luciferase test method in combination with other information are reported in literature (13) (16) (17) (18) (19).

This test method can be used to support the discrimination between skin sensitisers (i.e. UN GHS/CLP Category 1) and non-sensitisers in the context of IATA. This test method cannot be used on its own, neither to sub-categorise skin sensitisers into subcategories 1A and 1B as defined by the UN GHS/CLP nor to predict potency for safety assessment decisions. However, depending on the regulatory framework, a positive result may be used on its own to classify a chemical into UN GHS/CLP category 1.

Based on the dataset from the validation study and in-house testing used for the independent peer-review of the test method, the KeratinoSens™ assay proved to be transferable to laboratories experienced in cell culture. The level of reproducibility in predictions that can be expected from the test method is in the order of 85 % within and between laboratories (14). The accuracy (77 % - 155/201), sensitivity (78 % - 71/91) and specificity (76 % - 84/110) of the KeratinoSens™ assay for discriminating skin sensitisers (i.e. UN GHS/CLP Cat. 1) from non-sensitisers when compared to LLNA results were calculated by considering all of the data submitted to EURL ECVAM for evaluation and peer-review of the test method (14). These figures are similar to those recently published based on in-house testing of about 145 substances (77 % accuracy, 79 % sensitivity, 72 % specificity) (13). The KeratinoSens™ assay is more likely to under predict 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) (13) (14). Taken together, this information indicates the usefulness of the KeratinoSens™ assay to contribute to the identification of skin sensitisation hazard. However, the accuracy values given here for the KeratinoSens™ assay as a stand-alone test method are only indicative since the test method should be considered in combination with other sources of information in the context of an IATA and in accordance with the provisions of paragraph 9 above. Furthermore when evaluating non-animal methods for skin sensitisation, it should be kept in mind that the LLNA as well as other animal tests, may not fully reflect the situation in the species of interest i.e. humans.

The term “test chemical” is used in this test method to refer to what is being tested and is not related to the applicability of the ARE-Nrf2 luciferase test method to the testing of substances and/or mixtures. On the basis of the current data available the KeratinoSens™ assay was shown to be applicable to test chemicals covering a variety of organic functional groups, reaction mechanisms, skin sensitisation potency (as determined with *in vivo* studies) and physico-chemical properties (9) (12) (13) (14). Mainly mono-constituent substances were tested, although a limited amount of data also exist on the testing of mixtures (20). 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. Moreover, when testing multi-constituent substances or mixtures, consideration should be given to possible interference of cytotoxic constituents with the observed responses. The test 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 into different phases) either in water or DMSO (including all of the test chemical components in the case of testing a multi-constituent substance or a mixture). Test chemicals that do not

fulfil these conditions at the highest final required concentration of 2 000 µM (cf. paragraph 22) may still be tested at lower concentrations. In such a case, results fulfilling the criteria for positivity described in paragraph 39 could still be used to support the identification of the test chemical as a skin sensitiser, whereas a negative result obtained with concentrations < 1 000 µM should be considered as inconclusive (see prediction model in paragraph 39). In general substances with a LogP of up to 5 have been successfully tested whereas extremely hydrophobic substances with a LogP above 7 are outside the known applicability of the test method (14). For substances having a LogP falling between 5 and 7, only limited information is available.

Negative results should be interpreted with caution as substances with an exclusive reactivity towards lysine-residues can be detected as negative by the test method. Furthermore, because of the limited metabolic capability of the cell line used (21) and because of the experimental conditions, pro-haptens (i.e. chemicals requiring enzymatic activation for example via P450 enzymes) and pre-haptens (i.e. chemicals activated by auto-oxidation) in particular with a slow oxidation rate may also provide negative results. Test chemicals that do not act as a sensitiser but are nevertheless chemical stressors may lead on the other hand to false positive results (14). Furthermore, highly cytotoxic test chemicals cannot always be reliably assessed. Finally, test chemicals that interfere with the luciferase enzyme can confound the activity of luciferase in cell-based assays causing either apparent inhibition or increased luminescence (22). For example, phytoestrogen concentrations higher than 1 µM were reported to interfere with the luminescence signals in other luciferase-based reporter gene assays due to over-activation of the luciferase reporter gene (23). As a consequence, luciferase expression obtained at high concentrations of phytoestrogens or similar chemicals suspected of producing phytoestrogen-like over-activation of the luciferase reporter gene needs to be examined carefully (23). In cases where evidence can be demonstrated on the non-applicability of the test method to other specific categories of test chemicals, the test method should not be used for those specific categories.

In addition to supporting discrimination between skin sensitisers and non-sensitisers, the KeratinoSens™ assay also provides concentration-response information that may potentially contribute to the assessment of sensitising potency when used in integrated approaches such as IATA (19). However, further work preferably based on reliable human data is required to determine how KeratinoSens™ assay results can contribute to potency assessment (24) and to sub-categorisation of sensitisers according to UN GHS/CLP.

PRINCIPLE OF THE TEST

The ARE-Nrf2 luciferase test method makes use of an immortalised adherent cell line derived from HaCaT human keratinocytes stably transfected with a selectable plasmid. The cell line contains the luciferase gene under the transcriptional control of a constitutive promoter fused with an ARE element from a gene that is known to be up-regulated by contact sensitisers (25) (26). The luciferase signal reflects the activation by sensitisers of endogenous Nrf2 dependent genes, and the dependence of the luciferase signal in the recombinant cell line on Nrf2 has been demonstrated (27). This allows quantitative measurement (by luminescence detection) of luciferase gene induction, using well established light producing luciferase substrates, as an indicator of the activity of the Nrf2 transcription factor in cells following exposure to electrophilic substances.

Test chemicals are considered positive in the KeratinoSens™ assay if they induce a statistically significant induction of the luciferase activity above a given threshold (i.e. > 1,5 fold or 50 % increase), below a defined concentration which does not significantly affect cell viability (i.e. below 1 000 µM and at a concentration at which the cellular viability is above 70 % (9) (12)). For this purpose, the maximal fold induction of the luciferase activity over solvent (negative) control (I_{\max}) is determined. Furthermore, since cells are exposed to series of concentrations of the test chemicals, the concentration needed for a statistically significant induction of luciferase activity above the threshold (i.e. $EC_{1,5}$ value) should be interpolated from the dose-response curve (see paragraph 32 for calculations). Finally, parallel cytotoxicity measurements should be conducted to assess whether luciferase activity induction levels occur at sub-cytotoxic concentrations.

Prior to routine use of the ARE-Nrf2 luciferase assay that adheres to this test method, laboratories should demonstrate technical proficiency, using the ten Proficiency Substances listed in Appendix 2.

Performance standards (PS) (28) are available to facilitate the validation of new or modified *in vitro* ARE-Nrf2 luciferase test methods similar to the KeratinoSens™ assay and allow for timely amendment of this test method for their inclusion. Mutual Acceptance of Data (MAD) according to the OECD agreement will only be guaranteed for test methods validated according to the PS, if these test methods have been reviewed and included in the corresponding test guideline by OECD.

PROCEDURE

Currently, the only method covered by this test method is the scientifically valid KeratinoSens™ assay (9) (12) (13) (14). The Standard Operating Procedures (SOP) for the KeratinoSens™ assay is available and should be employed when implementing and using the test method in the laboratory (15). Laboratories willing to implement the test method can obtain the recombinant cell line used in the KeratinoSens™ assay by establishing a licence agreement with the test method developer. The following paragraphs provide with a description of the main components and procedures of the ARE-Nrf2 luciferase test method.

Preparation of the keratinocyte cultures

A transgenic cell line having a stable insertion of the luciferase reporter gene under the control of the ARE-element should be used (e.g. the KeratinoSens™ cell line). Upon receipt, cells are propagated (e.g. 2 to 4 passages) and stored frozen as a homogeneous stock. Cells from this original stock can be propagated up to a maximum passage number (i.e. 25 in the case of KeratinoSens™) and are employed for routine testing using the appropriate maintenance medium (in the case of KeratinoSens™ this represents DMEM containing serum and Geneticin).

For testing, cells should be 80-90 % confluent, and care should be taken to ensure that cells are never grown to full confluence. One day prior to testing cells are harvested, and distributed into 96-well plates (10 000 cells/well in the case of KeratinoSens™). Attention should be paid to avoid sedimentation of the cells during seeding to ensure homogeneous cell number distribution across wells. If this is not the case, this step may give rise to high well-to-well variability. For each repetition, three replicates are used for the luciferase activity measurements, and one parallel replicate used for the cell viability assay.

Preparation of the test chemical and control substances

The test chemical and control substances are prepared on the day of testing. For the KeratinoSens™ assay, test chemicals are dissolved in dimethyl sulfoxide (DMSO) to the final desired concentration (e.g. 200 mM). The DMSO solutions can be considered self-sterilising, so that no sterile filtration is needed. Test chemical not soluble in DMSO is dissolved in sterile water or culture medium, and the solutions sterilised by e.g. filtration. For a test chemical which has no defined molecular weight (MW), a stock solution is prepared to a default concentration (40 mg/mL or 4 % (w/v)) in the KeratinoSens™ assay. In case solvents other than DMSO, water or the culture medium are used, sufficient scientific rationale should be provided.

Based on the stock DMSO solutions of the test chemical, serial dilutions are made using DMSO to obtain 12 master concentrations of the chemical to be tested (from 0,098 to 200 mM in the KeratinoSens™ assay). For a test chemical not soluble in DMSO, the dilutions to obtain the master concentrations are made using sterile water or sterile culture medium. Independent of the solvent used, the master concentrations, are then further diluted 25 fold into culture medium containing serum, and finally used for treatment with a further 4 fold dilution factor so that the final concentrations of the tested chemical range from 0,98 to 2 000 µM in the KeratinoSens™ assay. Alternative concentrations may be used upon justification (e.g. in case of cytotoxicity or poor solubility).

The negative (solvent) control used in the KeratinoSens™ assay is DMSO (CAS No. 67-68-5, ≥ 99 % purity), for which six wells per plate are prepared. It undergoes the same dilution as described for the master concentrations in paragraph 22, so that the final negative (solvent) control concentration is 1 %, known not to affect cell viability and corresponding to the same concentration of DMSO found in the tested chemical and in the positive control. For a test chemical not soluble in DMSO, for which the dilutions were made in water, the DMSO level in all wells of the final test solution must be adjusted to 1 % as for the other test chemicals and control substances.

The positive control used in the case of the KeratinoSens™ assay is cinnamic aldehyde (CAS No. 14371-10-9, ≥ 98 % purity), for which a series of 5 master concentrations ranging from 0,4 to 6,4 mM are prepared in DMSO (from a 6,4 mM stock solution) and diluted as described for the master concentrations in paragraph 22, so that the final concentration of the positive control range from 4 to 64 µM. Other suitable positive controls, preferentially providing EC_{1,5} values in the mid-range, may be used if historical data are available to derive comparable run acceptance criteria.

Application of the test chemical and control substances

For each test chemical and positive control substance, one experiment is needed to derive a prediction (positive or negative), consisting of at least two independent repetitions containing each three replicates (i.e. n = 6). In case of discordant results between the two independent repetitions, a third repetition containing three replicates should be performed (i.e. n = 9). Each independent repetition is performed on a different day with fresh stock solution of test chemicals and independently harvested cells. Cells may come from the same passage however.

After seeding as described in paragraph 20, cells are grown for 24 hours in the 96-wells microtiter plates. The medium is then removed and replaced with fresh culture medium (150 µl culture medium containing serum but without Geneticin in the case of KeratinoSens™) to which 50 µl of the 25 fold diluted test chemical and control substances are added. At least one well per plate should be left empty (no cells and no treatment) to assess background values.

The treated plates are then incubated for about 48 hours at 37 ± 1 °C in the presence of 5 % CO₂ in the KeratinoSens™ assay. Care should be taken to avoid evaporation of volatile test chemicals and cross-contamination between wells by test chemicals by e.g. covering the plates with a foil prior to the incubation with the test chemicals.

Luciferase activity measurements

Three factors are critical to ensure appropriate luminescence readings:

- the choice of a sensitive luminometer,
- the use of a plate format with sufficient height to avoid light-cross-contamination; and
- the use of a luciferase substrate with sufficient light output to ensure sufficient sensitivity and low variability.

Prior to testing, a control experiment setup as described in Appendix 3 should be carried out to ensure that these three points are met.

After the 48 hour exposure time with the test chemical and control substances in the KeratinoSens™ assay, cells are washed with a phosphate buffered saline, and the relevant lysis buffer for luminescence readings added to each well for 20 min at room temperature.

Plates with the cell lysate are then placed in the luminometer for reading which in the KeratinoSens™ assay is programmed to: (i) add the luciferase substrate to each well (i.e. 50 µl), (ii) wait for 1 second, and (iii) integrate the luciferase activity for 2 seconds. In case alternative settings are used, e.g. depending on the model of luminometer used, these should be justified. Furthermore, a glow substrate may also be used provided that the quality control experiment of Appendix 3 is successfully fulfilled.

Cytotoxicity Assessment

For the KeratinoSens™ cell viability assay, medium is replaced after the 48 hour exposure time with fresh medium containing MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS No. 298-93-1) and cells incubated for 4 hours at 37 °C in the presence of 5 % CO₂. The MTT medium is then removed and cells are lysed (e.g. by adding 10 % SDS solution to each well) overnight. After shaking, the absorption is measured at i.e. 600 nm with a photometer.

DATA AND REPORTING

Data evaluation

The following parameters are calculated in the KeratinoSens™ assay:

- the maximal average fold induction of luciferase activity (I_{\max}) value observed at any concentration of the tested chemical and positive control;
- the $EC_{1,5}$ value representing the concentration for which induction of luciferase activity is above the 1,5 fold threshold (i.e. 50 % enhanced luciferase activity) was obtained; and
- the IC_{50} and IC_{30} concentration values for 50 % and 30 % reduction of cellular viability.
- Fold luciferase activity induction is calculated by Equation 1, and the overall maximal fold induction (I_{\max}) is calculated as the average of the individual repetitions.

Equation 1:

$$\text{Fold induction} = \frac{(L_{\text{sample}} - L_{\text{blank}})}{(L_{\text{solvent}} - L_{\text{blank}})}$$

where

L_{sample} is the luminescence reading in the test chemical well

L_{blank} is the luminescence reading in the blank well containing no cells and no treatment

L_{solvent} is the average luminescence reading in the wells containing cells and solvent (negative) control

$EC_{1,5}$ is calculated by linear interpolation according to Equation 2, and the overall $EC_{1,5}$ is calculated as the geometric mean of the individual repetitions.

Equation 2:

$$EC_{1,5} = (C_b - C_a) \times \left(\frac{1,5 - I_a}{I_b - I_a} \right) + C_a$$

where

C_a is the lowest concentration in μM with $> 1,5$ fold induction

C_b is the highest concentration in μM with $< 1,5$ fold induction

I_a is the fold induction measured at the lowest concentration with $> 1,5$ fold induction (mean of three replicate wells)

I_b is the fold induction at the highest concentration with $< 1,5$ fold induction (mean of three replicate wells)

Viability is calculated by Equation 3:

Equation 3:

$$\text{Viability} = \frac{(V_{\text{sample}} - V_{\text{blank}})}{V_{\text{solvent}} - V_{\text{blank}}} \times 100$$

where

V_{sample} is the MTT-absorbance reading in the test chemical well

V_{blank} is the MTT-absorbance reading in the blank well containing no cells and no treatment

V_{solvent} is the average MTT-absorbance reading in the wells containing cells and solvent (negative) control

IC_{50} and IC_{30} are calculated by linear interpolation according to Equation 4, and the overall IC_{50} and IC_{30} are calculated as the geometric mean of the individual repetitions.

Equation 4:

$$IC_x = (C_b - C_a) \times \left(\frac{(100 - x) - V_a}{V_b - V_a} \right) + C_a$$

where

X is the % reduction at the concentration to be calculated (50 and 30 for IC_{50} and IC_{30})

C_a is the lowest concentration in μM with $> x$ % reduction in viability

C_b is the highest concentration in μM with $< x$ % reduction in viability

V_a is the % viability at the lowest concentration with $> x$ % reduction in viability

V_b is the % viability at the highest concentration with $< x$ % reduction in viability

For each concentration showing $> 1,5$ fold luciferase activity induction, statistical significance is calculated (e.g. by a two-tailed Student's t-test), comparing the luminescence values for the three replicate samples with the luminescence values in the solvent (negative) control wells to determine whether the luciferase activity induction is statistically significant ($p < 0,05$). The lowest concentration with $> 1,5$ fold luciferase activity induction is the value determining the $EC_{1,5}$ value. It is checked in each case whether this value is below the IC_{30} value, indicating that there is less than 30 % reduction in cellular viability at the $EC_{1,5}$ determining concentration.

It is recommended that data are visually checked with the help of graphs. If no clear dose-response curve is observed, or if the dose-response curve obtained is biphasic (i.e. crossing the threshold of 1,5 twice), the experiment should be repeated to verify whether this is specific to the test chemical or due to an experimental artefact. In case the biphasic response is reproducible in an independent experiment, the lower $EC_{1,5}$ value (the concentration when the threshold of 1,5 is crossed the first time) should be reported.

In the rare cases where a statistically non-significant induction above 1,5 fold is observed followed by a higher concentration with a statistically significant induction, results from this repetition are only considered as valid and positive if the statistically significant induction above the threshold of 1,5 was obtained for a non-cytotoxic concentration.

Finally, for test chemicals generating a 1,5 fold or higher induction already at the lowest test concentration of 0,98 μM , the $EC_{1,5}$ value of < 0,98 is set based on visual inspection of the dose-response curve.

Acceptance criteria

The following acceptance criteria should be met when using the KeratinoSens™ assay. First, the luciferase activity induction obtained with the positive control, cinnamic aldehyde, should be statistically significant above the threshold of 1,5 (e.g. using a T-test) in at least one of the tested concentrations (from 4 to 64 μM).

Second, the $EC_{1,5}$ value should be within two standard deviations of the historical mean of the testing facility (e.g. between 7 μM and 30 μM based on the validation dataset) which should be regularly updated. In addition, the average induction in the three replicates for cinnamic aldehyde at 64 μM should be between 2 and 8. If the latter criterion is not fulfilled, the dose-response of cinnamic aldehyde should be carefully checked, and tests may be accepted only if there is a clear dose-response with increasing luciferase activity induction at increasing concentrations for the positive control.

Finally, the average coefficient of variation of the luminescence reading for the negative (solvent) control DMSO should be below 20 % in each repetition which consists of 6 wells tested in triplicate. If the variability is higher, results should be discarded.

Interpretation of results and prediction model

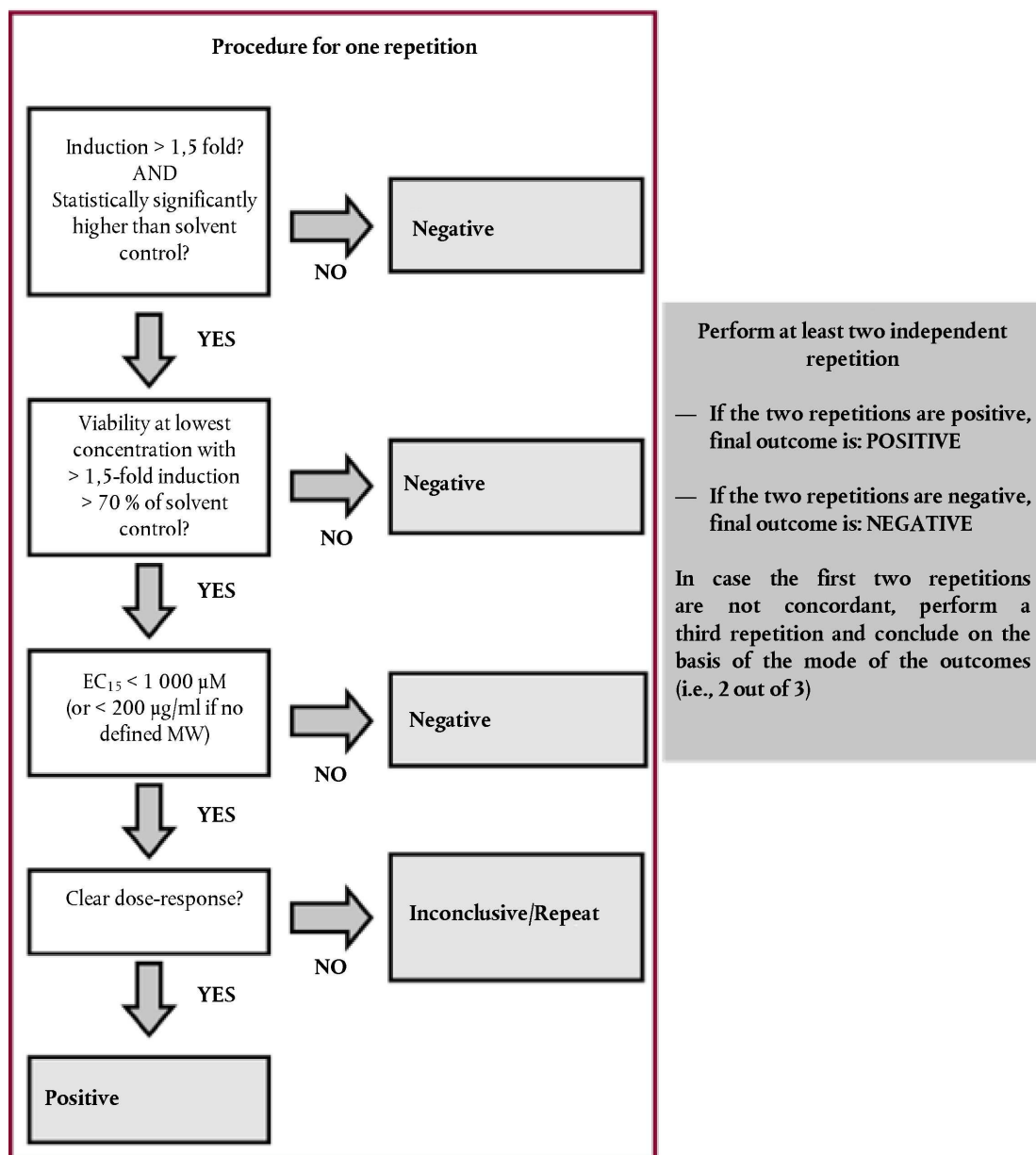
A KeratinoSens™ prediction is considered positive if the following 4 conditions are all met in 2 of 2 or in the same 2 of 3 repetitions, otherwise the KeratinoSens™ prediction is considered negative (Figure 1):

1. the I_{max} is higher than (>) 1,5 fold and statistically significantly different as compared to the solvent (negative) control (as determined by a two-tailed, unpaired Student's t-test);
2. the cellular viability is higher than (>) 70 % at the lowest concentration with induction of luciferase activity above 1,5 fold (i.e. at the $EC_{1,5}$ determining concentration);
3. the $EC_{1,5}$ value is less than (<) 1 000 μM (or < 200 $\mu\text{g/ml}$ for test chemicals with no defined MW);
4. there is an apparent overall dose-response for luciferase induction (or a biphasic response as mentioned under paragraph 33).

If in a given repetition, all of the three first conditions are met but a clear dose-response for the luciferase induction cannot be observed, then the result of that repetition should be considered inconclusive and further testing may be required (Figure 1). In addition, a negative result obtained with concentrations < 1 000 μM (or < 200 $\mu\text{g/ml}$ for test chemicals with no defined MW) should also be considered as inconclusive (see paragraph 11).

Figure 1

Prediction model used in the KeratinoSens™ assay. A KeratinoSens™ prediction should be considered in the framework of an IATA and in accordance with the provision of paragraphs 9 and 11



In rare cases, test chemicals which induce the luciferase activity very close to the cytotoxic levels can be positive in some repetitions at non-cytotoxic levels (i.e. EC_{15} determining concentration below (<) the IC_{30}), and in other repetitions only at cytotoxic levels (i.e. EC_{15} determining concentration above (>) the IC_{30}). Such test chemicals shall be retested with more narrow dose-response analysis using a lower dilution factor (e.g. 1,33 or $\sqrt{2}$ (= 1,41) fold dilution between wells), to determine if induction has occurred at cytotoxic levels or not (9).

Test report

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, 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.
- 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.

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, 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 (vehicle) 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 / vehicles than those mentioned in this test method 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 method conditions

- Name and address of the sponsor, test facility and study director;
- Description of test method used;
- Cell line used, its storage conditions and source (e.g. the facility from which they were obtained);
- Passage number and level of confluence of cells used for testing;
- Cell counting method used for seeding prior to testing and measures taken to ensure homogeneous cell number distribution (cf. paragraph 20);
- 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;
- The procedure used to demonstrate proficiency of the laboratory in performing the test method (e.g. by testing of proficiency substances) or to demonstrate reproducible performance of the test method over time.

Test procedure

- Number of repetitions and replicates used;
- Test chemical concentrations, application procedure and exposure time used (if different than the one recommended)
- Description of evaluation and decision criteria used;
- Description of study acceptance criteria used;
- Description of any modifications of the test procedure.

Results

- Tabulation of I_{\max} , $EC_{1,5}$ and viability values (i.e. IC_{50} , IC_{30}) obtained for the test chemical and for the positive control for each repetition as well as the mean values (I_{\max} : average; $EC_{1,5}$ and viability values: geometric mean) and SD calculated using data from all individual repetitions and an indication of the rating of the test chemical according to the prediction model;
- Coefficient of variation obtained with the luminescence readings for the negative control for each experiment;
- 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 KeratinoSens™ assay;
- Consideration of the test method results within the context of an IATA, if other relevant information is available.

Conclusion

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) 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.

- (3) Adler S., Basketter D., Creton S., Pelkonen O., van Benthem J., Zuang V., Andersen K.E., Angers-Loustau A., Aptula A., Bal-Price A., Benfenati E., Bernauer U., Bessems J., Bois F.Y., Boobis A., Brandon E., Bremer S., Broschard T., Casati S., Coecke S., Corvi R., Cronin M., Daston G., Dekant W., Felter S., Grignard E., Gundert-Remy U., Heinonen T., Kimber I., Kleinjans J., Komulainen H., Kreiling R., Kreysa J., Leite S.B., Loizou G., Maxwell G., Mazzatorta P., Munn S., Pfuhler S., Phrakonkham P., Piersma A., Poth A., Prieto P., Repetto G., Rogiers V., Schoeters G., Schwarz M., Serafimova R., Tähti H., Testai E., van Delft J., van Loveren H., Vinken M., Worth A., Zaldivar J.M. (2011). Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. *Archives of Toxicology* 85, 367-485.
- (4) Chapter B.42 of this Annex: Skin sensitization: Local Lymph Node assay.
- (5) Chapter B.6 of this Annex: Skin Sensitisation.
- (6) Chapter B.50 of this Annex: Skin sensitization: Local Lymph Node assay: DA.
- (7) Chapter B.51 of this Annex: Skin sensitization: Local Lymph Node assay: BrdU-ELISA.
- (8) Natsch A. (2010). The Nrf2-Keap1-ARE Toxicity Pathway as a Cellular Sensor for Skin Sensitizers-Functional Relevance and Hypothesis on Innate Reactions to Skin Sensitizers. *Toxicological Sciences* 113, 284-292.
- (9) Emter R., Ellis G., Natsch A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers *in vitro*. *Toxicology and Applied Pharmacology* 245, 281-290.
- (10) Dinkova-Kostova A.T., Holtzclaw W.D., Kensler T.W. (2005). The role of Keap1 in cellular protective responses. *Chem. Res. Toxicol.* 18, 1779-1791.
- (11) Kansanen E., Kuosmanen S.M., Leinonen H., Levonen A.L. (2013). The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer. *Redox Biol.* 1(1), 45-49.
- (12) Natsch A., Bauch C., Foertsch L., Gerberick F., Normann K., Hilberer A., Inglis H., Landsiedel R., Onken S., Reuter H., Schepky A., Emter R. (2011). The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers *in vitro*: results of a ring-study in five laboratories. *Toxicol. In Vitro* 25, 733-744.
- (13) Natsch A., Ryan C.A., Foertsch L., Emter R., Jaworska J., Gerberick G.F., Kern P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *Journal of Applied Toxicology*, 33, 1337-1352.
- (14) EURL-ECVAM (2014). Recommendation on the KeratinoSens™ assay for skin sensitisation testing, 42 pp. Available at: http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations/recommendation-keratinosens-skin-sensitisation.
- (15) DB-ALM (INVITTOX) (2013) Protocol 155: KeratinoSens™, 17 pp. Available: <http://ecvam-dbal.m.jrc.ec.europa.eu/beta/index.cfm/methodsAndProtocols/index>
- (16) Natsch A., Emter R., Ellis G. (2009). Filling the concept with data: integrating data from different *in vitro* and *in silico* assays on skin sensitizers to explore the battery approach for animal-free skin sensitization testing. *Toxicol. Sci.* 107, 106-121.
- (17) Ball N., Cagen S., Carrillo J.C., Certa H., Eigler D., Emter R., Faulhammer F., Garcia C., Graham C., Haux C., Kolle S.N., Kreiling R., Natsch A., Mehling A. (2011). Evaluating the sensitization potential of surfactants: integrating data from the local lymph node assay, guinea pig maximization test, and *in vitro* methods in a weight-of-evidence approach. *Regul. Toxicol. Pharmacol.* 60, 389-400.
- (18) Bauch C., Kolle S.N., Ramirez T., Eltze 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.

- (19) Jaworska J., Dancik Y., Kern P., Gerberick F., Natsch A. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. *J Appl Toxicol.* 33, 1353-1364.
 - (20) Andres E., Sa-Rocha V.M., Barrichello C., Haupt T., Ellis G., Natsch A. (2013). The sensitivity of the KeratinoSens™ assay to evaluate plant extracts: A pilot study. *Toxicology In Vitro* 27, 1220-1225.
 - (21) 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, 1683-1699.
 - (22) 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, 646-657.
 - (23) OECD (2012). BG1Luc Estrogen Receptor Transactivation Test Method for Identifying Estrogen Receptor Agonists and Antagonists. OECD Guidelines for Chemical Testing No. 457. OECD, Paris.
 - (24) ECETOC (2003). Contact sensitization: Classification according to potency. European Centre for Ecotoxicology & Toxicology of Chemicals (Technical Report No. 87).
 - (25) Gildea L.A., Ryan C.A., Foertsch L.M., Kennedy J.M., Dearman R.J., Kimber I., Gerberick G.F. (2006). Identification of gene expression changes induced by chemical allergens in dendritic cells: opportunities for skin sensitization testing. *J. Invest. Dermatol.*, 126, 1813-1822.
 - (26) Ryan C.A., Gildea L.A., Hulette B.C., Dearman R.J., Kimber I., Gerberick G.F. (2004). Gene expressing changes in peripheral blood-derived dendritic cells following exposure to a contact allergen. *Toxicol. Lett.* 150, 301-316.
 - (27) Emter R., van der Veen J.W., Adamson G., Ezendam J., van Loveren H., Natsch A. (2013). Gene expression changes induced by skin sensitizers in the KeratinoSens™ cell line: Discriminating Nrf2-dependent and Nrf2-independent events. *Toxicol. in vitro* 27, 2225-2232.
 - (28) OECD (2015). Performance Standards for assessment of proposed similar or modified *in vitro* skin sensitisation ARE-Nrf2 luciferase test methods. OECD Environment, Health and Safety publications, Series on Testing and Assessment N0 213, OECD, Paris.
 - (29) OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Environment, Health and Safety publications, Series on Testing and Assessment No.34. OECD, Paris, France.
 - (30) NAFTA (North American Free Trade Agreement) (2012). Technical Working Group on Pesticides — (Quantitative) Structure Activity Relationship ((Q)SAR) Guidance Document. 186 pp. <http://www.epa.gov/oppfead1/international/naftatwg/guidance/qsar-guidance.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 (29).

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 (2).

ARE: Antioxidant response element (also called EpRE, electrophile response element), is a response element found in the upstream promoter region of many cytoprotective and phase II genes. When activated by Nrf2, it mediates the transcriptional induction of these genes.

Chemical: A substance or a mixture.

Coefficient of variation: A measure of variability that is calculated for a group of replicate data by dividing the standard deviation by the mean. It can be multiplied by 100 for expression as a percentage.

EC_{1,5}: Interpolated concentration for a 1,5 fold luciferase induction.

IC₃₀: Concentration effecting a reduction of cellular viability by 30 %.

IC₅₀: Concentration effecting a reduction of cellular viability by 50 %.

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.

I_{max}: Maximal induction factor of luciferase activity compared to the solvent (negative) control measured at any test chemical concentration.

Keap1: Kelch-like ECH-associated protein 1, is a sensor protein that can regulate the Nrf2 activity. Under un-induced conditions the Keap1 sensor protein targets the Nrf2 transcription factor for ubiquitinylation and proteolytic degradation in the proteasome. Covalent modification of the reactive cysteine residues of Keap 1 by small molecules can lead to dissociation of Nrf2 from Keap1 (8) (10) (11).

Mixture: A mixture or a solution composed of two or more substances in which they do not react (1).

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.

Nrf2: Nuclear factor (erythroid-derived 2)-like 2, is a transcription factor involved in the antioxidant response pathway. When Nrf2 is not ubiquitinated, it builds up in the cytoplasm and translocates into the nucleus, where it combines to the ARE in the upstream promoter region of many cytoprotective genes, initiating their transcription (8) (10) (11).

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 (29).

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 (29).

Reproducibility: The agreement among results obtained from testing the same chemical using the same test protocol (see reliability) (29).

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 (29).

Solvent/vehicle control: A replicate containing all components of a test system except of the test chemical, but including the solvent that is used. It is used to establish the baseline response for the samples treated with the test chemical dissolved in the same solvent.

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 (29).

Substance: Chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the 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 (1).

Test chemical: The term “test chemical” is used to refer to what is being tested.

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).

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 (29).

Appendix 2

PROFICIENCY SUBSTANCES

In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method

Prior to routine use of this test method, laboratories should demonstrate technical proficiency by correctly obtaining the expected KeratinoSens™ prediction for the 10 Proficiency Substances recommended in Table 1 and by obtaining the EC_{1,5} and IC₅₀ values that fall within the respective reference range for at least 8 out of the 10 proficiency substances. These Proficiency Substances were selected to represent the range of responses for skin sensitisation hazards. Other selection criteria were commercial availability, availability of high quality *in vivo* reference, and availability of high quality *in vitro* data from the KeratinoSens™ assay.

Table 1

Recommended substances for demonstrating technical proficiency with the KeratinoSens™ assay

Proficiency Substances	CASRN	Physical Form	<i>In Vivo</i> Prediction ⁽¹⁾	KeratinoSens™ Prediction ⁽²⁾	EC _{1,5} (µM) Reference Range ⁽³⁾	IC ₅₀ (µM) Reference Range ⁽³⁾
Isopropanol	67-63-0	Liquid	Non-sensitiser	Negative	> 1 000	> 1 000
Salicylic acid	69-72-7	Solid	Non-sensitiser	Negative	> 1 000	> 1 000
Lactic acid	50-21-5	Liquid	Non-sensitiser	Negative	> 1 000	> 1 000
Glycerol	56-81-5	Liquid	Non-sensitiser	Negative	> 1 000	> 1 000
Cinnamyl alcohol	104-54-1	Solid	Sensitiser (weak)	Positive	25 - 175	> 1 000
Ethylene glycol dimethacrylate	97-90-5	Liquid	Sensitiser (weak)	Positive	5 - 125	> 500
2-Mercaptobenzothiazole	149-30-4	Solid	Sensitiser (moderate)	Positive	25 - 250	> 500
Methyldibromoglutaronitrile	35691-65-7	Solid	Sensitiser (strong)	Positive	< 20	20 - 100
4-Methylamino-phenol sulfate	55-55-0	Solid	Sensitiser (strong)	Positive	< 12,5	20 - 200
2,4-Dinitro-chlorobenzene	97-00-7	Solid	Sensitiser (extreme)	Positive	< 12,5	5 - 20

⁽¹⁾ The *in vivo* hazard (and potency) predictions are based on LLNA data (13). The *in vivo* potency is derived using the criteria proposed by ECETOC (24).

⁽²⁾ A KeratinoSens™ prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 9 and 11 of this test method.

⁽³⁾ Based on the historical observed values (12).

Appendix 3

QUALITY CONTROL OF LUMINESCENCE MEASUREMENTS

Basic experiment for ensuring optimal luminescence measurements in the KeratinoSens™ assay

The following three parameters are critical to ensure obtaining reliable results with the luminometer:

- having a sufficient sensitivity giving a stable background in control wells;
- having no gradient over the plate due to long reading times; and
- having no light contamination in adjacent wells from strongly active wells.

Prior to testing it is recommended to ensure having appropriate luminescence measurements, by testing a control plate set-up as described below (triplicate analysis).

Plate setup of first training experiment

	1	2	3	4	5	6	7	8	9	10	11	12
A	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
B	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
C	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
D	EGDMA 0,98	EGDMA 1,95	EGDMA 3,9	EGDMA 7,8	EGDMA 15,6	EGDMA 31,25	EGDMA 62,5	EGDMA 125	EGDMA 250	EGDMA 500	EGDMA 1000	EGDMA 2000
E	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
F	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
G	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
H	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	CA 4	CA 8	CA 16	CA 32	CA 64	Blank

EGDMA = Ethylene glycol dimethacrylate (CAS No.: 97-90-5) a strongly inducing chemical

CA = Cinnamic aldehyde, positive reference (CAS No.: 104-55-2)

The quality control analysis should demonstrate:

- a clear dose-response in row D, with the $I_{\max} > 20$ fold above background (in most cases I_{\max} values between 100 and 300 are reached);

- no dose-response in row C and E (no induction value above 1,5 (ideally not above 1,3) due to possible light contamination especially next to strongly active wells in the EGDMA row;
 - no statistically significant difference between the rows A, B, C, E, F and G. (i.e. no gradient over plate); and
 - variability in any of the rows A, B, C, E, F and G and in the DMSO wells in row H should be below 20 % (i.e. stable background).
-

B.61 Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants**INTRODUCTION**

This test method (TM) is equivalent to OECD test guideline (TG) 460 (2012). The Fluorescein Leakage (FL) test method is an *in vitro* test method that can be used under certain circumstances and with specific limitations to classify chemicals (substances and mixtures) as ocular corrosives and severe irritants, as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (Category 1), Regulation (EC) No 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (CLP) ⁽¹⁾ (Category 1), and the U.S. Environmental Protection Agency (EPA) (Category I) (1)(2). For the purpose of this test method, severe ocular irritants are defined as chemicals that cause tissue damage in the eye following test chemical administration that is not reversible within 21 days or causes serious physical decay of vision, while ocular corrosives are chemicals that cause irreversible tissue damage to the eye. These chemicals are classified as UN GHS Category 1, EU CLP Category 1, or U.S. EPA Category I.

While the FL test method is not considered valid as a complete replacement for the *in vivo* rabbit eye test, the FL is recommended for use as part of a tiered testing strategy for regulatory classification and labelling. Thus, the FL is recommended as an initial step within a Top-Down approach to identify ocular corrosives/severe irritants, specifically for limited types of chemicals (*i.e.* water soluble substances and mixtures) (3)(4).

It is currently generally accepted that, in the foreseeable future, no single *in vitro* eye irritation test will be able to replace the *in vivo* eye test (TM B.5 (5)) to predict across the full range of irritation for different chemical classes. However, strategic combinations of several alternative test methods within a (tiered) testing strategy may be able to replace the *in vivo* eye test (4). The Top-Down approach (4) is designed to be used when, based on existing information, a chemical is expected to have high irritancy potential.

Based on the prediction model detailed in paragraph 35, the FL test method can identify chemicals within a limited applicability domain as ocular corrosives/severe irritants (UN GHS Category 1; EU CLP Category 1; U.S. EPA Category I) without any further testing. The same is assumed for mixtures although mixtures were not used in the validation. Therefore, the FL test method may be used to determine the eye irritancy/corrosivity of chemicals, following the sequential testing strategy of TM B.5 (5). However, a chemical that is not predicted as ocular corrosive or severe irritant with the FL test method would need to be tested in one or more additional test methods (*in vitro* and/or *in vivo*) that are capable of accurately identifying i) chemicals that are *in vitro* false negative ocular corrosives/severe irritants in the FL (UN GHS Category 1; EU CLP Category 1; U.S. EPA Category I); ii) chemicals that are not classified for eye corrosion/irritation (UN GHS No Category; EU CLP No Category; U.S. EPA Category IV); and/or iii) chemicals that are moderate/mild eye irritants (UN GHS Categories 2A and 2B; EU CLP Category 2; U.S. EPA Categories II and III).

The purpose of this test method is to describe the procedures used to evaluate the potential ocular corrosivity or severe irritancy of a test chemical as measured by its ability to induce damage to an impermeable confluent epithelial monolayer. The integrity of trans-epithelial permeability is a major function of an epithelium such as that found in the conjunctiva and the cornea. Trans-epithelial permeability is controlled by various tight junctions. Increasing the permeability of the corneal epithelium *in vivo* has been shown to correlate with the level of inflammation and surface damage observed as eye irritation develops.

In the FL test method, toxic effects after a short exposure time to the test chemical are measured by an increase in permeability of sodium fluorescein through the epithelial monolayer of Madin-Darby Canine Kidney (MDCK) cells cultured on permeable inserts. The amount of fluorescein leakage that occurs is proportional to the chemical-induced damage to the tight junctions, desmosomal junctions and cell membranes, and can be used to estimate the ocular toxicity potential of a test chemical. Appendix 1 provides a diagram of MDCK cells grown on an insert membrane for the FL test method.

⁽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, 31.12.2008, p. 1.

Definitions are provided in Appendix 2.

INITIAL CONSIDERATIONS AND LIMITATIONS

This test method is based on the INVITTOX protocol No. 71 (6) that has been evaluated in an international validation study by the European Centre for the Validation of Alternative Methods (ECVAM), in collaboration with the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the Japanese Center for the Validation of Alternative Methods (JaCVAM).

The FL test method is not recommended for the identification of chemicals which should be classified as mild/moderate irritants or of chemicals which should not be classified for ocular irritation (substances and mixtures) (i.e. GHS Cat. 2A/2B, no category; EU CLP Cat. 2, no category; US EPA Cat. II/III/IV), as demonstrated by the validation study (3) (7).

The test method is only applicable to water soluble chemicals (substances and mixtures). The ocular severe irritation potential of chemicals that are water soluble and/or where the toxic effect is not affected by dilution is generally predicted accurately using the FL test method (7). To categorise a chemical as water soluble, under experimental conditions, it should be soluble in sterile calcium-containing (at a concentration of 1,0-1,8 mM), phenol red-free, Hanks' Balanced Salt Solution (HBSS) at a concentration ≥ 250 mg/ml (one dose above the cut-off of 100 mg/ml). However, if the test chemical is soluble below the concentration 100 mg/ml, but already induces a FL induction of 20 % at that concentration (meaning $FL_{20} < 100$ mg/ml), it can still be classified as GHS Cat. 1 or EPA Cat. I.

The identified limitations for this test method exclude strong acids and bases, cell fixatives and highly volatile chemicals from the applicability domain. These chemicals have mechanisms that are not measured by the FL test method, e.g. extensive coagulation, saponification or specific reactive chemistries. Other identified limitations for this method are based upon the results for the predictive capacity for coloured and viscous test chemical (7). It is suggested that both types of chemicals are difficult to remove from the monolayer following the short exposure period and that predictivity of the test method could be improved if a higher number of washing steps was used. Solid chemicals suspended in liquid have the propensity to precipitate out and the final concentration to cells can be difficult to determine. When chemicals within these chemical and physical classes are excluded from the database, the accuracy of FL across the EU, EPA, and GHS classification systems is substantially improved (7).

Based on the purpose of this test method (i.e. to identify ocular corrosives/severe irritants only), false negative rates (see Paragraph 13) are not critical since such chemicals would be subsequently tested with other adequately validated *in vitro* tests or in rabbits, depending on regulatory requirements, using a sequential testing strategy in a weight of evidence approach (5) (see also paragraphs 3 and 4).

Other identified limitations of the FL test method are based on false negative and false positive rates. When used as an initial step within a Top-Down approach to identify water soluble ocular corrosive/severe irritant substances and mixtures (UN GHS Category 1; EU CLP Category 1; U.S. EPA Category I), the false positive rate for the FL test method ranged from 7 % (7/103; UN GHS and EU CLP) to 9 % (9/99; U.S. EPA) and the false negative rate ranged from 54 % (15/28; U.S. EPA) to 56 % (27/48; UN GHS and EU CLP) when compared to *in vivo* results. Chemical groups showing false positive and/or false negative results in the FL test method are not defined here.

Certain technical limitations are specific to the MDCK cell culture. The tight junctions that block the passage of the sodium-fluorescein dye through the monolayer are increasingly compromised with increasing cell passage number. Incomplete formation of the tight junctions results in increased FL in the non-treated control. Therefore, a defined permissible maximal leakage in the non-treated controls is important (see paragraph 38: 0 % leakage). As with all *in vitro* assays there is the potential for the cells to become transformed over time, thus it is vital that passage number ranges for the assays are stated.

The current applicability domain might be increased in some cases, but only after analysing an expanded data set of studied test chemicals, preferably acquired through testing (3). This test method will be updated accordingly as new information and data are considered.

For any laboratory initially establishing this assay, the proficiency chemicals provided in Appendix 3 should be used. Laboratories can use these chemicals to demonstrate their technical competence in performing the FL test method prior to submitting FL assay data for regulatory hazard classification purposes.

PRINCIPLE OF THE TEST

The FL test method is a cytotoxicity and cell-function based *in vitro* assay that is performed on a confluent monolayer of MDCK CB997 tubular epithelial cells that are grown on semi-permeable inserts and model the non-proliferating state of the *in vivo* corneal epithelium. The MDCK cell line is well established and forms tight junctions and desmosomal junctions similar to those found on the apical side of conjunctival and corneal epithelia. Tight and desmosomal junctions *in vivo* prevent solutes and foreign materials penetrating the corneal epithelium. Loss of trans-epithelial impermeability, due to damaged tight junctions and desmosomal junctions, is one of the early events in chemical-induced ocular irritation.

The test chemical is applied to the confluent layer of cells grown on the apical side of the insert. A short 1 min exposure is routinely used to reflect the normal clearance rate in human exposures. An advantage of the short exposure period is that water-based substances and mixtures can be tested neat, if they can be easily removed after the exposure period. This allows more direct comparisons of the results with the chemical effects in humans. The test chemical is then removed and the non-toxic, highly fluorescent sodium-fluorescein dye is added to the apical side of the monolayer for 30 minutes. The damage caused by the test chemical to the tight junctions is determined by the amount of fluorescein which leaks through the cell layer within a defined period of time.

The amount of sodium-fluorescein dye that passes through the monolayer and the insert membrane into a set volume of solution present in the well (to which the sodium-fluorescein dye leaks in) is determined by measuring spectrofluorometrically the fluorescein concentration in the well. The amount of fluorescein leakage (FL) is calculated with reference to fluorescence intensity (FI) readings from two controls: a blank control, and a maximum leakage control. The percentage of leakage and therefore amount of damage to the tight junctions is expressed, relative to these controls, for each of the set concentrations of the test chemical. Then the FL₂₀ (i.e. concentration that causes 20 % FL relative to the value recorded for the untreated confluent monolayer and inserts without cells), is calculated. The FL₂₀ (mg/ml) value is used in the prediction model for identification of ocular corrosives and severe irritants (see paragraph 35).

Recovery is an important part of a test chemical's toxicity profile that is also assessed by the *in vivo* ocular irritation test. Preliminary analyses indicated that recovery data (up to 72 h following the chemical exposure) could potentially increase the predictive capacity of INVITTOX Protocol 71 but further evaluation is needed and would benefit from additional data, preferably acquired by further testing (6). This test method will be updated accordingly as new information and data are considered.

PROCEDURE

Preparation of the cellular monolayer

The monolayer of MDCK CB997 cells is prepared using sub-confluent cells growing in cell culture flasks in DMEM/ Nutrient Mix F12 (1x concentrate with L-glutamine, 15 mM HEPES, calcium (at a concentration of 1,0-1,8 mM) and 10 % heat-inactivated FCS/FBS). Importantly, all media/solutions used throughout the FL assay should contain calcium at a concentration between 1,8 mM (200 mg/l) and 1,0 mM (111 mg/l) to ensure tight junction formation and integrity. Cell passage number range should be controlled to ensure even and reproducible tight junctions formation. Preferably, the cells should be within the passage range 3-30 from thawing because cells within this passage range have similar functionality, which aids assay results to be reproducible.

Prior to performing the FL test method, the cells are detached from the flask by trypsinisation, centrifuged and an appropriate amount of cells is seeded into the inserts placed in 24-well plates (see Appendix 1). Twelve mm diameter inserts with membrane of mixed cellulose esters, a thickness of 80-150 µm and a pore size of 0,45 µm, should be used to seed the cells. In the validation study, Millicell-HA 12 mm inserts were used. The properties of the insert and membrane type are important as these may affect cell growth and chemical binding. Certain types of chemicals may bind to the Millicell-HA insert membrane, which could affect the interpretation of results. Proficiency chemicals (see Appendix 3) should be used to demonstrate equivalency if other membranes are used.

Chemical binding to the insert membrane is more common for cationic chemicals, such as benzalkonium chloride, which are attracted to the charged membrane (7). Chemical binding to the insert membrane may increase the chemical exposure period, leading to an over-estimation of the toxic potential of the chemical, but can also physically reduce the leakage of fluorescein through the insert by binding of the dye to the cationic chemical bound to the insert membrane, leading to an under-estimation of the toxic potential of the chemical. This can be readily monitored by exposing the membrane alone to the top concentration of the chemical tested and then adding sodium-fluorescein dye at the normal concentration for the standard time (no cell control). If binding of the sodium-fluorescein dye occurs, the insert membrane appears yellow after the test material has been washed-off. Thus, it is essential to know the binding properties of the test chemical in order to be able to interpret the effect of the chemical on the cells.

Cell seeding on inserts should produce a confluent monolayer at the time of chemical exposure. $1,6 \times 10^5$ cells should be added per insert (400 µl of a cell suspension with a density of 4×10^5 cells / ml). Under these conditions, a confluent monolayer is usually obtained after 96 hours in culture. Inserts should be examined visually prior to seeding, so as to ensure that any damages recorded at the visual control described at paragraph 30 is due to handling.

The MDCK cell cultures should be kept in incubators in a humidified atmosphere, at $5 \% \pm 1 \% \text{ CO}_2$ and $37 \pm 1 ^\circ\text{C}$. The cells should be free of contamination by bacteria, viruses, mycoplasma and fungi.

Application of the Test and Control Chemicals

A fresh stock solution of test chemical should be prepared for each experimental run and used within 30 minutes of preparation. Test chemicals should be prepared in calcium-containing (at a concentration of 1,0-1,8 mM), phenol red-free, HBSS to avoid serum protein binding. Solubility of the chemical at 250 mg/ml in HBSS should be assessed prior to testing. If at this concentration the chemical forms a stable suspension or emulsion (*i.e.* maintains uniformity and does not settle or separate into more than one phase) over 30 minutes, HBSS can still be used as solvent. However, if the chemical is found to be insoluble in HBSS at this concentration, the use of other test methods instead of FL should be considered. The use of light mineral oil as a solvent, in cases where the chemical is found to be insoluble in HBSS, should be considered with caution as there is not enough data available to conclude on the performance of the FL assay under such conditions.

All chemicals to be tested are prepared in sterile calcium-containing (at a concentration of 1,0-1,8 mM), phenol red-free, HBSS from the stock solution, at five fixed concentrations diluted on a weight per volume basis: 1, 25, 100, 250 mg/ml and a neat or a saturated solution. When testing a solid chemical, a very high concentration of 750 mg/ml should be included. This concentration of chemical may have to be applied on the cells using a positive displacement pipette. If the toxicity is found to be between 25 and 100 mg/ml, the following additional concentrations should be tested twice: 1, 25, 50, 75, 100 mg/ml. The FL_{20} value should be derived from these concentrations provided the acceptance criteria were met.

The test chemicals are applied to the confluent cell monolayers after removal of the cell culture medium and washing twice with sterile, warm (37 °C), calcium-containing (at a concentration of 1,0-1,8 mM), phenol red-free, HBSS. Previously, the filters have been visually checked for any pre-existing damages that could be falsely attributed to potential incompatibilities with test chemicals. At least three replicates should be used for each concentration of the test chemical and for the controls in each run. After 1 min of exposure at room temperature, the test chemical should be carefully removed by aspiration, the monolayer should be washed twice with sterile, warm (37 °C), calcium-containing (at a concentration of 1,0-1,8 mM), phenol red-free, HBSS, and the fluorescein leakage should be immediately measured.

Concurrent negative (NC) and positive controls (PC) should be used in each run to demonstrate that monolayer integrity (NC) and sensitivity of the cells (PC) are within a defined historical acceptance range. The suggested PC chemical is Brij 35 (CAS No. 9002-92-0) at 100 mg/ml. This concentration should give approximately 30 % fluorescein leakage (acceptable range 20-40 % fluorescein leakage, i.e. damage to cell layer). The suggested NC chemical is calcium-containing (at a concentration of 1,0-1,8 mM), phenol red-free, HBSS (untreated, blank control). A maximum leakage control should also be included in each run to allow for the calculation of FL₂₀ values. Maximum leakage is determined using a control insert without cells.

Determination of fluorescein permeability

Immediately after removal of the test and control chemicals, 400 µl of 0,1 mg/ml sodium-fluorescein solution (0,01 % (w/v) in calcium-containing [at a concentration of 1,0-1,8 mM], phenol red-free, HBSS) is added to the inserts (e.g. Millicell-HA). The cultures are kept for 30 minutes at room temperature. At the end of the incubation with fluorescein, the inserts are carefully removed from each well. Visual check is performed on each filter and any damage which may have occurred during handling is recorded.

The amount of fluorescein that leaked through the monolayer and the insert is quantified in the solution which remained in the wells after removal of the inserts. Measurements are done in a spectrofluorometer at excitation and emission wavelengths of 485 nm and 530 nm, respectively. The sensitivity of the spectrofluorometer should be set so that there is the highest numerical difference between the maximum FL (insert with no cells) and the minimum FL (insert with confluent monolayer treated with NC). Because of the differences in the used spectrofluorometer, it is suggested that a sensitivity is used which will give fluorescence intensity > 4 000 at the maximum fluorescein leakage control. The maximum FL value should not be greater than 9 999. The maximum fluorescence leakage intensity should fall within the linear range of the spectrofluorometer used.

Interpretation of results and Prediction model

The amount of FL is proportional to the chemical-induced damage to the tight junctions. The percentage of FL for each tested concentration of chemical is calculated from the FL values obtained for the test chemical with reference to FL values from the NC (reading from the confluent monolayer of cells treated with the NC) and a maximum leakage control (reading for the amount of FL through an insert without cells).

The mean maximum leakage fluorescence intensity = x

The mean 0 % leakage fluorescence intensity (NC) = y

The mean 100 % leakage is obtained by subtracting the mean 0 % leakage from the mean maximum leakage,

i.e. $x - y = z$

The percentage leakage for each fixed dose is obtained by subtracting the 0 % leakage to the mean fluorescence intensity of the three replicate readings (m), and dividing this value by the 100 % leakage, *i.e.* %FL = [(m-y) / z] × 100 %, where:

m = the mean fluorescence intensity of the three replicate measurements for the concentration involved

% FL = the percent of the fluorescein which leaks through the cell layer

The following equation for the calculation of the chemical concentration causing 20 % FL should be applied:

$$FL_D = [(A-B) / (C-B)] \times (M_C - M_B) + M_B$$

Where:

D = % of inhibition

A = % damage (20 % fluorescein leakage)

B = % fluorescein leakage < A

C = % fluorescein leakage > A

M_C = Concentration (mg/ml) of C

M_B = Concentration (mg/ml) of B

The cut-off value of FL₂₀ for predicting chemicals as ocular corrosives/severe irritants is given below:

FL ₂₀ (mg/ml)	UN GHS C&L	EU CLP C&L	U.S. EPA C&L
≤ 100	Category 1	Category 1	Category I

C&L: classification and labelling.

The FL test method is recommended only for the identification of water soluble ocular corrosives and severe irritants (UN GHS Category 1, EU CLP Category 1, U.S. EPA Category I) (see paragraphs 1 and 10).

In order to identify water soluble chemicals (substances and mixtures) (3) (6) (7) as “inducing serious eye damage” (UN GHS/EU CLP Category 1) or as an “ocular corrosive or severe irritant” (U.S. EPA Category I), the test chemical should induce an FL₂₀ value of ≤ 100 mg/ml.

Acceptance of results

The mean maximum fluorescein leakage value (x) should be higher than 4 000 (see paragraph 31), the mean 0 % leakage (y) should be equal or lower than 300, and the mean 100 % leakage (z) should fall between 3 700 and 6 000.

A test is considered acceptable if the positive control produced 20 % to 40 % damage to the cell layer (measure as % fluorescein leakage).

DATA AND REPORTING

Data

For each run, data from individual replicate wells (e.g. fluorescence intensity values and calculated percentage FL data for each test chemical, including classification) should be reported in tabular form. In addition, means ± SD of individual replicate measurements in each run should be reported.

Test Report

The test report should include the following information:

Test and Control Chemicals

- Chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known;
- Chemical CAS number, if known;
- Purity and composition of the substance or mixture (in percentage(s) by weight), to the extent this information is available;
- Physical-chemical properties relevant to the conduct of the study (e.g. physical state, volatility, pH, stability, water solubility, chemical class);
- Treatment of the test/control chemical prior to testing, if applicable (e.g. warming, grinding);
- Storage conditions;

Justification of the test method and Protocol Used

- Should include considerations regarding applicability domain and limitations of the test method;

Test Conditions

- Description of cell system used, including certificate of authenticity and the mycoplasma status of the cell line;
- Details of test procedure used;
- Test chemical concentration(s) used;
- Duration of exposure to the test chemical;
- Duration of incubation with fluorescein;
- Description of any modifications of the test procedure;
- Description of evaluation criteria used;
- Reference to historical data of the model (e.g. negative and positive controls, benchmark chemicals, if applicable);
- Information on the technical proficiency demonstrated by the laboratory;

Results

- Tabulation of data from individual test chemicals and controls for each run and each replicate measurement (including individual results, means and SDs);
- The derived classification(s) with reference to the prediction model and/or decision criteria used;
- Description of other effects observed;

Discussion of the Results

- Should include considerations regarding a non-conclusive outcome (paragraph 35: FL₂₀ > 100 mg/ml) and further testing;

Conclusions

LITERATURE:

- (1) UN (2009), United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Third 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_rev03/03files_e.html]
 - (2) U.S. EPA (1996), Label Review Manual: 2nd Edition, EPA737-B-96-001, Washington DC: U.S. Environmental Protection Agency.
 - (3) EC-ECVAM (2009), Statement on the scientific validity of cytotoxicity/cell-function based *in vitro* assays for eye irritation testing.
 - (4) 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
 - (5) Chapter B.5 of this Annex, *Acute Eye Irritation/Corrosion*
 - (6) EC-ECVAM (1999), INVITOX Protocol 71: Fluorescein Leakage Test, Ispra, Italy: European Centre for the Validation of Alternative Methods (ECVAM). Available at: [<http://ecvam-dbalm.jrc.ec.europa.eu>]
 - (7) EC-ECVAM (2008), Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing.
 - (8) OECD (2005), *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment*, OECD Series on Testing and Assessment No. 34. OECD, Paris.
-

Appendix 1

DIAGRAM OF MDCK CELLS GROWN ON AN INSERT MEMBRANE FOR THE FL TEST METHOD

A confluent layer of MDCK cells is grown on the semi-permeable membrane of an insert. The inserts are placed into the wells of 24 well plates.

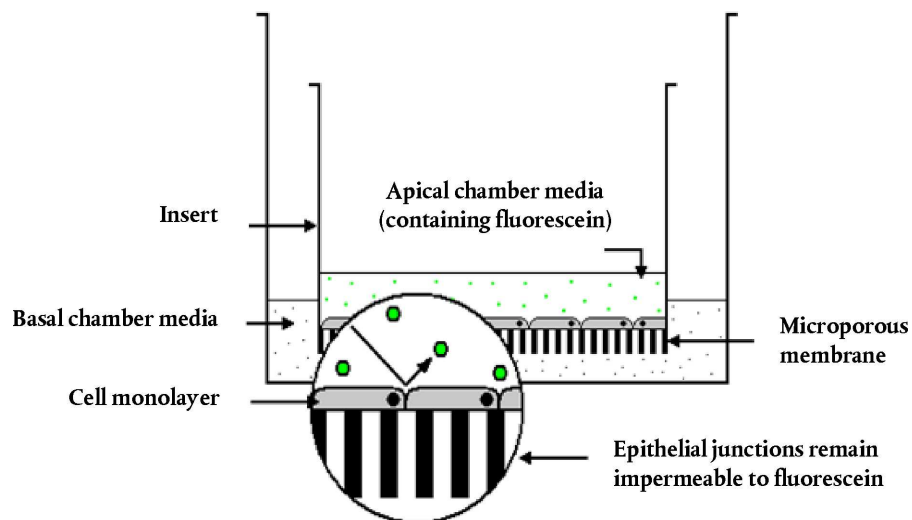


Figure taken from: Wilkinson, P.J. (2006), Development of an *in vitro* model to investigate repeat ocular exposure, Ph.D. Thesis, University of Nottingham, UK.

Appendix 2

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.

Chemical: A substance or a mixture.

EPA Category I: Chemicals that produce corrosive (irreversible destruction of ocular tissue) or corneal involvement or irritation persisting for more than 21 days (2).

EU CLP (Regulation (EC) No1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures): Implements in the European Union (EU) the UN GHS system for the classification of chemicals (substances and mixtures).

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.

FL₂₀: Can be estimated by the determination of the concentration at which the tested chemical causes 20 % of the fluorescein leakage through the cell layer.

Fluorescein leakage: the amount of fluorescein which passes through the cell layer, measured spectrofluorometrically.

GHS (Globally Harmonized System of Classification and Labeling of Chemicals by the United Nation (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.

GHS Category 1: 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.

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.

Mixture: Used in the context of the UN GHS as a mixture or solution composed of two or more substances in which they do not react.

Negative 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.

Not-classified: Chemicals that are not classified as UN GHS Categories 1, 2A, or 2B; EU CLP Categories 1 or 2; or U.S. EPA Categories I, II, or III ocular irritants.

Ocular corrosive: (a) A chemical that causes irreversible tissue damage to the eye. (b) Chemicals that are classified as UN GHS Category 1; EU CLP Category 1; or U.S. EPA Category I ocular irritants.

Ocular irritant: (a) A chemical that produces a reversible change in the eye following application to the anterior surface of the eye; (b) Chemicals that are classified as UN GHS Categories 2A, or 2B; EU CLP Category 2; or U.S. EPA Categories II or III ocular irritants.

Ocular severe irritant: (a) A chemical that causes tissue damage in the eye following application to the anterior surface of the eye that is not reversible within 21 days of application or causes serious physical decay of vision. (b) Chemicals that are classified as UN GHS Category 1; EU CLP Category 1; or U.S. EPA Category I ocular irritants.

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 extreme.

Proficiency Chemicals: A sub-set of the list of Reference Chemicals that can be used by a naïve laboratory to demonstrate proficiency with the validated reference test method.

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 (8).

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.

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.

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 (8).

Serious eye damage: Is 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.

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 negative 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.

Substance: Used in the context of the UN GHS as chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the 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.

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.

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 (8).

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 chemical.

Appendix 3

PROFICIENCY CHEMICALS FOR THE FL TEST METHOD

Prior to routine use of this test method, laboratories should demonstrate technical proficiency by correctly identifying the ocular corrosivity classification of the 8 chemicals recommended in Table 1. These chemicals were selected to represent the range of responses for local eye irritation/corrosion, which is based on results in the *in vivo* rabbit eye test (TG 405, TM B.5(5)) (i.e., Categories 1, 2A, 2B, or no classification according to the UN GHS). However, considering the validated usefulness of the FL assay (i.e., to identify ocular corrosives/severe irritants only), there are only two test outcomes for classification purposes (corrosive/severe irritant or non-corrosive/non-severe irritant) to demonstrate proficiency. Other selection criteria were that chemicals are commercially available, there are high quality *in vivo* reference data available, and there are high quality data from the FL test method. For this reason, the proficiency chemicals were selected from the “Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing” (8), which was used for the retrospective validation of the FL test method.

Table 1

Recommended chemicals for demonstrating technical proficiency with FL

Chemical	CAS NR	Chemical Class ⁽¹⁾	Physical Form	In Vivo Classification ⁽²⁾	In Vitro Classification ⁽³⁾
Benzalkonium chloride (5 %)	8001-54-5	Onium compound	Liquid	Category 1	Corrosive/Severe Irritant
Promethazine hydrochloride	58-33-3	Amine/Amidine, Heterocyclic, Organic sulphur compound	Solid	Category 1	Corrosive/Severe Irritant
Sodium hydroxide (10 %)	1310-73-2	Alkali	Liquid	Category 1	Corrosive/Severe Irritant
Sodium lauryl sulfate (15 %)	151-21-3	Carboxylic acid (salt)	Liquid	Category 1	Corrosive/Severe Irritant
4-carboxy-benzaldehyde	619-66-9	Carboxylic acid, Aldehyde	Solid	Category 2(A)	Non-corrosive/Non-severe irritant
Ammonium nitrate	6484-52-2	Inorganic salt	Solid	Category 2(A)	Noncorrosive/Non-severe irritant
Ethyl-2-methylacetoacetate	609-14-3	Ketone, Ester	Liquid	Category 2(B)	Noncorrosive/Non-severe irritant
Glycerol	56-81-5	Alcohol	Liquid	No Category	Noncorrosive/Non-severe irritant

Abbreviations: CAS NR = Chemical Abstracts Service Registry Number

⁽¹⁾ Chemical classes were assigned to each test chemical using a standard classification scheme, based on the National Library of Medicine Medical Subject Headings (MeSH) classification system (available at <http://www.nlm.nih.gov/mesh>)

⁽²⁾ Based on results from the *in vivo* rabbit eye test (OECD TG 405, TM B.5) and using the UN GHS and EU CLP.

⁽³⁾ Based on results obtained with FL (INVITTOX Protocol No. 71(6))

B.62 *In Vivo* Mammalian Alkaline Comet Assay

INTRODUCTION

This test method (TM) is equivalent to OECD test guideline (TG) 489 (2016). The *in vivo* alkaline comet (single cell gel electrophoresis) assay (hereafter called simply the comet assay) is used for the detection of DNA strand breaks in cells or nuclei isolated from multiple tissues of animals, usually rodents, that have been exposed to potentially genotoxic material(s). The comet assay has been reviewed and recommendations have been published by various expert groups (1) (2) (3) (4) (5) (6) (7) (8) (9) (10). This test method is part of a series of test methods on genetic toxicology. An OECD document that provides succinct information on genetic toxicology testing and an overview of the recent changes that were made to these Test Guidelines has been developed (11).

The purpose of the comet assay is to identify chemicals that cause DNA damage. Under alkaline conditions (> pH 13), the comet assay can detect single and double stranded breaks, resulting, for example, from direct interactions with DNA, alkali labile sites or as a consequence of transient DNA strand breaks resulting from DNA excision repair. These strand breaks may be repaired, resulting in no persistent effect, may be lethal to the cell, or may be fixed into a mutation resulting in a permanent viable change. They may also lead to chromosomal damage which is also associated with many human diseases including cancer.

A formal validation trial of the *in vivo* rodent comet assay was performed in 2006-2012, coordinated by the Japanese Center for the Validation of Alternative Methods (JaCVAM), in conjunction with the European Centre for the Validation of Alternative Methods (ECVAM), the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) (12). This test method includes the recommended use and limitations of the comet assay, and is based on the final protocol (12) used in the validation trial, and on additional relevant published and unpublished (laboratories proprietary) data.

Definitions of key terms are set out in Appendix 1. It is noted that many different platforms can be used for this assay (microscope slides, gel spots, 96-well plates etc.). For convenience the term “slide” is used throughout the remainder of this document but encompasses all of the other platforms.

INITIAL CONSIDERATIONS AND LIMITATIONS

The comet assay is a method for measuring DNA strand breaks in eukaryotic cells. Single cells/nuclei embedded in agarose on a slide are lysed with detergent and high salt concentration. This lysis step digests the cellular and nuclear membranes and allows the release of coiled DNA loops generally called nucleoids and DNA fragments. Electrophoresis at high pH results in structures resembling comets, which, by using appropriate fluorescent stains, can be observed by fluorescence microscopy; DNA fragments migrate away from the “head” into the “tail” based on their size, and the intensity of the comet tail relative to the total intensity (head plus tail) reflects the amount of DNA breakage (13) (14) (15).

The *in vivo* alkaline comet assay is especially relevant to assess genotoxic hazard in that the assay's responses are dependent upon *in vivo* ADME (absorption, distribution, metabolism and excretion), and also on DNA repair processes. These may vary among species, among tissues and among the types of DNA damage.

To fulfil animal welfare requirements, in particular the reduction in animal usage (3Rs — Replacement, Reduction, Refinement— principles), this assay can also be integrated with other toxicological studies, e.g. repeated dose toxicity studies (10) (16) (17), or the endpoint can be combined with other genotoxicity endpoints such as the *in vivo* mammalian erythrocyte micronucleus assay (18) (19) (20). The comet assay is most often performed in rodents, although it has been applied to other mammalian and non-mammalian species. The use of non-rodent species should be scientifically and ethically justified on a case-by-case basis and it is strongly recommended that the comet assay only be performed on species other than rodents as part of another toxicity study and not as a standalone test.

The selection of route of exposure and tissue(s) to be studied should be determined based on all available/existing knowledge of the test chemicals e.g. intended/expected route of human exposure, metabolism and distribution, potential for site-of-contact effects, structural alerts, other genotoxicity or toxicity data, and the purpose of the study. Thus, where appropriate, the genotoxic potential of the test chemicals can be assayed in the target tissue(s) of carcinogenic and/or other toxic effects. The assay is also considered useful for further investigation of genotoxicity detected by an *in vitro* system. It is appropriate to perform an *in vivo* comet assay in a tissue of interest when it can be reasonably expected that the tissue of interest will be adequately exposed.

The assay has been most extensively validated in somatic tissues of male rats in collaborative studies such as the JaCVAM trial (12) and in Rothfuss *et al.*, 2010 (10). The liver and stomach were used in the JaCVAM international validation trial. The liver, because it is the most active organ in metabolism of chemicals and also frequently a target organ for carcinogenicity. The stomach, because it is usually first site of contact for chemicals after oral exposure, although other areas of the gastro-intestinal tract such as the duodenum and jejunum should also be considered as site-of-contact tissues and may be considered more relevant for humans than the rodent glandular stomach. Care should be taken to ensure that such tissues are not exposed to excessively high test chemical concentrations (21). The technique is in principle applicable to any tissue from which analysable single cell/nuclei suspensions can be derived. Proprietary data from several laboratories demonstrate its successful application to many different tissues, and there are many publications showing the applicability of the technique to organs or tissues other than liver and stomach, e.g. jejunum (22), kidney (23) (24), skin (25) (26), or urinary bladder (27) (28), lungs and bronchoalveolar lavage cells (relevant for studies of inhaled chemicals) (29) (30), and tests have also been performed in multiple organs (31) (32).

Whilst there may be an interest in genotoxic effects in germ cells, it should be noted that the standard alkaline comet assay as described in this test method is not considered appropriate to measure DNA strand breaks in mature germ cells. Since high and variable background levels in DNA damage were reported in a literature review on the use of the comet assay for germ cell genotoxicity (33), protocol modifications together with improved standardization and validation studies are deemed necessary before the comet assay on mature germ cells (e.g. sperm) can be included in the test method. In addition, the recommended exposure regimen described in this test method is not optimal and longer exposures or sampling times would be necessary for a meaningful analysis of DNA strand breaks in mature sperm. Genotoxic effects as measured by the comet assay in testicular cells at different stages of differentiation have been described in the literature (34) (35). However, it should be noted that gonads contain a mixture of somatic and germ cells. For this reason, positive results in whole gonad (testis) are not necessarily reflective of germ cell damage; nevertheless, they indicate that tested chemical(s) and/or its metabolites have reached the gonad.

Cross-links cannot be reliably detected with the standard experimental conditions of the comet assay. Under certain modified experimental conditions, DNA-DNA and DNA-protein crosslinks, and other base modifications such as oxidized bases might be detected (23) (36) (37) (38) (39). But further work would be needed to adequately characterize the necessary protocol modifications. Thus detection of cross linking agents is not the primary purpose of the assay as described here. The assay is not appropriate, even with modifications, for detecting aneuploids.

Due to the current status of knowledge, several additional limitations (see Appendix 3) are associated with the *in vivo* comet assay. It is expected that the test method will be reviewed in the future and if necessary revised in light of experience gained.

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 METHOD

Animals are exposed to the test chemical by an appropriate route. A detailed description of dosing and sampling is given in paragraphs 36-40. At the selected sampling time(s), the tissues of interest are dissected and single cells/nuclei suspensions are prepared (*in situ* perfusion may be performed where considered useful e.g. liver) and embedded in soft agar so as to immobilize them on slides. Cells/nuclei are treated with lysis buffer to remove cellular and/or nuclear membrane, and exposed to strong alkali e.g. pH ≥ 13 to allow DNA unwinding and release of relaxed DNA loops and fragments. The nuclear DNA in the agar is then subjected to electrophoresis. Normal non-fragmented DNA molecules remain in the position where the nuclear DNA had been in the agar, while any fragmented DNA and relaxed DNA loops would migrate towards the anode. After electrophoresis, the DNA is visualized using an appropriate fluorescent stain. Preparations should be analysed using a microscope and full or semi-automated image analysis systems. The extent of DNA that has migrated during electrophoresis and the migration distance reflects the amount and size of DNA fragments. There are several endpoints for the comet assay. The DNA content in the tail (% tail DNA or % tail intensity) has been recommended to assess DNA damage (12) (40) (41) (42). After analysis of a sufficient number of nuclei, the data are analysed with appropriate methods to judge the assay results.

It should be noted that altering various aspects of the methodology, including sample preparation, electrophoresis conditions, visual analysis parameters (e.g. stain intensity, microscope bulb light intensity, and use of microscope filters and camera dynamics) and ambient conditions (e.g. background lighting), have been investigated and may affect DNA migration (43) (44) (45) (46).

VERIFICATION OF LABORATORY PROFICIENCY

Each laboratory should establish experimental competency in the comet assay by demonstrating the ability to obtain single cell or nuclei suspensions of sufficient quality for each target tissue(s) for each species used. The quality of the preparations will be evaluated firstly by the % tail DNA for vehicle treated animals falling within a reproducible low range. Current data suggest that the group mean % tail DNA (based on mean of medians — see paragraph 57 for details of these terms) in the rat liver should be preferably not exceed 6 %, which would be consistent with the values in the JaCVAM validation trial (12) and from other published and proprietary data. There are not enough data at this time to make recommendations about optimum or acceptable ranges for other tissues. This does not preclude the use of other tissues if justified. The test report should provide appropriate review of the performance of the comet assay in these tissues in relation to the published literature or from proprietary data. Firstly, a low range of % tail DNA in controls is desirable to provide sufficient dynamic range to detect a positive effect. Secondly, each laboratory should be able to reproduce expected responses for direct mutagens and pro-mutagens, with different modes of action as suggested in Table 1 (paragraph 29).

Positive substances may be selected, for example from the JaCVAM validation trial (12) or from other published data (see paragraph 9), if appropriate, with justification, and demonstrating clear positive responses in the tissues of interest. The ability to detect weak effects of known mutagens e.g. EMS at low doses, should also be demonstrated, for example by establishing dose-response relationships with appropriate numbers and spacing of doses. Initial efforts should focus on establishing proficiency with the most commonly used tissues e.g. the rodent liver, where comparison with existing data and expected results may be made (12). Data from other tissues e.g. stomach/duodenum/jejunum, blood etc. could be collected at the same time. The laboratory needs to demonstrate proficiency with each individual tissue in each species they are planning to study, and will need to demonstrate that an acceptable positive response with a known mutagen (e.g. EMS) can be obtained in that tissue.

Vehicle/negative control data should be collected so as to demonstrate reproducibility of negative data responses, and to ensure that the technical aspects of the assay were properly controlled or to suggest the need to re-establish historical control ranges (see paragraph 22).

It should be noted, that whilst multiple tissues can be collected at necropsy and processed for comet analysis, the laboratory needs to be proficient in harvesting multiple tissues from a single animal, thereby ensuring that any potential DNA lesion is not lost and comet analysis is not compromised. The length of time from euthanasia to removal of tissues for processing may be critical (see paragraph 44).

Animal welfare must be considered whilst developing proficiency in this test and therefore tissues from animals used in other tests can be used when developing competence in the various aspects of the test. Furthermore, it may not be necessary to conduct a full study during the stages of establishing a new test method in a laboratory and fewer animals or test concentrations can be used when developing the necessary skills.

Historical control data

During the course of the proficiency investigations, the laboratory should build a historical database to establish positive and negative control ranges and distributions for relevant tissues and species. 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 (47). Different tissues and different species, as well as different vehicles and routes of administrations, may give different negative control % tail DNA values. It is therefore important to establish negative control ranges for each tissue and species. Laboratories should use quality control methods, such as control charts (e.g. C-charts or X-bar charts (48)), to identify how variable their data are, and to show that the methodology is 'under control' in their laboratory. Selection of appropriate positive control substances, dose ranges and experimental conditions (e.g. electrophoresis conditions) may need also to be optimised for the detection of weak effects (see paragraph 17).

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.

DESCRIPTION OF THE METHOD

Preparations

Selection of animal species

Common laboratory strains of healthy young adult rodents (6-10 weeks old at start of treatment though slightly older animals are also acceptable) are normally used. The choice of rodent species should be based on (i) species used in other toxicity studies (to be able to correlate data and to allow integrated studies), (ii) species that developed tumours in a carcinogenicity study (when investigating the mechanism of carcinogenesis), or (iii) species with the most relevant metabolism for humans, if known. Rats are routinely used in this test. However, other species can be used if ethically and scientifically justified.

Animal housing and feeding conditions

For rodents, the temperature in the experimental animal room ideally should be 22 °C (\pm 3 °C). The relative humidity ideally should be 50-60 %, being at least 30 % and preferably not exceeding 70 % other than during room cleaning. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with 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 (usually no more than five) of the same sex if no aggressive behaviour is expected. Animals may be housed individually only if scientifically justified. Solid floors should be used wherever possible as mesh floors can cause serious injury (49). Appropriate environmental enrichment must be provided.

Preparation of the animals

Animals are randomly assigned to the control and treatment groups. The animals are identified uniquely and acclimated to the laboratory conditions for at least five days before the start of treatment. The least invasive method of uniquely identifying animals must be used. Appropriate methods include ringing, tagging, micro-chipping and biometric identification. Toe and ear clipping are not scientifically justified in these tests. Cages should be arranged in such a way that possible effects due to cage placement are minimized. At the commencement of the study, the weight variation of animals should be minimal and not exceed $\pm 20\%$.

Preparation of doses

Solid test chemicals should be dissolved or suspended in appropriate 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 (50) (51).

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

Vehicle

The vehicle should not produce toxic effects at the dose volumes used, and should not be suspected of chemical reaction with the test chemicals. If other than well-known vehicles are used, their inclusion should be supported with reference data indicating their compatibility in terms of test animals, route of administration and endpoint. It is recommended that wherever possible, the use of an aqueous solvent/vehicle should be considered first. It should be noted that some vehicles (particularly viscous vehicles) can induce inflammation and increase background levels of DNA strand breaks at the site of contact, particularly with multiple administrations.

Controls

Positive controls

At this time, a group of a minimum of 3 analysable animals of one sex, or of each sex if both are used (see paragraph 32), treated with a positive control substance should normally be included with each test. In future, it may be possible to demonstrate adequate proficiency to reduce the need for positive controls. If multiple sampling times are used (e.g. with a single administration protocol) it is only necessary to include positive controls at one of the sampling times, but a balanced design should be ensured (see paragraph 48). It is not necessary to administer concurrent positive control substances by the same route as the test chemical, although it is important that the same route should be used when measuring site-of-contact effects. The positive control substances should be shown to induce DNA strand breaks in all of the tissues of interest for the test chemical, and EMS is likely to be the positive control of choice since it has produced DNA strand breaks in all tissues that have been studied. The doses of the positive control substances should be selected so as to produce moderate effects that critically assess the performance and sensitivity of the assay and could be based on dose-response curves established by the laboratory during the demonstration of proficiency. The % tail DNA in concurrent positive control animals should be consistent with the pre-established laboratory range for each individual tissue and sampling time for that species (see paragraph 16). Examples of positive control substances and some of their target tissues (in rodents) are included in Table 1. Substances other than those given in Table 1 can be selected if scientifically justified.

Table 1

Examples of positive control substances and some of their target tissues

Substances and CAS RN No.
Ethyl methanesulfonate (CAS RN 62-50-0) for any tissue
Ethyl nitrosourea (CAS RN 759-73-9) for liver and stomach, duodenum or jejunum
Methyl methanesulfonate (CAS RN 66-27-3) for liver, stomach, duodenum or jejunum, lung and bronchoalveolar lavage (BAL) cells, kidney, bladder, lung, testis and bone marrow/blood
N-Methyl-N'-nitro-N-nitrosoguanidine (CAS RN: 70-25-7) for stomach, duodenum or jejunum
1,2-Dimethylhydrazine 2HCl (CAS RN 306-37-6) for liver and intestine
N-methyl-N-nitrosourea (CAS RN 684-93-5) for liver, bone marrow, blood, kidney, stomach, jejunum, and brain.

Negative controls

A group of negative control animals, treated with vehicle alone, and otherwise treated in the same way as the treatment groups, should be included with each test for every sampling time and tissue. The % tail DNA in negative control animals should be within the pre-established laboratory background range for each individual tissue and sampling time for that species (see paragraph 16). In the absence of historical or published control data showing that no deleterious or genotoxic effects are induced by the chosen vehicle, by the number of administrations or by the route of administration, initial studies should be performed prior to conducting the full study, in order to establish acceptability of the vehicle control.

PROCEDURE

Number and Sex of Animals

Although there is little data on female animals from which to make comparison between sexes in relation to the comet assay, in general, other *in vivo* genotoxicity responses are similar between male and female animals and therefore most studies could be performed in either sex. Data demonstrating relevant differences between males and females (e.g. differences in systemic toxicity, metabolism, bioavailability, etc. including e.g. in a range-finding study) encourage the use of both sexes. In this case, it may be appropriate to perform a study in both sexes e.g. as part of a repeated dose toxicity study. It might be appropriate to use the factorial design in case both sexes are used. Details on how to analyse the data using this design are given in Appendix 2.

Group sizes at study initiation (and during establishment of proficiency) should be established with the aim of providing a minimum of 5 analysable animals of one sex, or of each sex if both are used, per group (less in the concurrent positive control group — see paragraph 29). Where human exposure to chemicals may be sex-specific, as for example with some pharmaceuticals, the test should be performed with the appropriate sex. As a guide to maximum typical animal requirements, a study conducted according the parameters established in paragraph 33 with three dose groups and concurrent negative and positive controls (each group composed of five animals of a single sex) would require between 25 and 35 animals.

TREATMENT SCHEDULE

Animals should be given daily treatments over a duration of 2 or more days (i.e. two or more treatments at approximately 24 hour intervals), and samples should be collected once at 2-6 h (or at the T_{max}) after the last treatment (12). Samples from extended dose regimens (e.g. 28-day daily dosing) are acceptable. Successful combination of the comet and the erythrocyte micronucleus test has been demonstrated (10) (19). However careful consideration should be given to the logistics involved in tissue sampling for comet analysis alongside the requirements of tissue sampling for other types of toxicological assessments. Harvest 24 hours after the last dose, which is typical of a general toxicity study, is not appropriate in most cases (see paragraph 40 on sampling time). The use of other treatment and sampling schedules should be justified (see Appendix 3). For example single treatment with multiple sampling could be used however, it should be noted that more animals will be required for a study with a single administration study because of the need for multiple sampling times, but on occasions this may be preferable, e.g. when the test chemical induces excessive toxicity following repeated administrations.

Whatever way the test is performed, it is acceptable as long as the test chemical gives a positive response or, for a negative study, as long as direct or indirect evidence supportive of exposure of, or toxicity to, the target tissue(s) has been demonstrated or if the limit dose is achieved (see paragraph 36).

Test chemicals also may be administered as a split dose, i.e., two treatments on the same day separated by no more than 2-3 hours, to facilitate administering a large volume. Under these circumstances, the sampling time should be scheduled based on the time of the last dosing (see paragraph 40).

Dose Levels

If a preliminary range-finding study is performed because there are no suitable data available from other relevant studies 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 according to current approaches for conducting dose range-finding studies. The 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, clear clinical signs such as abnormal behaviour or reactions, minor body weight depression or target tissue cytotoxicity), but not death or evidence of pain, suffering or distress necessitating euthanasia. For a non-toxic test chemical, with an administration period of 14 days or more, the maximum (limit) dose is 1 000 mg/kg bodyweight/day. For administration periods of less than 14 days the maximum (limit) dose is 2 000 mg/kg bodyweight/day. For certain types of test chemicals (e.g. human pharmaceuticals) covered by specific regulations these limits may vary.

Chemicals that exhibit saturation of toxicokinetic properties, or induce detoxification processes that may lead to a decrease in exposure after long-term administration, may be exceptions to the dose-setting criteria and should be evaluated on a case-by-case basis.

For both acute and sub-acute versions of the comet assay, in addition to the maximum dose (MTD, maximum feasible dose, maximum exposure or limit dose) a descending sequence of at least two additional appropriately spaced dose levels (preferably separated by less than $\sqrt{10}$) should be selected for each sampling time to demonstrate dose-related responses. However, the dose levels used should also preferably cover a range from the maximum to one producing little or no toxicity. When target tissue toxicity is observed at all dose levels tested, further study at non-toxic doses is advisable (see paragraphs 54-55). Studies intending to more fully investigate the shape of the dose-response curve may require additional dose group(s).

Administration of Doses

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, intratracheal, 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 generally not recommended since it is not a typical relevant route of human exposure, and should only be used with specific justification (e.g. some positive control substances, for investigative purposes, or for some drugs that are administered by the intraperitoneal route). The maximum volume of liquid that can be administered by gavage or injection at one time depends on the size of the test animal. The volume should not exceed 1 ml/100 g body weight, except in the case of aqueous solutions where 2 ml/100g body weight may be used. The use of volumes greater than this (if permitted by animal welfare legislation) should be justified. Wherever possible different dose levels should be achieved by adjusting the concentration of the dosing formulation to ensure a constant volume in relation to body weight at all dose levels.

Sampling Time

The sampling time is a critical variable because it is determined by the period needed for the test chemicals to reach maximum concentration in the target tissue and for DNA strand breaks to be induced but before those breaks are removed, repaired or lead to cell death. The persistence of some of the lesions that lead to the DNA strand breaks detected by the comet assay may be very short, at least for some chemicals tested *in vitro* (52) (53). Accordingly, if such transient DNA lesions are suspected, measures should be taken to mitigate their loss by ensuring that tissues are sampled sufficiently early, possibly earlier than the default times given below. The optimum sampling time(s) may be chemical- or route-specific resulting in, for example, rapid tissue exposure with intravenous administration or inhalation exposure. Accordingly, where available, sampling times should be determined from kinetic data (e.g. the time (T_{max}) at which the peak plasma or tissue concentration (C_{max}) is achieved, or at the steady state for multiple administrations). In the absence of kinetic data a suitable compromise for the measurement of genotoxicity is to sample at 2-6 h after the last treatment for two or more treatments, or at both 2-6 and 16-26 h after a single administration, although care should be taken to necropsy all animals at the same time after the last (or only) dose. Information on the appearance of toxic effects in target organs (if available) may also be used to select appropriate sampling times.

Observations

General clinical observations related to the health of the animals should be made and recorded at least once a day preferably at the same time(s) each day and considering the peak period of anticipated effects after dosing (54). At least twice daily, all animals should be observed for morbidity and mortality. For longer duration studies, all animals should be weighed at least once a week, and at completion of the test period. Food consumption should be measured at each change of food and 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 excessive toxicity should be euthanized prior to completion of the test period, and are generally not used for comet analysis.

Tissue Collection

Since it is possible to study induction of DNA strand breaks (comets) in virtually any tissue, the rationale for selection of tissue(s) to be collected should be clearly defined and based upon the reason for conducting the study together with any existing ADME, genotoxicity, carcinogenicity or other toxicity data for the test chemicals under investigation. Important factors for consideration should include the route of administration (based on likely human exposure route(s)), the predicted tissue distribution and absorption, the role of metabolism and the possible mechanism of action of the test chemicals. The liver has been the tissue most frequently studied and for which there are the most data. Therefore, in the absence of any background information, and if no specific tissues of

interest are identified, sampling the liver would be justified as this is a primary site of xenobiotic metabolism and is often highly exposed to both parent substance(s) and metabolite(s). In some cases, examination of a site of direct contact (for example, for orally-administered chemicals the glandular stomach or duodenum/jejunum, or for inhaled chemicals the lungs) may be most relevant. Additional or alternative tissues should be selected based on the specific reasons for the test is being conducted but it may be useful to examine multiple tissues in the same animals providing the laboratory has demonstrated proficiency with those tissues and competency in handling multiple tissues at the same time.

Preparation of specimens

For the processes described in the following paragraphs (44-49) it is important that all solutions or stable suspensions should be used within their expiration date, or should be freshly prepared if needed. Also in the following paragraphs, the times taken to (i) remove each tissue after necropsy, (ii) process each tissue into cell/nuclei suspensions, and (iii) process the suspension and prepare the slides are all considered critical variables (see Definitions, Appendix 1), and acceptable lengths of time for each of these steps should have been determined during establishment of the method and demonstration of proficiency.

Animals will be euthanised, consistent with effective animal welfare legislation and 3Rs principles, at the appropriate time(s) after the last treatment with a test chemical. Selected tissue(s) is removed, dissected, and a portion is collected for the comet assay, whilst at the same time a section from the same part of the tissue should be cut and placed in formaldehyde solution or appropriate fixative for possible histopathology analysis (see paragraph 55) according to standard methods (12). The tissue for the comet assay is placed into mincing buffer, rinsed sufficiently with cold mincing buffer to remove residual blood, and stored in ice-cold mincing buffer until processed. *In situ* perfusion may also be performed, e.g. for liver, kidney.

Many published methods exist for cell/nuclei isolation. These include mincing of tissues such as liver and kidney, scraping mucosal surfaces in the case of the gastro-intestinal tract, homogenization and enzymic digestion. The JaCVAM validation trial only studied isolated cells, and therefore in terms of establishing the method and being able to refer to the JaCVAM trial data for demonstration of proficiency, isolated cells are preferred. However, it has been shown that there was no essential difference in the assay result whether isolated cells or nuclei were used (8). Also different methods to isolate cells/nuclei (e.g. homogenizing, mincing, enzymic digestion and mesh filtration) gave comparable results (55). Consequently, either isolated cells or isolated nuclei can be used. A laboratory should thoroughly evaluate and validate tissue-specific methods of single cell/nuclei isolation. As discussed in paragraph 40, the persistence of some of the lesions that lead to the DNA strand breaks detected by the comet assay may be very short (52) (53). Therefore, whatever method is used to prepare the single cell/nuclei suspensions, it is important that tissues are processed as soon as possible after the animals have been euthanised and placed in conditions that reduce the removal of lesions (e.g. by maintaining the tissue at low temperature). The cell suspensions should be kept ice-cold until ready for use, so that minimal inter-sample variation and appropriate positive and negative control responses can be demonstrated.

PREPARATION OF SLIDES

Slide preparation should be done as soon as possible (ideally within one hour) after single cell/nuclei preparation, but the temperature and time between animal death and slide preparation should be tightly controlled and validated under the laboratory's conditions. The volume of the cell suspension added to low melting point agarose (usually 0.5-1.0 %) to make the slides should not reduce the percentage of low melting point agarose to less than 0.45 %. The optimum cell density will be determined by the image analysis system used for scoring comets.

Lysis

Lysis conditions are also a critical variable and may interfere with the strand breaks resulting from specific types of DNA modifications (certain DNA alkylations and base adducts). It is therefore recommended that the lysis conditions be kept as constant as possible for all slides within an experiment. Once prepared, the slides should be immersed in chilled lysing solution for at least one hour (or overnight) at around 2-8 °C under subdued lighting conditions e.g. yellow light (or light proof) that avoid exposure to white light that may contain UV components. After this incubation period, the slides should be rinsed to remove residual detergent and salts prior to the alkali unwinding step. This can be done using purified water, neutralization buffer or phosphate buffer. Electrophoresis buffer can also be used. This would maintain the alkaline conditions in the electrophoresis chamber.

Unwinding and electrophoresis

Slides should be randomly placed onto the platform of a submarine-type electrophoresis unit containing sufficient electrophoresis solution such that the surfaces of the slides are completely covered (the depth of covering should also be consistent from run to run). In another type of comet assay electrophoresis units i.e. with active cooling, circulation and high capacity power supply a higher solution covering will result in higher electric current while the voltage is kept constant. A balanced design should be used to place slides in the electrophoresis tank to mitigate the effects of any trends or edge effect within the tank and to minimize batch-to-batch variability, i.e., in each electrophoresis run, there should be the same number of slides from each animal in the study and samples from the different dosage groups, negative and positive controls, should be included. The slides should be left for at least 20 minutes for the DNA to unwind, and then subjected to electrophoresis under controlled conditions that will maximize the sensitivity and dynamic range of the assay (i.e. lead to acceptable levels of % tail DNA for negative and positive controls that maximize sensitivity). The level of DNA migration is linearly associated with the duration of electrophoresis, and also with the potential (V/cm). Based on the JaCVAM trial this could be 0,7 V/cm for at least 20 minutes. The duration of electrophoresis is considered a critical variable and the electrophoresis time should be set to optimize the dynamic range. Longer electrophoresis times (e.g. 30 or 40 minutes to maximize sensitivity) usually lead to stronger positive responses with known mutagens. However longer electrophoresis times may also lead to excessive migration in control samples. In each experiment the voltage should be kept constant, and the variability in the other parameters should be within a narrow and specified range, for example in the JaCVAM trial 0,7 V/cm delivered a starting current of 300 mA. The depth of buffer should be adjusted to achieve the required conditions and maintained throughout the experiment. The current at the start and end of the electrophoresis period should be recorded. The optimum conditions should therefore be determined during the initial demonstration of proficiency in the laboratory concerned with each tissue studied. The temperature of the electrophoresis solution through unwinding and electrophoresis should be maintained at a low temperature, usually 2-10 °C (10). The temperature of the electrophoresis solution at the start of unwinding, the start of electrophoresis, and the end of electrophoresis should be recorded.

After completion of electrophoresis, the slides should be immersed/rinsed in the neutralization buffer for at least 5 minutes. Gels can be stained and scored "fresh" (e.g. within 1-2 days) or can be dehydrated for later scoring (e.g. within 1-2 weeks after staining) (56). However, the conditions should be validated during the demonstration of proficiency and historical data should be obtained and retained separately for each of these conditions. In case of the latter, slides should be dehydrated by immersion into absolute ethanol for at least 5 minutes, allowed to air dry, and then stored, either at room temperature or in a container in a refrigerator until scored.

Methods of Measurement

Comets should be scored quantitatively using an automated or semi-automated image-analysis system. The slides will be stained with an appropriate fluorescent stain e.g. SYBR Gold, Green I, propidium iodide or ethidium bromide and measured at a suitable magnification (e.g. 200x) on a microscope equipped with epi-fluorescence and appropriate detectors or a digital (e.g. CCD) camera.

Cells may be classified into three categories as described in the atlas of comet images (57), namely scorable, non-scorable and “hedgehog” (see paragraph 56 for further discussion). Only scorable cells (clearly defined head and tail with no interference with neighbouring cells) should be scored for % tail DNA to avoid artefacts. There is no need to report the frequency of non-scorable cells. The frequency of hedgehogs should be determined based on the visual scoring (since the absence of a clearly-defined head will mean they are not readily detected by image analysis) of at least 150 cells per sample (see paragraph 56 for further discussion) and separately documented.

All slides for analysis, including those of positive and negative controls, should be independently coded and scored “blinded” so the scorer is unaware of the treatment condition. For each sample (per tissue per animal), at least 150 cells (excluding hedgehogs — see paragraph 56) should be analysed. Scoring 150 cells per animal in at least 5 animals per dose (less in the concurrent positive control — see paragraph 29) provides adequate statistical power according to the analysis of Smith *et al.*, 2008 (5). If slides are used, this could be from 2 or 3 slides scored per sample when five animals per group are used. Several areas of the slide should be observed at a density that ensures there is no overlapping of tails. Scoring at the edge of slides should be avoided.

DNA strand breaks in the comet assay can be measured by independent endpoints such as % tail DNA, tail length and tail moment. All three measurements can be made if the appropriate image software analyser system is used. However, the % tail DNA (also known as % tail intensity) is recommended for the evaluation and interpretation of results (12) (40) (41) (42), and is determined by the DNA fragment intensity in the tail expressed as a percentage of the cell's total intensity (13).

Tissue damage and cytotoxicity

Positive findings in the comet assay may not be solely due to genotoxicity, target tissue toxicity may also result in increases in DNA migration (12) (41). Conversely, low or moderate cytotoxicity is often seen with known genotoxins (12), showing that it is not possible to distinguish DNA migration induced by genotoxicity versus that induced by cytotoxicity in the comet assay alone. However, where increases in DNA migration are observed, it is recommended that an examination of one or more indicators of cytotoxicity is performed as this can aid in interpretation of the findings. Increases in DNA migration in the presence of clear evidence of cytotoxicity should be interpreted with caution.

Many measures of cytotoxicity have been proposed and of these histopathological changes are considered a relevant measure of tissue toxicity. Observations such as inflammation, cell infiltration, apoptotic or necrotic changes have been associated with increases in DNA migration, however, as demonstrated by the JaCVAM validation trial (12) no definitive list of histopathological changes that are always associated with increased DNA migration is available. Changes in clinical chemistry measures (e.g. AST, ALT), can also provide useful information on tissue damage and additional indicators such as caspase activation, TUNEL stain, Annexin V stain, etc. may also be considered. However, there are limited published data where the latter have been used for *in vivo* studies and some may be less reliable than others.

Hedgehogs (or clouds, ghost cells) are cells that exhibit a microscopic image consisting of a small or non-existent head, and large diffuse tails and are considered to be heavily damaged cells, although the etiology of the hedgehogs is uncertain (see Appendix 3). Due to their appearance, % tail DNA measurements by image analysis are unreliable and therefore hedgehogs should be evaluated separately. The occurrence of hedgehogs should be noted and reported and any relevant increase thought to be due to the test chemicals should be investigated and interpreted with care. Knowledge of the potential mode of action of the test chemicals may help with such considerations.

DATA AND REPORTING

Treatment of Results

The animal is the experimental unit and therefore both individual animal data and summarized results should be presented in tabular form. Due to the hierarchical nature of the data it is recommended that the median %tail DNA for each slide is determined and the mean of the median values is calculated for each animal (12). The mean of the individual animal means is then determined to give a group mean. All of these values should be included in the report. Alternative approaches (see paragraph 53) may be used if scientifically and statistically justified. Statistical analysis can be done using a variety of approaches (58) (59) (60) (61). When selecting the statistical methods to be used, the need for transformation (e.g. log or square root) of the data and/or addition of a small number (e.g. 0,001) to all (even non-zero) values to mitigate the effects of zero cell values, should be considered as discussed in the above references. Details of analysis of treatment/sex interactions when both sexes are used, and subsequent analysis of data where either differences or no differences are found is given in Appendix 2. Data on toxicity and clinical signs should also be reported.

Acceptability Criteria

Acceptance of a test is based on the following criteria:

- (a) The concurrent negative control is considered acceptable for addition to the laboratory historical negative control database as described in paragraph 16
- (b) Concurrent positive controls (see paragraph 29) should induce responses that are compatible with those generated in the historical positive control database and produce a statistically significant increase compared with the concurrent negative control.
- (c) Adequate numbers of cells and doses have been analysed (paragraphs 52 and 36-38).
- (d) The criteria for the selection of highest dose are consistent with those described in paragraph 36.

Evaluation and Interpretation of Results

Providing that all acceptability criteria are fulfilled, a test chemical is considered to be clearly positive if:

- (a) at least one of the test doses exhibits a statistically significant increase compared with the concurrent negative control,
- (b) the increase is dose-related when evaluated with an appropriate trend test,
- (c) any of the results are outside the distribution of the historical negative control data for a given species, vehicle, route, tissue, and number of administrations.

When all of these criteria are met, the test chemical is then considered able to induce DNA strand breakage in the tissues studied in this test system. If only one or two of these criteria are satisfied, see paragraph 62.

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

- (a) none of the test concentrations exhibits a statistically significant increase compared with the concurrent negative control,
- (b) there is no concentration-related increase when evaluated with an appropriate trend test.
- (c) all results are inside the distribution of the historical negative control data for a given species, vehicle, route, tissue, and number of administrations.
- (d) direct or indirect evidence supportive of exposure of, or toxicity to, the target tissue(s) has been demonstrated.

The test chemical is then considered unable to induce DNA strand breakage in the tissues studied in this test system.

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

In case the response is neither clearly negative nor clearly positive (i.e. not all the criteria listed in paragraphs 59 or 60 are met) and in order to assist in establishing the biological relevance of a result, the data should be evaluated by expert judgement and/or further investigations conducted, if scientifically justified. Scoring additional cells (where appropriate) or performing a repeat experiment possibly using optimised experimental conditions (e.g. dose spacing, other routes of administration, other sampling times or other tissues) could be useful.

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.

To assess the biological relevance of a positive or equivocal result, information on cytotoxicity at the target tissue is required (see paragraphs 54-55). Where positive or equivocal findings are observed solely in the presence of clear evidence of cytotoxicity, the study would be concluded as equivocal for genotoxicity unless there is enough information that is supportive of a definitive conclusion. In cases of a negative study outcome where there are signs of toxicity at all doses tested, further study at non-toxic doses may be advisable.

Test Report

The test report should include the following information:

Test chemical:

- source, lot number if available;
- stability of the test chemical, limit date for use, or date for re-analysis 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.

Solvent/vehicle:

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

Test animals:

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

- source, housing conditions, diet, enrichment, etc.;
- individual weight of the animals at the start and at the end of the test, including body weight range, mean and standard deviation for each group.

Test conditions:

- positive and negative (vehicle/solvent) control data;
- results from the range-finding study (if conducted);
- rationale for dose level selection;
- details of test chemical preparation;
- details of the administration of the test chemical;
- rationale for route of administration;
- site of injection (for subcutaneous or intravenous studies);
- methods for sample preparation, where available, histopathological analyses, especially for a chemical giving a positive comet response;
- rationale for tissue selection;
- 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 diet and water quality;
- detailed description of treatment and sampling schedules and justifications for the choices (e.g. toxicokinetic data, where available);
- method of pain relief, analgesia;
- method of euthanasia;
- procedures for isolating and preserving tissues;
- methods for preparing single cell/nucleus suspension;
- source and lot numbers of all reagents (where possible);
- methods for evaluating cytotoxicity;
- electrophoresis conditions;
- staining techniques used; and
- methods for scoring and measuring comets.

Results:

- General clinical observations, if any, prior to and throughout the test period for each animal;
- evidence of cytotoxicity if performed;
- for studies longer than one week: Individual body weights during the study, including body weight range, mean and standard deviation for each group; food consumption;

- dose-response relationship, where evident;
- for each tissue/animal, the % tail DNA (or other measures, if chosen) and median values per slide, mean values per animal and mean values per group;
- concurrent and historical negative control data with ranges, means/medians and standard deviations for each tissue evaluated;
- concurrent and historical positive control data;
- for tissues other than liver, a dose-response curve using the positive control. This can be from data collected during the demonstration of proficiency (see paragraphs 16-17) and should be accompanied by a justification, with citations to current literature, for the appropriateness of the magnitude and scatter of the responses to the controls in that tissue;
- statistical analyses and methods applied; and criteria for considering a response as positive, negative or equivocal;
- frequency of hedgehogs in each group and per animal.

Discussion of the results

Conclusion

References

LITERATURE:

- (1) Kirkland, D., G. Speit (2008), Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens III. Appropriate follow-up testing *in vivo*, *Mutation Research*, Vol. 654/2, pp. 114-32.
- (2) Brendler-Schwaab, S. *et al.* (2005), The *in vivo* Comet assay: use and status in genotoxicity testing, *Mutagenesis*, Vol. 20/4, pp. 245-54.
- (3) Burlinson, B. *et al.* (2007), Fourth International Workgroup on Genotoxicity Testing: result of the *in vivo* Comet assay workgroup, *Mutation Research*, Vol. 627/1, pp. 31-5.
- (4) Burlinson, B. (2012), The *in vitro* and *in vivo* Comet assays, *Methods in Molecular Biology*, Vol. 817, pp. 143-63.
- (5) Smith, C.C. *et al.* (2008), Recommendations for design of the rat Comet assay, *Mutagenesis*, Vol. 23/3, pp. 233-40.
- (6) Hartmann, A. *et al.* (2003), Recommendations for conducting the *in vivo* alkaline Comet assay, *Mutagenesis*, Vol. 18/1, pp. 45-51.
- (7) McKelvey-Martin, V.J. *et al.* (1993), The single cell gel electrophoresis assay (Comet assay): a European review, *Mutation Research*, Vol. 288/1, pp. 47-63.
- (8) Tice, R.R. *et al.* (2000), Single cell gel/Comet assay: guidelines for *in vitro* and *in vivo* genetic toxicology testing, *Environmental and Molecular Mutagenesis*, Vol. 35/3, pp. 206-21.
- (9) Singh, N.P. *et al.* (1988), A simple technique for quantitation of low levels of DNA damage in individual cells, *Experimental Cell Research*, Vol. 175/1, pp. 184-91.
- (10) Rothfuss, A. *et al.* (2010), Collaborative study on fifteen compounds in the rat-liver Comet assay integrated into 2- and 4-week repeat-dose studies, *Mutation Research*, Vol., 702/1, pp. 40-69.
- (11) 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.

- (12) OECD (2014), *Reports of the JaCVAM initiative international pre-validation and validation studies of the in vivo rodent alkaline comet assay for the detection of genotoxic carcinogens*, Series on Testing and Assessment, Nos. 195 and 196, OECD Publishing, Paris.
- (13) Olive, P.L., J.P. Banath, R.E. Durand (1990), Heterogeneity in radiation-induced DNA damage and repair in tumor and normal cells using the "Comet" assay, *Radiation Research*, Vol. 122/1, pp. 86-94.
- (14) Tice, R.R., G.H. Strauss (1995), The single cell gel electrophoresis/Comet assay: a potential tool for detecting radiation-induced DNA damage in humans, *Stem Cells*, Vol. 13/1, pp. 207-14.
- (15) Collins, A.R (2004), The Comet assay for DNA damage and repair: principles, applications, and limitations, *Molecular Biotechnology*, Vol. 26/3, pp. 249-61.
- (16) Rothfuss, A. *et al.* (2011), Improvement of *in vivo* genotoxicity assessment: combination of acute tests and integration into standard toxicity testing, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 723/2, pp. 108-20.
- (17) Kushwaha, S. *et al.* (2010), Evaluation of multi-organ DNA damage by Comet assay from 28 days repeated dose oral toxicity test in mice: A practical approach for test integration in regulatory toxicity testing, *Regulatory Toxicology and Pharmacology*, Vol. 58/1, pp. 145-54.
- (18) Vasquez, M.Z. (2010), Combining the *in vivo* Comet and micronucleus assays: a practical approach to genotoxicity testing and data interpretation, *Mutagenesis*, Vol. 25/2, pp. 187-99.
- (19) Bowen, D.E. (2011), Evaluation of a multi-endpoint assay in rats, combining the bone-marrow micronucleus test, the Comet assay and the flow-cytometric peripheral blood micronucleus test, *Mutation Research*, Vol. 722/1, pp. 7-19.
- (20) Recio, L. *et al.* (2010), Dose-response assessment of four genotoxic chemicals in a combined mouse and rat micronucleus (MN) and Comet assay protocol, *The Journal of Toxicological Science*, Vol. 35/2, pp. 149-62.
- (21) O'Donovan, M., B. Burlinson (2013), Maximum dose levels for the rodent comet assay to examine damage at the site of contact or to the gastrointestinal tract, *Mutagenesis*, Vol. 28/6, pp. 621-3.
- (22) Hartmann, A. (2004), **Use of the alkaline *in vivo* Comet assay for mechanistic genotoxicity investigations**, *Mutagenesis*, Vol. 19/1, pp. 51-9.
- (23) Nessler, F. (2007), *In vivo* Comet assay on isolated kidney cells to distinguish genotoxic carcinogens from epigenetic carcinogens or cytotoxic compounds, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 630/1, pp. 28-41.
- (24) Brendler-Schwaab, S.Y., B.A. Herbold (1997), A new method for the enrichment of single renal proximal tubular cells and their first use in the Comet assay, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 393/1-2, pp. 175-8.
- (25) Toyozumi, T. *et al.* (2011), Use of the *in vivo* skin Comet assay to evaluate the DNA-damaging potential of chemicals applied to the skin, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 726/2, pp. 175-80.
- (26) Struwe, M. *et al.* (2008), Detection of photogenotoxicity in skin and eye in rat with the photo Comet assay, *Photochemical and Photobiological Sciences*, Vol. 7/2, pp. 240-9.
- (27) Wada, K. *et al.* (2012), A comparison of cell-collecting methods for the Comet assay in urinary bladders of rats, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 742/1-2, pp. 26-30.
- (28) Wang, A. *et al.* (2007), Measurement of DNA damage in rat urinary bladder transitional cells: improved selective harvest of transitional cells and detailed Comet assay protocols, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 634/ 1-2, pp. 51-9.

- (29) Burlinson, B. *et al.* (2007), *In Vivo Comet Assay Workgroup*, part of the Fourth International Workgroup on Genotoxicity Testing. Fourth International Workgroup on Genotoxicity testing: results of the *in vivo* Comet assay workgroup, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 627/1, pp. 31-5.
- (30) Jackson, P. *et al.* (2012), Pulmonary exposure to carbon black by inhalation or instillation in pregnant mice: effects on liver DNA strand breaks in dams and offspring, *Nanotoxicology*, Vol. 6/5, pp. 486-500.
- (31) Sasaki, Y.F. *et al.* (2000), The comet assay with multiple mouse organs: comparison of Comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database, *Critical Reviews in Toxicology*, Vol. 30/6, pp. 629-799.
- (32) Sekihashi, K. *et al.* (2002), Comparative investigations of multiple organs of mice and rats in the Comet assay, *Mutation Research*, Vol. 517/1-2, pp. 53-74.
- (33) Speit, G, M. Vasquez, A. Hartmann (2009), **The comet assay as an indicator test for germ cell genotoxicity**, *Mutation Research*, Vol. 681/1, pp. 3-12.
- (34) Zheng, H., P.L. Olive (1997), Influence of oxygen on radiation-induced DNA damage in testicular cells of C3H mice, *International Journal of Radiation Biology*, Vol. 71/3, pp. 275-282.
- (35) Cordelli, E. *et al.* (2003), Evaluation of DNA damage in different stages of mouse spermatogenesis after testicular X irradiation, *Journal of Radiation Research*, Vol. 160/4, pp. 443-451.
- (36) Merk, O., G. Speit (1999), Detection of crosslinks with the Comet assay in relationship to genotoxicity and cytotoxicity, *Environmental and Molecular Mutagenesis*, Vol. 33/2, pp. 167-72.
- (37) Pfuhler, S., H.U. Wolf (1996), Detection of DNA-crosslinking agents with the alkaline Comet assay, *Environmental and Molecular Mutagenesis*, Vol. 27/3, pp. 196-201.
- (38) Wu, J.H., N.J. Jones (2012), Assessment of DNA interstrand crosslinks using the modified alkaline Comet assay, *Methods in Molecular Biology*, Vol. 817, pp. 165-81.
- (39) Spanswick, V.J., J.M. Hartley, J.A. Hartley (2010), **Measurement of DNA interstrand crosslinking in individual cells using the Single Cell Gel Electrophoresis (Comet) assay**, *Methods in Molecular Biology*, Vol. 613, pp. 267-282.
- (40) Kumaravel, T.S., A.N. Jha (2006), Reliable Comet assay measurements for detecting DNA damage induced by ionizing radiation and chemicals, *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, Vol. 605 (1-2), pp. 7-16.
- (41) Burlinson, B. *et al.* (2007), Fourth International Workgroup on Genotoxicity Testing: result of the *in vivo* Comet assay workgroup, *Mutation Research*, Vol.627/1, pp. 31-5.
- (42) Kumaravel, T.S. *et al.* (2009), Comet Assay measurements: a perspective, *Cell Biology and Toxicology*, Vol. 25/1, pp. 53-64.
- (43) Ersson, C., L. Möller (2011), The effects on DNA migration of altering parameters in the Comet assay protocol such as agarose density, electrophoresis conditions and durations of the enzyme or the alkaline treatments, *Mutagenesis*, Vol. 26/6, pp. 689-95.
- (44) Møller, P. *et al.* (2010), Assessment and reduction of Comet assay variation in relation to DNA damage: studies from the European Comet Assay Validation Group, *Mutagenesis*, Vol. 25/2, pp. 109-11.
- (45) Forchhammer, L. *et al.* (2010), Variation in the measurement of DNA damage by Comet assay measured by the ECVAG inter-laboratory validation trial, *Mutagenesis*, Vol. 25/2, pp. 113-23.
- (46) Azqueta, A. *et al.* (2011), Towards a more reliable comet assay: Optimising agarose concentration, unwinding time and electrophoresis conditions, *Mutation Research*, Vol. 724/1-2, pp. 41-45.

- (47) Hayashi, M. *et al.* (2011), Compilation and use of genetic toxicity historical control data, *Mutation Research*, Vol. 723/2, pp. 87-90.
- (48) Ryan, T. P. (2000), *Statistical Methods for Quality Improvement*, John Wiley and Sons, New York 2nd ed.
- (49) Appendix A of the European Convention for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes (ETS No. 123)
- (50) Chapter B.8 of this Annex: *Subacute Inhalation Toxicity: 28-Day Study*.
- (51) Chapter B.29 of this Annex: *Subchronic Inhalation Toxicity: 90-day Study*.
- (52) Blakey, D.H., G.R. Douglas (1984), Transient DNA lesions induced by benzo[a]pyrene in Chinese hamster ovary cells, *Mutation Research*, Vol. 140/2-3, pp. 141-45.
- (53) Blakey, D.H., G.R. Douglas (1990), The role of excision repair in the removal of transient benzo[a]pyrene-induced DNA lesions in Chinese hamster ovary cells, *Mutation Research*, Vol. 236/1, pp. 35-41.
- (54) OECD (2002), "Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 19, OECD Publishing, Paris.
- (55) Nakajima, M. (2012), Tissue sample preparation for *in vivo* rodent alkaline Comet assay, *Genes and Environment*, Vol. 34/1, pp. 50-4.
- (56) Hartmann, A. *et al.* (2003), Recommendations for conducting the *in vivo* alkaline Comet assay, *Mutagenesis*, Vol.18/1, pp.45-51.
- (57) Atlas of Comet Assay Images, Scientist Press Co., Ltd., Tokyo, Japan.
- (58) Lovell, D.P., G. Thomas, R. Dubow (1999), Issues related to the experimental design and subsequent statistical analysis of *in vivo* and *in vitro* Comet studies, *Teratogenesis Carcinogenesis Mutagenesis*, Vol. 19/2, pp. 109-19.
- (59) Wiklund, S.J., E. Agurell (2003), Aspects of design and statistical analysis in the Comet assay, *Mutagenesis*, Vol. 18/2, pp. 167-75.
- (60) Bright, J. *et al.* (2011), Recommendations on the statistical analysis of the Comet assay, *Pharmaceutical Statistics*, Vol. 10/6, pp. 485-93.
- (61) Lovell, D.P., T. Omori (2008), Statistical issues in the use of the Comet assay, *Mutagenesis*, Vol. 23/3, pp. 171-82.
-

Appendix 1

DEFINITIONS:

Alkaline single cell gel electrophoresis: Sensitive technique for the detection of primary DNA damage at the level of individual cell/nucleus.

Chemical: A substance or a mixture.

Comet: The shape that nucleoids adopt after submitted to one electrophoretic field, due to its similarity to comets: the head is the nucleus and the tail is constituted by the DNA migrating out of the nucleus in the electric field.

A critical variable/parameter: This is a protocol variable for which a small change can have a large impact on the conclusion of the assay. Critical variables can be tissue-specific. Critical variables should not be altered, especially within a test, without consideration of how the alteration will alter an assay response, for example as indicated by the magnitude and variability in positive and negative controls. The test report should list alterations of critical variables made during the test or compared to the standard protocol for the laboratory and provide a justification for each alteration.

Tail intensity or % tail DNA: This corresponds to the intensity of the comet tail relative to the total intensity (head plus tail). It reflects the amount of DNA breakage, expressed as a percentage.

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

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

*Appendix 2***THE FACTORIAL DESIGN FOR IDENTIFYING SEX DIFFERENCES IN THE IN VIVO COMET ASSAY****The factorial design and its analysis**

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

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

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

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

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

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

References

There are many statistical textbooks which discuss the theory, design, methodology, analysis and interpretation of factorial designs ranging from the simplest two factor analyses to the more complex forms used in Design of Experiment methodology. The following is a non-exhaustive list. Some books provide worked examples of comparable designs, in some cases with code for running the analyses using various software packages.

- (1) Box, G.E.P, Hunter, W.G. and Hunter, J.S. (1978). Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building. New York: John Wiley & Sons.
- (2) Box G.E.P. & Draper, N.R. (1987) Empirical model-building and response surfaces. John Wiley & Sons Inc.
- (3) Doncaster, C.P. & Davey, A.J.H. (2007) Analysis of Variance and Covariance: How to choose and Construct Models for the Life Sciences. Cambridge University Press.

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

-
- (4) Mead, R. (1990) The Design of Experiments. Statistical principles for practical application. Cambridge University Press.
 - (5) Montgomery D.C. (1997) Design and Analysis of Experiments. John Wiley & Sons Inc.
 - (6) Winer, B.J. (1971) Statistical Principles in Experimental Design. McGraw Hill.
 - (7) Wu, C.F.J & Hamada, M.S. (2009) Experiments: Planning, Analysis and Optimization. John Wiley & Sons Inc.
-

Appendix 3

CURRENT LIMITATIONS OF THE ASSAY

Due to the current status of knowledge, several limitations are associated with the *in vivo* comet assay. It is expected that these limitations will be reduced or more narrowly defined as there is more experience with application of the assay to answer safety issues in a regulatory context.

1. Some types of DNA damage may be short-lived, i.e. may be repaired too quickly to be observed 24 hours or more after the last dose. There is no identifiable list of the types of short-lived damages, nor of the chemicals which are likely to cause this type of damage, nor is it known over what time period this type of damage can be detected. The optimum sampling time(s) may also be chemical- or route-specific and sampling times should be determined from kinetic data (for example the time, T_{max} , at which the peak plasma or tissue concentration is achieved), when such data are available. Most of the validation studies supporting this test method specified necropsy 2 or 3 hours following administration of the final dose. Most studies in the published literature describe administration of the final dose between 2 and 6 hours prior to sacrifice. Therefore, these experiences were used as the basis for the recommendation in the test method that, in the absence of data indicating otherwise, the final dose should be administered at a specified time point between 2 and 6 hours prior to necropsy.
2. There are no identifiable study data that examine the sensitivity of the test for the detection of short-lived DNA damage following administration in food or drinking water compared to administration by gavage. DNA damage has been detected following administration in feed and drinking water, but there are relatively few such reports compared to the much greater experience with gavage and i.p. administration. Thus the sensitivity of the assay may be reduced for chemicals which induce short-lived damage administered through feed or drinking water.
3. No inter-laboratory studies have been conducted in tissues other than liver and stomach, therefore no recommendation has been established for how to achieve a sensitive and reproducible response in tissues other than liver, such as expected positive and negative control ranges. For the liver, agreement on setting a lower limit to the negative control value also could not be reached.
4. Although there are several publications demonstrating the confounding effect of cytotoxicity *in vitro*, very little data have been published *in vivo* and therefore no single measure of cytotoxicity could be recommended. Histopathological changes such as inflammation, cell infiltration, apoptotic or necrotic changes have been associated with increases in DNA migration however, as demonstrated by the JaCVAM validation trial (OECD, 2014), these changes do not always result in positive comet findings and consequently no definitive list of histopathological changes that are always associated with increased DNA migration is available. Hedgehogs (or clouds, ghost cells) have previously been suggested as an indicator of cytotoxicity, however, the etiology of the hedgehogs is uncertain. Data exist which suggest that they can be caused by chemical-related cytotoxicity, mechanical/enzyme-induced damage initiated during sample preparation (Guerard *et al.*, 2014) and/or a more extreme effect of test chemical genotoxicity. Other data seem to show they are due to extensive, but perhaps repairable DNA damage (Lorenzo *et al.*, 2013).
5. Tissues or cell nuclei have been successfully frozen for later analysis. This usually results in a measurable effect on the response to the vehicle and positive control (Recio *et al.*, 2010; Recio *et al.*, 2012; Jackson *et al.*, 2013). If used, the laboratory should demonstrate competency in freezing methodologies and confirm acceptable low ranges of % tail DNA in target tissues of vehicle treated animals, and that positive responses can still be detected. In the literature, the freezing of tissues has been described using different methods. However, currently there is no agreement on how to best freeze and thaw tissues, and how to assess whether a potentially altered response may affect the sensitivity of the test.
6. Recent work demonstrates that the list of critical variables is expected to continue to become shorter and the parameters for critical variables more precisely defined (Guerard *et al.*, 2014).

References

- (1) Guerard, M., C. Marchand, U. Plappert-Helbig (2014), Influence of Experimental Conditions on Data Variability in the Liver Comet Assay, *Environmental and Molecular Mutagenesis*, Vol. 55/2, pp. 114-21.
- (2) Jackson, P. *et al.* (2013), Validation of use of frozen tissues in high-throughput comet assay with fully-automatic scoring, *Mutagenesis*, Vol. 28/6, pp. 699-707.
- (3) Lorenzo, Y. *et al.* (2013), The comet assay, DNA damage, DNA repair and cytotoxicity: hedgehogs are not always dead, *Mutagenesis*, Vol. 28/4, pp. 427-32.
- (4) OECD (2014), *Reports of the JaCVAM initiative international pre-validation and validation studies of the in vivo rodent alkaline comet assay for the detection of genotoxic carcinogens*, Series on Testing and Assessment, Nos. 195 and 196, OECD Publishing, Paris.
- (5) Recio L, Hobbs C, Caspary W, Witt KL, (2010), Dose-response assessment of four genotoxic chemicals in a combined mouse and rat micronucleus (MN) and Comet assay protocol, *J. Toxicol. Sci.* 35:149-62.
- (6) Recio, L. *et al.* (2012), Comparison of Comet assay dose-response for ethyl methanesulfonate using freshly prepared versus cryopreserved tissues, *Environmental and Molecular Mutagenesis*, Vol. 53/2, pp. 101-13.

(16) In Part C, Chapter C.13 is replaced by the following:

'C.13 Bioaccumulation in Fish: Aqueous and Dietary Exposure

INTRODUCTION

This test method (TM) is equivalent to OECD test guideline (TG) 305 (2012). The major goal of this revision of test method is two-fold. Firstly, it is intended to incorporate a dietary bioaccumulation ⁽¹⁾ test suitable for determining the bioaccumulation potential of substances with very low water solubility. Secondly, it is intended to create a test method that, when appropriate, utilises fewer fish for animal welfare reasons, and that is more cost-effective.

In the years since adoption of the consolidated test method C.13 (1), numerous substances have been tested, and considerable experience has been gained both by laboratories and by regulatory authorities. This has led to the conviction that the complexity of the test can be reduced if specific criteria are met (*cf.* paragraph 88), and that a tiered approach is possible. Experience has also shown that biological factors such as growth and fish lipid content can have a strong impact on the results and may need to be taken into account. In addition, it has been recognised that testing very poorly water soluble substances may not be technically feasible. In addition, for substances with very low water solubility in the aquatic environment, exposure via water may be of limited importance in comparison to the dietary route. This has led to the development of a test method in which fish are exposed via their diet (*cf.* paragraph 7-14 and 97 onwards). Validation (ring test) of the dietary exposure test was conducted in 2010 (51).

The main changes include:

- The testing of only one test concentration can be considered sufficient, when it is likely that the bioconcentration factor (BCF) is independent of the test concentration.
- A minimised aqueous exposure test design in which a reduced number of sample points is possible, if specific criteria are met.

⁽¹⁾ See Appendix 1 for definitions and units

- Fish lipid content should be measured so that BCF can be expressed on a 5 % lipid content basis.
- Greater emphasis on kinetic BCF estimation (when possible) next to estimating the BCF at steady state.
- For certain groups of substances, a dietary exposure test will be proposed, where this is considered more suitable than an aqueous exposure test.
- Fish weight should be measured so that BCF_k can be corrected for growth dilution.

Before carrying out any of the bioaccumulation tests, the following information about the test substance should be known:

- (a) Sensitivity of the analytical technique for measuring tissue and aqueous or food concentrations of both the test substance and possible metabolites (*cf.* paragraph 65).
- (b) Solubility in water [TM A.6; (2)]; this should be determined in accordance with a method that is appropriate for the (estimated) range of the solubility to obtain a reliable value. For hydrophobic substances, this will generally be the column elution method.
- (c) n-Octanol-water partition coefficient, K_{ow} ⁽¹⁾ [TMs A.8 (4), A.24 (5), A.23 (6)]; or other suitable information on partitioning behaviour (e.g. sorption to lipids, K_{oc}); this should be determined in accordance with a method that is appropriate for the (estimated) range of the K_{ow} to obtain a reliable value. For hydrophobic substances, this will generally be the slow-stirring method [TM A.23 (6)];
- (d) Substance stability in water (hydrolysis [TM C.7 (7)]);
- (e) Substance stability in food (specifically when a dietary exposure test approach is chosen);
- (f) Information on phototransformation relevant for the irradiation conditions in the test (8);
- (g) Surface tension (*i.e.* for substances where the $\log K_{ow}$ cannot be determined) [TM A.5 (9)];
- (h) Vapour pressure [TM A.4 (10)];
- (i) Any information on biotic or abiotic degradation in water, such as (but not restricted to) ready biodegradability [TMs C.4 parts II to VII (11), C.29 (12)], where appropriate;
- (j) Information on metabolites: structure, $\log K_{ow}$, formation and degradability, where appropriate;
- (k) Acid dissociation constant (pK_a) for substances that might ionise. If necessary, the pH of the test water should be adjusted to ensure that the substance is in the unionised form in the test if compatible with fish species.

Independent of the chosen exposure method or sampling scheme, this test method describes a procedure for characterising the bioaccumulation potential of substances in fish. Although flow-through test regimes are much to be preferred, semi-static regimes are permissible, provided that the validity criteria (*cf.* paragraphs 24 and 113) are satisfied. In the dietary exposure route, the flow-through system is not necessary to maintain aqueous concentrations of the tested substance, but will help maintain adequate dissolved oxygen concentrations and help ensure clean water and remove influences of e.g. excretion products.

⁽¹⁾ Sometimes denoted by P_{ow} ; determined by a shake-flask method in TM A.8 (4), an HPLC method in TM A.24 (5) and a slow-stirring method in TM A.23 (6). The generator-column technique is occasionally used for the determination of $\log K_{ow}$. A limited number of studies are available that makes use of this technique, primarily for chlorinated biphenyls and dibenzodioxins (e.g. Li and Doucette, 1993) (3). For substances that might ionise, $\log K_{ow}$ should refer to the unionised form.

Independent of the chosen test method, sufficient details are given in this test method for performing the test while allowing adequate freedom for adapting the experimental design to the conditions in particular laboratories and for varying characteristics of test substances. The aqueous exposure test is most appropriately applied to stable organic substances with $\log K_{ow}$ values between 1,5 and 6,0 (13) but may still be applied to strongly hydrophobic substances (having $\log K_{ow} > 6,0$), if a stable and fully dissolved concentration of the test substance in water can be demonstrated. If a stable concentration of the test substance in water cannot be demonstrated, an aqueous study would not be appropriate thus the dietary approach for testing the substance in fish would be required (although interpretation and use of the results of the dietary test may depend on the regulatory framework). Pre-estimates of the bioconcentration factor (BCF, sometimes denoted as K_b) for organic substances with $\log K_{ow}$ values up to about 9,0 can be obtained using the equation of Bintein *et al.* (14). The pre-estimate of the bioconcentration factor for such strongly hydrophobic substances may be higher than the steady-state bioconcentration factor (BCF_{ss}) value expected to be obtained from laboratory experiments, especially when a simple linear model is used for the pre-estimate. Parameters which characterise the bioaccumulation potential include the uptake rate constant (k_1), loss rate constants including the depuration rate constant (k_2), the steady-state bioconcentration factor (BCF_{ss}), the kinetic bioconcentration factor (BCF_k) and the dietary biomagnification factor (BMF) ⁽¹⁾.

Radiolabelled test substances can facilitate the analysis of water, food and fish samples, and may be used to determine whether identification and quantification of metabolites will be necessary. If total radioactive residues are measured alone (e.g. by combustion or tissue solubilisation), the BCF or BMF is based on the total of the parent substance, any retained metabolites and also assimilated carbon. BCF or BMF values based on total radioactive residues may not, therefore, be directly comparable to a BCF or BMF derived by specific chemical analysis of the parent substance only. Separation procedures, such as TLC, HPLC or GC ⁽²⁾ may be employed before analysis in radiolabelled studies in order to determine BCF or BMF based on the parent substance. When separation techniques are applied, identification and quantification of parent substance and relevant metabolites should be performed ⁽³⁾ (cf. paragraph 65) if BCF or BMF is to be based upon the concentration of the parent substance in fish and not upon total radiolabelled residues. It is also possible to combine a fish metabolism or *in vivo* distribution study with a bioaccumulation study by analysis and identification of the residues in tissues. The possibility of metabolism can be predicted by suitable tools (e.g. OECD QSAR toolbox (15) and proprietary QSAR programs).

The decision on whether to conduct an aqueous or dietary exposure test, and in what set-up, should be based on the factors in paragraph 3 considered together with the relevant regulatory framework. For example, for substances, which have a high $\log K_{ow}$ but still show appreciable water solubility with respect to the sensitivity of available analytical techniques, an aqueous exposure test should be considered in the first instance. However it is possible that information on water solubility is not definitive for these hydrophobic types of substances, so the possibility of preparing stable, measurable dissolved aqueous concentrations (stable emulsions are not allowed) applicable for an aqueous exposure study should be investigated before a decision is made on which test method to use (16). It is not possible to give exact prescriptive guidance on the method to be used based on water solubility and octanol-water partition coefficient “cut off” criteria, as other factors (analytical techniques, degradation, adsorption, etc.) can have a marked influence on method applicability for the reasons given above. However, a $\log K_{ow}$ above 5 and a water solubility below $\sim 0,01 - 0,1$ mg/l mark the range of substances where testing via aqueous exposure may become increasingly difficult.

Other factors that may influence test choice should be considered, including the substance's potential for adsorption to test vessels and apparatus, its stability in aqueous solution versus its stability in fish food (17) (18), etc.

⁽¹⁾ See Appendix 1 for definitions and units

⁽²⁾ TLC: thin layer chromatography; HPLC: high pressure liquid chromatography; GC: gas chromatography

⁽³⁾ In some regulatory frameworks analysis of metabolites may be obligatory when certain conditions are met (cf. paragraph 65).

Information on such practical aspects may be available from other completed aqueous studies. Further information on the evaluation of aspects relating to the performance of bioaccumulation studies is available in the literature (e.g. (19)).

For substances where the solubility or the maintenance of the aqueous concentration as well as the analysis of these concentrations do not pose any constraints to the realization of an aqueous exposure method, this method is preferred to determine the bioconcentration potential of the substance. In any case, it should be verified that the aqueous exposure concentration(s) to be applied are within the aqueous solubility in the test media. Different methods for maintaining stable concentrations of the dissolved test substance can be used, such as the use of stock solutions or passive dosing systems (e.g. column elution method), as long as it can be demonstrated that stable concentrations can be maintained and the test media are not altered from that recommended in paragraph 27.

For strongly hydrophobic substances ($\log K_{ow} > 5$ and a solubility below $\sim 0,01\text{-}0,1$ mg/l), testing via aqueous exposure may become increasingly difficult. Reasons for constraints may be that the aqueous concentration cannot be maintained at a level that is considered to be sufficiently constant (e.g. due to sorption to the glass of exposure containers or rapid uptake by the fish) or that the aqueous concentrations to be applied are so low that they are in the same range as or below the analytical limit of quantification ⁽¹⁾. For these highly hydrophobic substances the dietary test is recommended, provided that the test is consistent with the relevant regulatory framework and risk assessment needs.

For surfactants it should be considered whether the aqueous bioconcentration test is feasible, given the substance properties, otherwise the dietary study is probably more appropriate. Surfactants are surface acting agents, which lower the interfacial tension between two liquids. Their amphiphilic nature (i.e. they contain both a hydrophilic and a hydrophobic part) causes them to accumulate at interfaces such as the water-air interface, the water-food interface, and glass walls, which hampers the determination of their aqueous concentration.

The dietary test can circumvent some of the exposure aspects for complex mixtures with components of differing water solubility limits, in that comparable exposure to all components of the mixture is more likely than in the aqueous method (cf. (20)).

It should be noted that the dietary approach yields a dietary biomagnification factor (BMF) rather than a bioconcentration factor (BCF) ⁽²⁾. Approaches are available to estimate a kinetic bioconcentration factor (BCF_k) from data generated in the dietary study (as discussed in Appendix 8, but these approaches should be used with caution. In general, these approaches assume first order kinetics, and are only applicable to certain groups of compounds. It is unlikely that such approaches can be applied for surfactants (see paragraph 12).

A minimised aqueous exposure test set-up with fewer sampling points to reduce the number of animals and/or resources (cf. paragraph 83 onwards) should only be applied to those substances where there is reason to expect that uptake and depuration will follow approximately first order kinetics (i.e. in general non-ionized organic substances, cf. paragraph 88).

⁽¹⁾ In general, measured concentrations in water during the uptake phase should be at least an order of magnitude above the limit of quantification so that more than one half-life of body burden can be measured in the depuration phase of the study.

⁽²⁾ See Appendix 1 for definitions and units

C.13 - I: Aqueous Exposure Bioconcentration Fish Test

PRINCIPLE OF THE TEST

The test consists of two phases: the exposure (uptake) and post-exposure (depuration) phases. During the uptake phase, a group of fish of one species is exposed to the test substance at one or more chosen concentrations, depending on the properties of the test substance (cf. paragraph 49). They are then transferred to a medium free of the test substance for the depuration phase. A depuration phase is always necessary unless uptake of the substance during the uptake phase has been insignificant. The concentration of the test substance in/on the fish (or specified tissue thereof) is followed through both phases of the test. In addition to the exposed group, a control group of fish is held under identical conditions except for the absence of the test substance, to relate possible adverse effects observed in the bioconcentration test to a matching control group and to obtain background concentrations of test substance ⁽¹⁾.

In the aqueous exposure test, the uptake phase is usually run for 28 days. The duration can be lengthened if necessary (cf. paragraph 18), or shortened if it is demonstrated that steady-state has been reached earlier (see Appendix 1, definitions and units). A prediction of the length of the uptake phase and the time to steady-state can be made from equations in Appendix 5. The depuration period is then begun when the fish are no longer exposed to the test substance, by transferring the fish to the same medium but without the test substance in a clean vessel. Where possible the bioconcentration factor is calculated preferably both as the ratio of concentration in the fish (C_f) and in the water (C_w) at steady-state (BCF_{ss} ; see Appendix 1, definition) and as a kinetic bioconcentration factor (BCF_k ; see Appendix 1, definitions and units), which is estimated as the ratio of the rate constants of uptake (k_1) and depuration (k_2) assuming first order kinetics ⁽²⁾.

If a steady-state is not achieved within 28 days, either the BCF is calculated using the kinetic approach (cf. paragraph 38) or the uptake phase can be extended. Should this lead to an impractically long uptake phase to reach steady-state (cf. paragraphs 37 and 38, Appendix 5), the kinetic approach is preferred. Alternatively, for highly hydrophobic substances the conduction of a dietary study should be considered ⁽³⁾, provided that the dietary test is consistent with the relevant regulatory framework.

The uptake rate constant, the depuration (loss) rate constant (or constants, where more complex models are involved), the bioconcentration factor (steady-state and/or kinetic), and where possible, the confidence limits of each of these parameters are calculated from the model that best describes the measured concentrations of test substance in fish and water (cf. Appendix 5).

The increase in fish mass during the test will result in a decrease of test substance concentration in growing fish (so-called growth dilution), and thus the kinetic BCF will be underestimated if not corrected for growth (cf. paragraphs 72 and 73).

The BCF is based on the total concentration in the fish (i.e. per total wet weight of the fish). However, for special purposes, specified tissues or organs (e.g. muscle, liver), may be used if the fish are sufficiently large or the fish may be divided into edible (fillet) and non-edible (viscera) fractions. Since, for many organic substances, there is a clear relationship between the potential for bioconcentration and hydrophobicity, there is also a corresponding relationship between the lipid content of the test fish and the observed bioconcentration of such substances. Thus, to reduce this source of variability in test results for those substances with high lipophilicity (i.e. with $\log K_{ow} > 3$), bioconcentration should be expressed as normalised to a fish with a 5 % lipid content (based on whole body wet weight) in addition to that derived directly from the study. This is necessary to provide a basis from which results for different substances and/or test species can be compared against one another. The figure of 5 % lipid content has been widely used as this represents the average lipid content of fish commonly used in this test method (21).

⁽¹⁾ For most test substances, there should ideally be no detections in the control water. Background concentrations should only be relevant to naturally occurring materials (e.g., some metals) and substances that are ubiquitous in the environment.

⁽²⁾ If first order kinetics is obviously not obeyed, more complex models should be employed (see references in Appendix 5 and advice from a biostatistician sought).

⁽³⁾ Uptake may be limited by low exposure concentrations because of low water solubility in the bioconcentration test, whereas far higher exposure concentrations can be achieved with the dietary test.

INFORMATION ON THE TEST SUBSTANCE

In addition to the properties of the test substance given in the Introduction (paragraph 3), other information required is the toxicity to the fish species to be used in the test, preferably the asymptotic LC_{50} (i.e. time-independent) and/or toxicity estimated from long-term fish tests (e.g. TMs C.47 (22), C.15 (23), C.14 (24)).

An appropriate analytical method, of known accuracy, precision, and sensitivity, for the quantification of the substance in the test solutions and in biological material should be available, together with details of sample preparation and storage. The analytical quantification limit of the test substance in both water and fish tissues should also be known. When a radiolabelled test substance is used, it should be of the highest purity (e.g. preferably > 98 %) and the percentage of radioactivity associated with impurities should be known.

VALIDITY OF THE TEST

For a test to be valid the following conditions apply:

The water temperature variation is less than ± 2 °C, because large deviations can affect biological parameters relevant for uptake and depuration as well as cause stress to animals;

The concentration of dissolved oxygen does not fall below 60 % saturation;

The concentration of the test substance in the chambers is maintained within ± 20 % of the mean of the measured values during the uptake phase;

The concentration of the test substance is below its limit of solubility in water, taking into account the effect that the test water may have on effective solubility ⁽¹⁾;

The mortality or other adverse effects/disease in both control and treated fish is less than 10 % at the end of the test; where the test is extended over several weeks or months, death or other adverse effects in both sets of fish should be less than 5 % per month and not exceed 30 % in all. Significant differences in average growth between the test and the control groups of sampled fish could be an indication of a toxic effect of the test substance.

REFERENCE SUBSTANCES

The use of reference substances of known bioconcentration potential and low metabolism would be useful in checking the experimental procedure, when required (e.g. when a laboratory has no previous experience with the test or experimental conditions have been changed).

DESCRIPTION OF THE METHOD

Apparatus

Care should be taken to avoid the use of materials — for all parts of the equipment — that can dissolve, sorb or leach and have an adverse effect on the fish. Standard rectangular or cylindrical tanks, made of chemically inert material and of a suitable capacity in compliance with loading rate (cf. paragraph 43), can be used. The use of soft plastic tubing should be minimised. Polytetrafluoroethylene, stainless steel and/or glass tubing should be used. Experience has shown that for test substances with high adsorption coefficient, such as the synthetic pyrethroids, silanised glass may be required. In such situations the equipment should be discarded after use. It is preferable to expose test systems to concentrations of the test substance to be used in the study for as long as is required to demonstrate the maintenance of stable exposure concentrations prior to the introduction of test organisms.

⁽¹⁾ For multi-constituent substances, UVCBs and mixtures, the water solubility of each relevant component should be considered to determine the appropriate exposure concentrations.

Water

Natural water is generally used in the test and should be obtained from uncontaminated and uniform quality source. Yet, reconstituted water (*i.e.* demineralised water with specific nutrients added in known amounts) may be more suitable to guarantee uniform quality over time. The dilution water, which is the water that is mixed with the test substance before entering the test vessel (*cf.* paragraph 30), should be of a quality that will allow the survival of the chosen fish species for the duration of the acclimation and test periods without them showing any abnormal appearance or behaviour. Ideally, it should be demonstrated that the test species can survive, grow and reproduce in the dilution water (*e.g.* in laboratory culture or a life-cycle toxicity test). The dilution water should be characterised at least by pH, hardness, total solids, total organic carbon (TOC ⁽¹⁾) and, preferably also ammonium, nitrite and alkalinity and, for marine species, salinity. The parameters which are important for optimal fish well-being are not fully known, but Appendix 2 gives recommended maximum concentrations of a number of parameters for fresh and marine test waters.

The dilution water should be of constant quality during the period of a test. The pH value should be within the range 6,0 to 8,5 at test start, but during a given test it should be within a range of $\pm 0,5$ pH units. In order to ensure that the dilution water will not unduly influence the test result (for example, by complexation of the test substance) or adversely affect the performance of the stock of fish, samples should be taken at intervals for analysis, at least at the beginning and end of the test. Determination of heavy metals (*e.g.* Cu, Pb, Zn, Hg, Cd, and Ni), major anions and cations (*e.g.* Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , and SO_4^{2-}), pesticides (*e.g.* total organophosphorous and total organochlorine pesticides), total organic carbon and suspended solids should be conducted, for example, every three months where dilution water is known to be relatively constant in quality. If dilution water quality has been demonstrated to be constant over at least one year, determinations can be less frequent and intervals extended (*e.g.* every six months).

The natural particle content as well as the total organic carbon of the dilution water should be as low as possible to avoid adsorption of the test substance to organic matter, which may reduce its bioavailability and therewith result in an underestimation of the BCF. The maximum acceptable value is 5 mg/l for particulate matter (dry matter, not passing a 0,45 μm filter) and 2 mg/l for total organic carbon (*cf.* Appendix 2). If necessary, the dilution water should be filtered before use. The contribution to the organic carbon content in test water from the test fish (excreta) and from the food residues should be kept as low as possible (*cf.* paragraph 46).

Test Solutions

Prepare a stock solution of the test substance at a suitable concentration. The stock solution should preferably be prepared by simply mixing or agitating the test substance in the dilution water. An alternative that may be appropriate in some cases is the use of a solid phase desorption dosing system. The use of solvents and dispersants (solubilising agents) is not generally recommended (*cf.* (25)); however, the use of these materials may be acceptable in order to produce a suitably concentrated stock solution, but every effort should be made to minimise the use of such materials and their critical micelle concentration should not be exceeded (if relevant). Solvents which may be used are acetone, ethanol, methanol, dimethyl formamide and triethylene glycol; dispersants that have been used are Tween 80, methylcellulose 0,01 % and HCO-40. The solvent concentration in the final test medium should be the same in all treatments (*i.e.* regardless of test substance concentration) and should not exceed the corresponding toxicity thresholds determined for the solvent under the test conditions. The maximum level is a concentration of 100 mg/l (or 0,1 ml/l). It is unlikely that a solvent concentration of 100 mg/l will significantly alter the maximum dissolved concentration of the test substance which can be achieved in the medium (25). The solvent's contribution (together with the test substance) to the overall content of organic carbon in the test water should be known. Throughout the test, the concentration of total organic carbon in the test vessels should not exceed the concentration of organic carbon originating from the test substance, and solvent or solubilising agent ⁽²⁾, if used, by more

⁽¹⁾ TOC includes organic carbon from particles and dissolved organic carbon, *i.e.* $\text{TOC} = \text{POC} + \text{DOC}$.

⁽²⁾ Although not generally recommended, if a solvent or solubilising agent is used the organic carbon originating from this agent should be added to the organic carbon from the test substance to evaluate the concentration of organic carbon in the test vessels.

than 10 mg/l ($\pm 20\%$). Organic matter content can have a significant effect on the amount of freely dissolved test substance during flow-through fish tests, especially for highly lipophilic substances. Solid-phase microextraction (*cf.* paragraph 60) can provide important information on the ratio between bound and freely dissolved compounds, of which the latter is assumed to represent the bioavailable fraction. The test substance concentration should be below the solubility limit of the test substance in the test media in spite of the use of a solvent or solubilising agent. Care should be taken when using readily biodegradable solvents as these can cause problems with bacterial growth in flow-through tests. If it is not possible to prepare a stock solution without the use of a solubilising agent, consideration should be given to the appropriateness of an aqueous exposure study as opposed to a dietary exposure study.

For flow-through tests, a system which continuously dispenses and dilutes a stock solution of the test substance (e.g. metering pump, proportional diluter, saturator system) or a solid phase desorption dosing system is required to deliver the test concentrations to the test chambers. Preferably allow at least five volume replacements through each test chamber per day. The flow-through mode is to be preferred, but where this is not possible (e.g. when the test organisms are adversely affected) a semi-static technique may be used provided that the validity criteria are satisfied (*cf.* paragraph 24). The flow rates of stock solutions and dilution water should be checked both 48 hours before and then at least daily during the test. Include in this check the determination of the flow-rate through each test chamber and ensure that it does not vary by more than 20 % either within or between chambers.

Selection of species

Important criteria in the selection of species are that they are readily available, can be obtained in convenient sizes and can be satisfactorily maintained in the laboratory. Other criteria for selecting fish species include recreational, commercial, ecological importance as well as comparable sensitivity, past successful use etc. Recommended test species are given in Appendix 3. Other species may be used but the test procedure may have to be adapted to provide suitable test conditions. The rationale for the selection of the species and the experimental method should be reported in this case. In general, the use of smaller fish species will shorten the time to steady-state, but more fish (samples) may be needed to adequately analyse lipid content and test substance concentrations in the fish. In addition it is possible that differences in respiration rate and metabolism between young and older fish may hamper comparisons of results between different tests and test species. It should be noted that fish species tested during a (juvenile) life-stage with rapid growth can complicate data interpretation.

Holding of fish (relevant for aqueous and dietary exposure)

The stock population of fish should be acclimated for at least two weeks in water (*cf.* paragraph 28) at the test temperature and feed throughout on a sufficient diet (*cf.* paragraph 45). Both water and diet should be of the same type as those to be used during the test.

Following a 48-hour settling-in period, mortalities are recorded and the following criteria applied:

- Mortalities exceeding 10 % of the population in seven days: reject the entire batch;
- Mortalities of between 5 and 10 % of the population in seven days: acclimate for seven additional days — if more than 5 % mortality during the second seven days, reject the entire batch;
- Mortalities below 5 % of the population in seven days: accept the batch.

Fish used in tests should be free from observable diseases and abnormalities. Any diseased fish should be discarded. Fish should not receive treatment for disease in the two weeks preceding the test, or during the test.

PERFORMANCE OF THE TEST

Preliminary test

It may be useful to conduct a preliminary experiment in order to optimise the test conditions of the definitive test, e.g. selection of test substance concentration(s), duration of the uptake and depuration phases, or to determine whether a full test need be conducted. The design of the preliminary test should be such as to obtain the information required. It can be considered if a minimised test may be sufficient to derive a BCF, or if a full study is needed (*cf.* paragraphs 83-95 on the minimised test).

Conditions of Exposure*Duration of uptake phase*

A prediction of the duration of the uptake phase can be obtained from practical experience (e.g. from a previous study or an accumulation study on a structurally related substance) or from certain empirical relationships utilising knowledge of either the aqueous solubility or the octanol/water partition coefficient of the test substance (provided that uptake follows first order kinetics, *cf.* Appendix 5).

The uptake phase should be run for 28 days unless it can be demonstrated that steady-state has been reached earlier (see Appendix 1, definitions and units). A steady-state is reached in the plot of test substance in fish (C_f) against time when the curve becomes parallel to the time axis and three successive analyses of C_f made on samples taken at intervals of at least two days are within $\pm 20\%$ of each other, and there is no significant increase of C_f in time between the first and last successive analysis. When pooled samples are analysed, at least four successive analyses are required. For test substances which are taken up slowly the intervals would more appropriately be seven days. If steady-state has not been reached by 28 days, either the BCF is calculated using only the kinetic approach, which is not reliant on steady-state being reached, or the uptake phase can be extended, taking further measurements, until steady-state is reached or for 60 days, whichever is shorter. Also, the test substance concentration in the fish at the end of the uptake phase needs to be sufficiently high to ensure a reliable estimation of k_2 from the depuration phase. If no significant uptake is shown after 28 days, the test can be stopped.

Duration of the depuration phase

For substances following first order kinetics, a period of half the duration of the uptake phase is usually sufficient for an appropriate (e.g. 95 %) reduction in the body burden of the substance to occur (*cf.* Appendix 5 for explanation of the estimation). If the time required to reach 95 % loss is impractically long, exceeding for example twice the normal duration of the uptake phase (*i.e.* more than 56 days) a shorter period may be used (e.g. until the concentration of test substance is less than 10 % of steady-state concentration). However, longer depuration periods may be necessary for substances having more complex patterns of uptake and depuration than are represented by a one-compartment fish model that yields first order kinetics. If such complex patterns are observed and/or anticipated, it is advised to seek advice from a biostatistician and/or pharmacokineticist to ensure a proper test set-up. As the depuration period is extended, numbers of fish to sample may become limiting and growth differences between fish can influence the results. The period will also be governed by the period over which the concentration of the test substance in the fish remains above the analytical limit of quantification.

Numbers of test fish

Select the numbers of fish per test concentration such that a minimum of four fish are available at each sampling point. Fish should only be pooled if analysis of single fish is not feasible. If higher precision in curve fitting (and derived parameters) is intended or if metabolism studies are required (e.g. to distinguish between metabolites and parent substance when using radiolabelled test substances), more fish per sampling point will be necessary. The lipid content should be determined on the same biological material as is used to determine the concentration of the test substance. Should this not be feasible, additional fish may be needed (*cf.* paragraphs 56 and 57).

If adult (*i.e.* sexually mature) fish are used, they should not be in a spawning state or recently spent (*i.e.* already spawned) either before or during the test. It should also be reported whether male or female, or both are used in the experiment. If both sexes are used, differences in growth and lipid content between sexes should be documented to be non-significant before the start of the exposure, in particular if it is anticipated that pooling of male and female fish will be necessary to ensure detectable substance concentrations and/or lipid content.

In any one test, select fish of similar weight such that the smallest are no smaller than two-thirds of the weight of the largest. All should be of the same year-class and come from the same source. Since weight and age of a fish may have a significant effect on BCF values (12) these details should be recorded accurately. It is recommended that a sub-sample of the stock of fish is weighed shortly before the start of the test in order to estimate the mean weight (*cf.* paragraph 61).

Loading

High water-to-fish ratios should be used in order to minimise the reduction in the concentration of the test compound in water caused by the addition of the fish at the start of the test and also to avoid decreases in dissolved oxygen concentration. It is important that the loading rate is appropriate for the test species used. In any case, a fish-to-water loading rate of 0,1-1,0 g of fish (wet weight) per litre of water per day is normally recommended. Higher fish-to-water loading rates can be used if it is shown that the required concentration of test substance can be maintained within ± 20 % limits, and that the concentration of dissolved oxygen does not fall below 60 % saturation (*cf.* paragraph 24).

In choosing appropriate loading regimes, take into account the normal habitat of the fish species. For example, bottom-living fish may demand a larger bottom area of the aquarium for the same volume of water compared to pelagic fish species.

Feeding

During the acclimation and test periods, feed an appropriate diet of known lipid and total protein content to the fish in an amount sufficient to keep them in a healthy condition and to maintain body weight (some growth is allowed). Feed daily throughout the acclimation and test periods at a set level depending on the species used, experimental conditions and calorific value of the food (for example for rainbow trout between approximately 1 to 2 % of body weight per day). The feeding rate should be selected such that fast growth and large increase of lipid content are avoided. To maintain the same feeding rate, the amount of feed should be re-calculated as appropriate, for example once per week. For this calculation, the weight of the fish in each test chamber can be estimated from the weight of the fish sampled most recently in that chamber. Do not weigh the fish remaining in the chamber.

Uneaten food and faeces should be siphoned daily from the test chambers shortly after feeding (30 minutes to one hour). The chambers should be kept as clean as possible throughout the test to keep the concentration of organic matter as low as possible (*cf.* paragraph 29), since the presence of organic carbon may limit the bioavailability of the test substance (12).

Since many feeds are derived from fishmeal, it should be ensured that the feed will not influence the test results or induce adverse effects, *e.g.* by containing (traces of) pesticides, heavy metals and/or the test substance itself.

Light and temperature

A 12- to 16-hour photoperiod is recommended and the temperature (± 2 °C) should be appropriate for the test species (*cf.* Appendix 3). The type and characteristics of illumination should be known. Caution should be given to the possible phototransformation of the test substance under the irradiation conditions of the study. Appropriate illumination should be used avoiding exposure of fish to unnatural photoproducts. In some cases it may be appropriate to use a filter to screen out UV irradiation below 290 nm.

Test concentrations

The test was originally designed for non-polar organic substances. For this type of substance, the exposure of fish to a single concentration is expected to be sufficient, as no concentration effects are expected, although two concentrations may be required for the relevant regulatory framework. If substances outside this domain are tested, or other indications of possible concentration dependence are known, the test should be run with two or more concentrations. If only one concentration is tested, justification for the use of one concentration should be given (*cf.* paragraph 79). Also, the tested concentration should be as low as is practical or technically possible (*i.e.* not close to the solubility limit).

In some cases it can be anticipated that the bioconcentration of a substance is dependent on the water concentration (e.g. for metals, where the uptake in fish may be at least partly regulated). In such a case it is necessary that at least two, but preferably more, concentrations are tested (*cf.* paragraph 49) which are environmentally relevant. Also for substances where the concentrations tested have to be near the solubility limit for practical reasons, testing at least two concentrations is recommended, because this can give insight into the reliability of the exposure concentrations. The choice of the test concentrations should incorporate the environmentally realistic concentration as well as the concentration that is relevant to the purpose of the specific assessment.

The concentration(s) of the test substance should be selected to be below its chronic effect level or 1 % of its acute asymptotic LC_{50} , within an environmentally relevant range and at least an order of magnitude above its limit of quantification in water by the analytical method used. The highest permissible test concentration can also be determined by dividing the acute 96 h LC_{50} by an appropriate acute/ chronic ratio (e.g. appropriate ratios for some substances are about three, but a few are above 100). If a second concentration is used, it should differ from the one above by a factor of ten. If this is not possible because of the toxicity criterion (that limits the upper test concentration) and the analytical limit (that limits the lower test concentration), a lower factor than ten can be used and use of radiolabelled test substance (of the highest purity, e.g. preferably > 98 %) should be considered. Care should be taken that no concentration used is above the solubility limit of the test substance in the test media.

Controls

One dilution water control or if relevant (*cf.* paragraphs 30 and 31), one control containing the solvent should be run in addition to the test series.

Frequency of Water Quality Measurements

During the test, dissolved oxygen, TOC, pH and temperature should be measured in all test and control vessels. Total hardness and salinity (if relevant) should be measured in the control(s) and one vessel. If two or more concentrations are tested, measure these parameters at the higher (or highest) concentration. As a minimum, dissolved oxygen and salinity (if relevant) should be measured three times — at the beginning, around the middle and end of the uptake period — and once a week in the depuration period. TOC should be measured at the beginning of the test (24 h and 48 h prior to test initiation of uptake phase) before addition of the fish and at least once a week during both uptake and depuration phases. Temperature should be measured and recorded daily, pH at the beginning and end of each period and hardness once each test. Temperature should preferably be monitored continuously in at least one vessel.

Sampling and Analysis of Fish and Water

Fish and water sampling schedule

Water should be sampled from the test chambers for the determination of test substance concentration before addition of the fish and during both uptake and depuration phases. The water should be sampled the before feeding, at the same time as the fish sampling. More frequent sampling may be useful to ensure stable concentrations after introduction of the fish. During the uptake phase, the concentrations of test substance should be determined in order to check compliance with the validity criteria (paragraph 24). If water sample analyses at the beginning of the depuration phase show that the test substance is not detected, this can be used as a justification not to measure test and control water for the test substance for the remainder of the depuration phase.

Fish should be sampled on at least five occasions during the uptake phase and on at least four occasions during the depuration phase for test substance. Since on some occasions it will be difficult to calculate a reasonably precise estimate of the BCF value based on this number of samples (especially when other than simple first order uptake and depuration kinetics are indicated), it may be advisable to take samples at a higher frequency in both periods (*cf.* Appendix 4).

The lipid content should be determined on the same biological material as is used to determine the concentration of the test substance at least at the start and end of the uptake phase and at the end of the depuration phase. Should this not be feasible, at least three independent fish should be sampled to determine lipid content at each of the same three time-points. The number of fish per tank at the start of the experiment should be adjusted accordingly ⁽¹⁾. Alternatively, if no significant amounts of the test substance are detected in control fish (*i.e.* fish from the stock population), the control fish from the test can be analysed for lipid content only and test substance analysis in the test group(s) (and the related uptake rate constant, depuration rate constant and BCF values) can be corrected for changes according to control group lipid content during the test ⁽²⁾.

Dead or diseased fish should not be analysed for test substance or lipid concentration.

An example of an acceptable sampling schedule is given in Appendix 4. Other schedules can readily be calculated using other assumed values of K_{ow} to calculate the exposure time for 95 % uptake (refer to Appendix 5 for calculations).

Sampling should be continued during the uptake phase until a steady-state has been established (see Appendix 1, definitions and units) or the uptake phase is otherwise terminated (after 28 or 60 days, *cf.* paragraphs 37 and 38). Before beginning the depuration phase, the fish should be transferred to clean vessels.

Sampling and sample preparation

Water samples should be obtained for analysis e.g. by siphoning through inert tubing from a central point in the test chamber. Neither filtration nor centrifuging appears always to separate the non-bioavailable fraction of the test substance from that which is bioavailable. If a separation technique is applied, a justification for, or validation of, the separation technique should always be provided in the test report given the bioavailability difficulties (25). Especially for highly hydrophobic substances (*i.e.* those substances with a $\log K_{ow} > 5$) (12) (26), where adsorption to the filter matrix or centrifugation containers could occur, samples should not be subjected to those treatments. Instead, measures should be taken to keep the tanks as clean as possible (*cf.* paragraph 46) and the content of total organic carbon should be monitored during both the uptake and depuration phases (*cf.* paragraph 53). To avoid possible issues with reduced bioavailability, sampling by solid phase microextraction techniques may be used for poorly soluble and highly hydrophobic substances.

⁽¹⁾ If the lipid content is not analysed in the same fish as the test substance, fish should at least be of the similar weight, and (if relevant) the same sex.

⁽²⁾ This alternative is only valid if the fish in all test groups are held in similar group sizes, fish are removed according to the same pattern and fed in the same way. This ensures that fish growth in all test groups is similar, if the tested concentration is below the toxic range. If the growth is similar, also the lipid content is expected to be similar. A different growth in the control would indicate a substance effect and invalidate the study.

The sampled fish should be euthanised instantly, using the most appropriate and humane method (for whole fish measurements, no further processes than rinsing with water (cf. paragraph 28) and blot drying the fish should be done). Weigh and measure total length ⁽¹⁾. In each individual fish, the measured weight and length should be linked to the analysed substance concentration (and lipid content, if applicable), for example using a unique identifier code for each sampled fish.

It is preferable to analyse fish and water immediately after sampling in order to prevent degradation or other losses and to calculate approximate uptake and depuration rate constants as the test proceeds. Immediate analysis also avoids delay in determining when a plateau (steady-state) has been reached.

Failing immediate analysis, the samples should be stored by an appropriate method. Before the beginning of the study, information should be obtained on the proper method of storage for the particular test substance — for example, deep-freezing, holding at 4 °C, extraction, etc. The duration of storage should be selected to ensure that the substance has not degraded while in storage.

Quality of analytical method

Since the whole procedure is governed essentially by the accuracy, precision and sensitivity of the analytical method used for the test substance, check experimentally that the accuracy, precision and reproducibility of the substance analysis, as well as recovery of the test substance from both water and fish are satisfactory for the particular method. This should be part of preliminary tests. Also, check that the test substance is not detectable in the dilution water used. If necessary, correct the values of test substance concentration in water and fish obtained from the test for the recoveries and background values of controls. The fish and water samples should be handled throughout in such a manner as to minimise contamination and loss (e.g. resulting from adsorption by the sampling device).

Analysis of fish samples

If radiolabelled materials are used in the test, it is possible to analyse for total radiolabel (*i.e.* parent and metabolites) or the samples may be cleaned up so that the parent substance can be analysed separately. If the BCF is to be based on the parent substance, the major metabolites should be characterised, as a minimum at the end of the uptake phase (cf. paragraph 6). Major metabolites are those representing ≥ 10 % of total residues in fish tissues, those representing ≥ 5 % at two consecutive sampling points, those showing increasing levels throughout the uptake phase, and those of known toxicological concern. If the BCF for the whole fish in terms of total radiolabelled residues is ≥ 500 , it may be advisable — and for certain categories of substances such as pesticides strongly recommended — identifying and quantifying major metabolites. Quantification of such metabolites may be required by some regulatory authorities. If degradates representing ≥ 10 % of total radiolabelled residues in the fish tissue are identified and quantified, then it is also recommended to identify and quantify degradates in the test water. Should this not be feasible, this should be explained in the report.

The concentration of the test substance should usually be determined for each weighed individual fish. If this is not possible, pooling of the samples on each sampling occasion may be done but pooling does restrict the statistical procedures which can be applied to the data, so an adequate number of fish to accommodate the desired pooling, statistical procedure and power should be included in the test. References (27) and (28) may be used as an introduction to relevant pooling procedures.

⁽¹⁾ In addition to weight, total length should be recorded because comparison of the extent of length increase during the test is a good indicator of whether an adverse effect has occurred.

BCF should be expressed as normalised to a fish with a 5 % lipid content (based on wet weight) in addition to that derived directly from the study (*cf.* paragraph 21), unless it can be argued that the test substance does not accumulate primarily in lipids. The lipid content of the fish should be determined on each sampling occasion if possible, preferably on the same extract as that produced for analysis for the test substance, since the lipids often have to be removed from the extract before it can be analysed chromatographically. However, analysis of test substances often requires specific extraction procedures which might be in contradiction to the test methods for lipid determination. In this case (until suitable non-destructive instrumental methods are available), it is recommended to employ a different strategy to determine the fish lipid content (*cf.* paragraph 56). Suitable methods should be used for determination of lipid content (20). The chloroform/methanol extraction technique (29) may be recommended as standard method (30), but the Smedes-method (31) is recommended as an alternative technique. This latter method is characterised by a comparable efficiency of extraction, high accuracy, the use of less toxic organic solvents and ease of performance. Other methods for which accuracy compares favourably to the recommended methods could be used if properly justified. It is important to give details of the method used.

Fish growth measurement

At the start of the test, five to ten fish from the stock population need to be weighed individually and their total length measured. These can be the same fish used for lipid analysis (*cf.* paragraph 56). The weight and length of fish used for each sampling event from both test and control groups should be measured before chemical or lipid analysis is conducted. The measurements of these sampled fish can be used to estimate the weight and length of fish remaining in the test and control tanks (*cf.* paragraph 45).

DATA AND REPORTING

Treatment of results

The uptake curve of the test substance should be obtained by plotting its concentration in/on fish (or specified tissues) in the uptake phase against time on arithmetic scales. If the curve has reached a plateau, that is, become approximately asymptotic to the time axis, the steady-state BCF (BCF_{ss}) should be calculated from:

$$\frac{C_f \text{ at steady-state (mean)}}{C_w \text{ at steady-state (mean)}}$$

The development of C_f may be influenced by fish growth (*cf.* paragraphs 72 and 73). The mean exposure concentration (C_w) is influenced by variation over time. It can be expected that a time-weighted average concentration is more relevant and precise for bioaccumulation studies, even if variation is within the appropriate validity range (*cf.* paragraph 24). A time weighted average (TWA) water concentration can be calculated according to Appendix 5, section 1.

The kinetic bioconcentration factor (BCF_k) should be determined as the ratio k_1/k_2 , the two first order kinetic rate constants. Rate constants k_1 and k_2 and BCF_k can be derived by simultaneously fitting both the uptake and the depuration phase. Alternatively, k_1 and k_2 can be determined sequentially (see Appendix 5 for a description and comparison of these methods). The depuration rate constant (k_2) may need correction for growth dilution (*cf.* paragraphs 72 and 73). If the uptake and/or depuration curve is obviously not first order, then more complex models should be employed (see references in Appendix 5 and advice from a biostatistician and/or pharmacokineticist sought).

Fish weight/length data

Individual fish wet weights and total lengths for all sampling intervals are tabulated separately for test and control groups during the uptake (including stock population for start of uptake) and depuration phases. In each individual fish the measured weight and length should be linked to the analysed chemical concentration, for example using a unique identifier code for each sampled fish. Weight is the preferred measure of growth for the purposes of correcting kinetic BCF values for growth dilution (see paragraph 73 and Appendix 5 for the method used to correct data for growth dilution).

Growth-Dilution Correction and Lipid Normalisation

Fish growth during the depuration phase can lower measured chemical concentrations in the fish with the effect that the overall depuration rate constant (k_2) is greater than would arise from removal processes (e.g. respiration, metabolism, egestion) alone. Kinetic bioconcentration factors should be corrected for growth dilution. A BCF_{ss} will also be influenced by growth, but no agreed procedure is available to correct a BCF_{ss} for growth. In cases of significant growth, the BCF_K , corrected for growth (BCF_{Kg}), should also be derived as it may be a more relevant measure of the bioconcentration factor. Lipid contents of test fish (which are strongly associated with the bioaccumulation of hydrophobic substances) can vary enough in practice such that normalisation to a set fish lipid content (5 % w/w) is necessary to present both kinetic and steady-state bioconcentration factors in a meaningful way, unless it can be argued that the test substance does not primarily accumulate in lipid (e.g. some perfluorinated substances may bind to proteins). Equations and examples for these calculations can be found in Appendix 5.

To correct a kinetic BCF for growth dilution, the depuration rate constant should be corrected for growth. This growth-corrected depuration rate constant (k_{2g}) is calculated by subtracting the growth rate constant (k_g , as obtained from the measured weight data) from the overall depuration rate constant (k_2). The growth-corrected kinetic bioconcentration factor is then calculated by dividing the uptake rate constant (k_1) by the growth-corrected depuration rate constant (k_{2g}) (cf. Appendix 5). In some cases this approach is compromised. For example, for very slowly depurating substances tested in fast growing fish, the derived k_{2g} may be very small and so the error in the two rate constants used to derive it becomes critical, and in some cases k_g estimates can be larger than k_2 . An alternative approach that circumvents the need for growth dilution correction involves using mass of test substance per fish (whole fish basis) depuration data rather than the usual mass of test substance per unit mass of fish (concentration) data. This can be easily achieved as tests according to this TM should link recorded tissue concentrations to individual fish weights. The simple procedure for doing this is outlined in Appendix 5. Note that k_2 should still be reported even if this alternative approach is used.

Kinetic and steady-state bioconcentration factors should also be reported relative to a default fish lipid content of 5 % (w/w), unless it can be argued that the test substance does not primarily accumulate in lipid. Fish concentration data, or the BCF, are normalised according to the ratio between 5 % and the actual (individual) mean lipid content (in % wet weight) (cf. Appendix 5).

If chemical and lipid analyses have been conducted on the same fish, then individual fish lipid normalised data should be used to calculate a lipid-normalised BCF. Alternatively, if the growth in control and exposed fish is similar, the lipid content of control fish alone may be used for lipid-correction (cf. paragraph 56). A method for calculating a lipid-normalised BCF is described in Appendix 5.

Interpretation of results

The results should be interpreted with caution where measured concentrations of test solutions occur at levels near the detection limit of the analytical method.

Average growth in both test and control groups should in principle not be significantly different to exclude toxic effects. The growth rate constants or the growth curves of the two groups should be compared by an appropriate procedure ⁽¹⁾.

⁽¹⁾ A t-test on growth rate constants can be performed, to test whether growth differs between control and test groups, or an F-test in case of analysis of variance. If needed, an F-test or likelihood ratio test can be used to assist in the choice of the appropriate growth model (OECD monograph 54, (32)).

Clearly defined uptake and depuration curves are an indication of good quality bioconcentration data. For the rate constants, the result of a χ^2 goodness-of-fit-test should show a good fit (*i.e.* small measurement error percentage (32)) for the bioaccumulation model, so that the rate constants can be considered reliable (*cf.* Appendix 5). If more than one test concentration is used, the variation in uptake/depuration constants between the test concentrations should be less than 20 % ⁽¹⁾. If not, concentration dependence could be indicated. Observed significant differences in uptake/ depuration rate constants between the applied test concentrations should be recorded and possible explanations given. Generally, the 95 % confidence limit of BCFs from well-designed studies approach ± 20 % of the derived BCF.

If two or more concentrations are tested, the results of both or all concentrations are used to examine whether the results are consistent and to show whether there is concentration dependence. If only one concentration is tested to reduce the use of animals and/or resources, justification of the use of one concentration should be given.

The resulting BCF_{ss} is doubtful if the BCF_k is significantly larger than the BCF_{ss} , as this can be an indication that steady-state has not been reached or growth dilution and loss processes have not been taken into account. In cases where the BCF_{ss} is very much higher than the BCF_k , the derivation of the uptake and depuration rate constants should be checked for errors and re-evaluated. A different fitting procedure might improve the estimate of BCF_k (*cf.* Appendix 5).

Test Report

Apart from the test substance information indicated in paragraph 3, the test report includes the following information:

Test substance:

Physical nature and, where relevant, physicochemical properties;

- Chemical identification data, 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).
- For multi-constituent substances and UVCB (chemical substances of Unknown or Variable composition, Complex reaction products and Biological materials) description, as far as possible, of the chemical identity of the individual constituents and, for each, of its percentage of the total mass of the substance. How the analytical method used in the test reflects a measure of the concentration of the substance should be summarised; all analytical procedures should be described including the accuracy of the method, method detection limit, and limit of quantification.
- If radiolabelled, the precise position of the labelled atom(s) and the percentage of radioactivity associated with impurities.
- Information on the test substance toxicity to fish (ideally the test species). The toxicity should be reported as an acute 96 h LC_{50} and a NOAEC & LOAEC from a chronic study (*i.e.*, an early life stage test or a full life cycle test, if available).
- Storage conditions of the test chemical or test substance and stability of the test chemical or test substance under storage conditions if stored prior to use.

Test species:

Scientific name, strain, source, any pre-treatment, acclimation, age, sex (if relevant), size-range (weight and length), etc.

⁽¹⁾ These percentages assume that the analytical methods are reliable and the half life is < 14 days. If the analytical methods are less reliable or the half life is (greatly) increased these numbers will become larger.

Test conditions:

- Test procedure used (e.g. flow-through or semi-static); regular study or minimised design (including rationale and justification).
- Type and characteristics of illumination used and photoperiod(s).
- Test design (e.g. number and size of test chambers, water volume replacement rate, loading rate, number of replicates, number of fish per replicate, number of test concentrations, length of uptake and depuration phases, sampling frequency for fish and water samples).
- Method of preparation of stock solutions and frequency of renewal (the solvent, its concentration and its contribution to the organic carbon content of test water should be given, when used) or description of alternative dosing system.
- The nominal test concentrations, the means of the measured values and their standard deviations in the test vessels and the method and frequency by which these were attained.
- Source of the dilution water, description of any pre-treatment, results of any demonstration of the ability of test fish to live in the water, and water characteristics: pH, hardness, temperature, dissolved oxygen concentration, residual chlorine levels (if measured), total organic carbon, suspended solids, salinity of the test medium (if appropriate) and any other measurements made.
- Water quality within test vessels, pH, hardness, TOC, temperature and dissolved oxygen concentration; methods used and frequency of measurements.
- Detailed information on feeding, e.g. type of food(s), source, composition (at least lipid and protein content if possible), selected feeding rate, amount given and frequency;
- Information on the treatment of fish and water samples, including details of preparation, storage, extraction and analytical procedures (and precision) for the test substance and lipid content.
- Methods used for treatment randomisation and assignment of fish to test vessels.
- Date of introduction of test organisms to test solutions and test duration.
- Description of range-finding tests and results, if available.

Results:

- Results from any preliminary study performed.
- Mortality of the control fish and the fish in each exposure chamber and any observed abnormal behaviour.
- Information on any adverse effects observed.
- Complete description of all chemical analysis procedures employed including limits of detection and quantification, variability and recovery.
- The lipid content of the fish, including the method used, and if derived, lipid normalisation factor (L_n , factor to express results relative to fish lipid content of 5 %).
- Tabulated fish weight (and length) data, linked to individual fish chemical concentrations (and lipid content, if applicable), both for control and exposure groups (for example using unique identifiers for each sampled fish) and calculations for derived growth rate constant(s).
- Tabulated test substance concentration data in fish (C_f , linked to individual fish) and water (C_w) (with mean values for test group and control, standard deviation and range, if appropriate) for all sampling times (C_f expressed in mg/kg wet weight of whole body or specified tissues thereof, e.g. lipid, and C_w in mg/l). C_w values for the control series (background should also be reported).

- Curves (including all measured data), showing the following (if applicable, concentrations may be expressed in relation to the whole body and the lipid content normalised to 5 % of the animal or specified tissues thereof):
 - growth, *i.e.* fish weight vs. time or natural logarithm transformed weight vs. time (including the derived growth rate constant, k_g);
 - the uptake and depuration of the test substance in the fish (on one graph);
 - the time to steady-state (if achieved);
 - natural logarithm transformed concentration vs. uptake time (including the derived uptake rate constant k_1);
 - natural logarithm transformed concentration (ln concentration) vs. depuration time (including the derived depuration rate constant k_2); and
 - both uptake and depuration phase curves, showing both the data and the fitted model.
- If a visual inspection of a plot shows obvious outliers, a statistically valid outlier test may be applied to remove spurious data points as well as documented justification for their omission.
- The steady-state bioconcentration factor, (BCF_{ss}), if steady-state is (almost) achieved.
- Kinetic bioconcentration factor (BCF_k) and derived uptake and depuration rate constants k_1 and k_2 , together with the variances in k_2 (slope and intercept) if sequential fitting is used.
- Confidence limits, standard deviation (as available) and methods of computation/data analysis for each parameter for each concentration of test substance used.
 - Any information concerning radiolabelled test substance metabolites and their accumulation.
 - Growth rate constant(s) (including 95 % confidence interval(s)) and calculated growth-corrected depuration rate constant (k_{2g}), half-life and BCF (BCF_{kg}) values.
 - Anything unusual about the test, any deviation from these procedures, and any other relevant information.
- A summary table of relevant measured and calculated data, as hereafter:

Substance Uptake and Depuration Rate Constants and Bioconcentration Factors (BCF)	
k_g (growth rate constant; day^{-1}):	Insert Value (95 % CI) ⁽¹⁾
k_1 (overall uptake rate constant; $\text{l kg}^{-1} \text{ day}^{-1}$):	Insert Value (95 % CI) ⁽¹⁾
k_2 (overall depuration rate constant; day^{-1}):	Insert Value (95 % CI) ⁽¹⁾
k_{2g} (growth-corrected depuration rate constant; day^{-1}):	Insert Value (95 % CI) ⁽¹⁾
C_f (chemical concentration in the fish at steady-state; mg kg^{-1}):	Insert Value \pm SD ⁽²⁾
C_w (chemical concentration in the water; mg l^{-1}):	Insert Value \pm SD ⁽²⁾
L_n (lipid normalisation factor):	Insert Value ⁽³⁾

Substance Uptake and Depuration Rate Constants and Bioconcentration Factors (BCF)	
BCF _{ss} (steady-state BCF; l kg ⁻¹)	Insert Value ± SD ⁽²⁾
BCF _{SSL} (lipid normalised steady-state BCF; l kg ⁻¹):	Insert Value ⁽³⁾ ± SD ⁽²⁾
BCF _k (kinetic BCF; l kg ⁻¹)	Insert Value (95 % CI) ⁽¹⁾
BCF _{kg} (growth-corrected kinetic BCF; l kg ⁻¹)	Insert Value (95 % CI) ⁽¹⁾
t _{1/2g} (growth-corrected half-life; day):	Insert Value (95 % CI) ⁽¹⁾
BCF _{KL} (lipid-normalised kinetic BCF; l kg ⁻¹):	Insert Value
BCF _{KLg} (lipid-normalised growth corrected kinetic BCF; l kg ⁻¹):	Insert Value
⁽¹⁾ CI: confidence interval (where possible to estimate) ⁽²⁾ SD: Standard deviation (where possible to estimate)	

Results reported as “not detected/quantified at the limit of detection/quantification” by pre-test method development and experimental design should be avoided, since such results cannot be used for rate constant calculations.

C.13 - II: Minimised Aqueous Exposure Fish Test

INTRODUCTION

The growing experience that has been gained in conducting and interpreting the full test, both by laboratories and regulatory bodies, shows that — with some exceptions — first order kinetics apply for estimating uptake and depuration rate constants. Thus, uptake and depuration rate constants can be estimated with a minimum of sampling points, and the kinetic BCF derived.

The initial purpose of examining alternative designs for BCF studies was to develop a small test to be used in an intermediate testing step to refute or confirm BCF estimates based on K_{ow} and QSARs and so eliminate the need for a full study for many substances, and to minimise cost and animal use via reduction in sampling and in the number of analytical sequences performed. While following the main design of the previous test method to allow integration of test results with existing BCF data, and to ease performance of testing and data interpretation, the aim was to provide BCF estimates of adequate accuracy and precision for risk assessment decisions. Many of the same considerations apply as in the full test, e.g. validity criteria (*cf.* paragraph 24) and stopping a test if insignificant uptake is seen at the end of the uptake phase (*cf.* paragraphs 16 and 38).

Substances that would be eligible for the minimised test design should belong to the general domain that this test method was developed for, *i.e.* non-polar organic substances (*cf.* paragraph 49). If there is any indication that the substance to be tested might show a different behaviour (e.g. a clear deviation from first-order kinetics), a full test should be conducted for regulatory purposes.

Typically, the minimised test is not run over a shorter period than the standard BCF test, but comprises less fish sampling (see Appendix 6 for the rationale). However, the depuration period may be shortened for rapidly depurating substances to avoid concentrations in the fish falling below the limit of detection/quantification before the end of the test. A minimised exposure fish test with a single concentration can be used to determine the need for a full test, and if the resulting data used to calculate rate constants and BCF are robust (*cf.* paragraph 93), the full test may be waived if the resulting BCF is far from regulatory values of concern.

In some cases, it may be advantageous to perform the minimised test design with more than one test concentration as a preliminary test to determine whether BCF estimates for a substance are concentration dependent. If the BCF estimates from the minimised test show concentration dependence, the performance of the full test will be necessary. If, based on such a minimised test, BCF estimates are not concentration dependent but the results are not considered definitive, then any subsequent full test could be performed at a single concentration, thereby reducing animal use in comparison to a two (or more) concentration full test.

Substances potentially eligible for the minimised test should:

- Be likely to exhibit approximate first order uptake and depuration kinetics, e.g. derived from read-across with similar substances;
- Have a $\log K_{ow} < 6$ unless rapid metabolism is expected ⁽¹⁾;
- Be sufficiently water-soluble with respect to the analytical technique (*cf.* paragraph 24);
- Be clearly quantifiable (*i.e.* concentrations should be at least one order of magnitude above the limit of quantification), both in fish and water, radioactive labelling is recommended (*cf.* paragraph 23); and
- Have a depuration period greater than its predicted half-life (*cf.* Appendix 5 for calculations), or the duration of depuration should be adjusted accordingly (*cf.* paragraph 91). An exception to this rule is allowed if rapid metabolism of the substance is expected.

SAMPLING SCHEDULE FOR STUDIES FOLLOWING THE MINIMISED DESIGN

Fish sampling

Fish sampling is reduced to four sampling points:

- At the middle and end of the uptake phase (the latter being the beginning of depuration as well), e.g. after 14 and 28 days (33).
- At the middle of the depuration phase and at termination of the study (where substance concentration is < 10 % of the maximum concentration, or at least clearly past one half-life of the substance), e.g. after 7 and 14 days of depuration (33). If rapid depuration is expected or observed, it may be necessary to shorten the depuration period to avoid concentrations in the fish falling below the limit of quantification.
- Lipid measurement as in full study.
- Growth correction as in full study.
- The BCF is calculated as a kinetic BCF.

Water sampling

For the minimised design, water is sampled as in full study (*cf.* paragraph 54) or at least five times equally divided over the uptake phase, and weekly in the depuration phase.

⁽¹⁾ The minimised test may in fact be used to demonstrate rapid metabolism when it is known that rapid metabolism is likely.

Design modifications

Taking into account the test substance properties, valid QSAR predictions and the specific purpose of the study, some modifications in the design of the study can be considered:

- If greater precision is needed, more fish (6 or 8 instead of 4) could be used for the sample at the end of the uptake phase.
- Inclusion of an 'extra' group of fish to be used if depuration at 14 days (or the predicted end of the depuration phase) has not been sufficient for adequate depuration (*i.e.* > 50 %). If the predicted duration of the depuration phase is shorter or longer than 14 days, the sampling schedule should be adapted (*i.e.* one group of fish at the predicted end of the depuration phase, and one group after half that time).
- Use of two test concentrations to explore possible concentration dependence. If the results of the minimised test, conducted with two test concentrations, show that the BCF is not concentration dependent (*i.e.* differ less than 20 %), one test concentration may be considered sufficient in a full test, if it is conducted.
- It seems likely that models of bioaccumulation processes such as those proposed by Arnot *et al.* (35) can be used to assist in planning the length of uptake and depuration phases (see also Appendix 5).

Calculations

The rationale for this approach is that the bioconcentration factor in a full test can either be determined as a steady-state bioconcentration factor (BCF_{ss}) by calculating the ratio of the concentration of the test substance in the fish's tissue to the concentration of the test substance in the water, or by calculating the kinetic bioconcentration factor (BCF_k) as the ratio of the uptake rate constant k_1 to the depuration rate constant k_2 . The BCF_k is valid even if a steady-state concentration of a substance is not achieved during uptake, provided that uptake and depuration act approximately according to first order kinetic processes. As an absolute minimum two data points are required to estimate uptake and depuration rate constants, one at the end of the uptake phase (*i.e.* at the beginning of the depuration phase) and one at the end (or after a significant part) of the depuration phase. The intermediate sampling point is recommended as a check on the uptake and depuration kinetics ⁽¹⁾. For calculations, see Appendixes 5 and 6.

Interpretation of the results

To assess the validity and informative value of the test, verify that the depuration period exceeds one half-life. Also, the BCF_{km} (kinetic BCF derived from a minimised test) should be compared to the minimised BCF_{ss} value (which is the BCF_{ss} calculated at the end of the uptake phase, assuming that steady-state has been reached. This can only be assumed, as the number of sampling points is not sufficient for proving this). If the $BCF_{km} < \text{minimised } BCF_{ss}$, the minimised BCF_{ss} should be the preferred value. If BCF_{km} is less than 70 % of the minimised BCF_{ss} , the results are not valid, and a full test should be conducted.

If the minimised test gives a BCF_{km} in the region of any value of regulatory concern, a full test should be conducted. If the result is far from any regulatory value of concern (well above or below), a full test may not be necessary, or a single concentration full test may be conducted if required by the relevant regulatory framework.

If a full test is found to be necessary after a minimised test at one concentration, this can be conducted at a second concentration. If the results are consistent, a further full test at a different concentration can be waived, as the bioconcentration of the substance is not expected to be concentration dependent. If the minimised test has been conducted at two concentrations, and the results show no concentration dependence, the full test may be conducted with only one concentration (*cf.* paragraph 87).

⁽¹⁾ When only two data points are measured, estimates of the confidence limits for BCF_{km} can be made using bootstrap methods. When intermediate data points are also available confidence limits for BCF_{km} can be calculated as in the full test.

Test report

The test report for the minimised test should include all the information demanded for the full test (cf. paragraph 81), except that which is not possible to elaborate (i.e. a curve showing the time to steady-state and the steady-state bioconcentration factor; for the latter the minimised BCF_{ss} should be given instead). Additionally, it should also include the reasoning for using the minimised test and the resulting BCF_{km} .

C.13 - III: Dietary Exposure Bioaccumulation Fish Test

INTRODUCTION

The method described in this section should be used for substances where the aqueous exposure methodology is not practicable (for example because stable, measurable water concentrations cannot be maintained, or adequate body burdens cannot be achieved within 60 days of exposure; see previous sections on the aqueous exposure method). It should be realised though that the endpoint from this test will be a dietary biomagnification factor (BMF) rather than a bioconcentration factor (BCF) ⁽¹⁾.

In May 2001 a new method for the bioaccumulation testing of poorly water soluble organic substances was presented at the SETAC Europe conference held in Madrid (36). This work built on various reported bioaccumulation studies in the literature using a dosing method involving spiked feed (e.g. (37)). Early in 2004 a draft protocol (38), designed to measure the bioaccumulation potential of poorly water soluble organic substances for which the standard water exposure bioconcentration method was not practicable, together with a supporting background document (39), was submitted to an EU PBT working group. Further justification given for the method was that potential environmental exposure to such poorly soluble substances (i.e. $\log K_{ow} > 5$) may be largely via the diet (cf. (40) (41) (42) (43) (44)). For this reason, dietary exposure tests are referred to in some published chemicals regulations ⁽²⁾. It should be realised however, that in the method described here exposure via the aqueous phase is carefully avoided and thus a BMF value from this test method cannot directly be compared to a BMF value from a field study (in which both water and dietary exposure may be combined).

This section of the present test method is based on this protocol (38) and is a new method that did not appear in the previous version of TM C.13. This alternative test allows the dietary exposure pathway to be directly investigated under controlled laboratory conditions.

Potential investigators should refer to paragraphs 1 to 14 of this test method for information on when the dietary exposure test may be preferred over the aqueous exposure test. Information on the various substance considerations is laid out, and should be considered before a test is conducted.

The use of radiolabelled test substances can be considered with similar considerations as for the aqueous exposure method (cf. paragraphs 6 and 65).

The dietary method can be used to test more than one substance in a single test, so long as certain criteria are fulfilled; these are explored further in paragraph 112. For simplicity the methodology here describes a test using only one test substance.

The dietary test is similar to the aqueous exposure method in many respects with the obvious exception of the exposure route. Hence many aspects of the method described here overlap with the aqueous exposure method described in the previous section. Cross-reference to relevant paragraphs in the previous section has been made as far as possible, but in the interests of readability and understanding a certain amount of duplication is unavoidable.

⁽¹⁾ See Appendix 1 for definitions and units

⁽²⁾ For the purpose of Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) (OJ L 396, 30.12.2006, p. 1), this issue is addressed in the "Guidance on Information Requirements and Chemical Safety Assessment", chapter R.7c, R.7.10.3.1; R.7.10.4.1; and figure R7.10-2.

PRINCIPLE OF THE TEST

Flow-through or semi-static conditions can be employed (*cf.* paragraph 4); flow-through conditions are recommended to limit potential exposure of test substance via water as a result of any desorption from spiked food or faeces. The test consists of two phases: uptake (test substance-spiked feed) and depuration (clean, untreated feed) (*cf.* paragraph 16). In the uptake phase, a “test” group of fish are fed a set diet of a commercial fish food of known composition, spiked with the test substance, on a daily basis. Fish ideally should consume all of the offered food (*c.f.* paragraph 141). Fish are then fed the pure, untreated commercial fish food during the depuration phase. As for the aqueous exposure method, more than one test group with different spiked test substance concentrations can be used if necessary, but for the majority of highly hydrophobic organic test substances one test group is sufficient (*cf.* paragraphs 49 and 107). If semi-static conditions are used fish should be transferred to a new medium and/or a new test chamber at the end of the uptake phase (in case the medium and/or apparatus used in the uptake phase has been contaminated with the test substance through leaching). The concentrations of the test substance in the fish are measured in both phases of the test. In addition to the group of fish fed the spiked diet (the test group), a control group of fish is held under identical conditions and fed identically except that the commercial fish food diet is not spiked with test substance. This control group allows background levels of test substance to be quantified in unexposed fish and serves as a comparison for any treatment-related adverse effects noted in the test group(s) ⁽¹⁾. It also allows comparison of growth rate constants between groups as a check that similar quantities of offered diet have been consumed (potential differences in palatability between diets should also be considered in explaining different growth rate constants; *cf.* paragraph 138). It is important that during both the uptake and depuration phases, diets of nutritional equivalency are fed to the test and control groups.

An uptake phase that lasts 7-14 days is generally sufficient, based on experience from the method developers (38) (39). This range should minimise the cost of undertaking the test whilst still ensuring sufficient exposure for most substances. However, in some cases the uptake phase may be extended (*cf.* paragraph 127). During the uptake phase the substance concentration in the fish may not reach steady-state so data treatment and results from this method are usually based on a kinetic analysis of tissue residues. (Note: Equations for estimating time to steady-state can be applied here as for the aqueous exposure test — see Appendix 5). The depuration phase begins when the fish are first fed unspiked diet and typically lasts for up to 28 days or until the test substance can no longer be quantified in whole fish, whichever is the sooner. The depuration phase can be shortened or lengthened beyond 28 days, depending on the change with time in measured chemical concentrations and fish size.

This method allows the determination of the substance-specific half-life ($t_{1/2}$, from the depuration rate constant, k_2), the assimilation efficiency (absorption across the gut; a), the kinetic dietary biomagnification factor (BMF_k), the growth-corrected kinetic dietary biomagnification factor (BMF_{kg}), and the lipid-corrected ⁽²⁾ kinetic dietary biomagnification factor (BMF_{kgl}) (and/or the growth- and lipid-corrected kinetic dietary biomagnification factor, BMF_{kggl}) for the test substance in fish. As for the aqueous exposure method, increase in fish mass during the test will result in dilution of test substance in growing fish and thus the (kinetic) BMF will be underestimated if not corrected for growth (*cf.* paragraphs 162 and 163). In addition, if it is estimated that steady-state was reached in the uptake phase an indicative steady-state BMF can be calculated. Approaches are available that make it feasible to estimate a kinetic bioconcentration factor (BCF_k) from data generated in the dietary study (e.g. (44) (45) (46) (47) (48)). Pros and cons of such approaches are discussed in Appendix 8.

The test was designed primarily for poorly soluble non-polar organic substances that follow approximately first order uptake and depuration kinetics in fish. In case a substance is tested that does not follow approximately first order uptake and depuration kinetics, then more complex models should be employed (see references in Appendix 5) and advice from a biostatistician and/or pharmacokineticist sought.

⁽¹⁾ For most test substances, there should ideally be no detections in the control water. Background concentrations should only be relevant to naturally occurring materials (e.g., some metals) and substances that are ubiquitous in the environment.

⁽²⁾ As the BMF is defined as the ratio of the concentration of a substance in an organism to that in the organism's food at steady-state, lipid is taken into account by correcting for the contents of lipid in the organism and in the food, hence it is described more accurately as a “correction”. This approach differs from “normalisation” to a set organism lipid content as is done in the aqueous exposure bioconcentration test.

The BMF is normally determined using test substance analysis of whole fish (wet weight basis). If relevant for the objectives of the study, specific tissues (e.g. muscle, liver) can be sampled if the fish is divided into edible and non-edible parts (*cf.* paragraph 21). Furthermore, removal and separate analysis of the gastrointestinal tract may be employed to determine the contribution to whole fish concentrations for sample points at the end of the uptake phase and near the beginning of the depuration phase, or as part of a mass balance approach.

Lipid content of sampled whole fish should be measured so that concentrations can be lipid-corrected, taking account of lipid content of both the diet and the fish (*cf.* paragraphs 56 and 57, and Appendix 7).

Fish weight of sampled individuals should be measured and recorded, and be linked to the analysed chemical concentration for that individual (e.g. reported using a unique identifier code for each fish sampled), for the purpose of calculating growth that may occur during the test. Fish total length should also be measured where possible ⁽¹⁾. Weight data are also necessary for estimating BCF using depuration data from the dietary test.

INFORMATION ON THE TEST SUBSTANCE

Information on the test substance as described in paragraphs 3 and 22 should be available. An analytical method for test substance concentrations in water is not usually necessary; methods with suitable sensitivity for measuring concentrations in fish food and fish tissue are required.

The method can be used to test more than one substance in a single test. However, test substances should be compatible with one another such that they do not interact or change their chemical identity upon spiking into fish food. The aim is that measured results for each substance tested together should not differ greatly from the results that would be given if individual tests had been run on each test substance. Preliminary analytical work should establish that each substance can be recovered from a multiply-spiked food and fish tissue sample with i) high recoveries (e.g. > 85 % of nominal) and ii) the necessary sensitivity for testing. The total dose of substances tested together should be below the combined concentration that might cause toxic effects (*cf.* paragraph 51). Furthermore, possible adverse effects in fish and the potential for interactive effects (e.g. metabolic effects) associated with testing multiple substances simultaneously should be taken into consideration in the experimental design. Simultaneous testing of ionisable substances should be avoided. In terms of exposure, the method is also suitable for complex mixtures (*cf.* paragraph 13, although the same limitations in analysis as for any other method will apply).

VALIDITY OF THE TEST

For a test to be valid the following conditions apply (*cf.* paragraph 24):

- Water temperature variation is less than ± 2 °C in treatment or control groups
- Concentration of dissolved oxygen does not fall below 60 % of the air saturation value
- The concentration of the test substance in fish food before and at the end of the uptake phase is within a range of ± 20 % (based on at least three samples at both time points)
- A high degree of homogeneity of substance in food should be demonstrated in preliminary analytical work on the spiked diet; at least three sample concentrations for the substance taken at test start should not vary more than ± 15 % from the mean

⁽¹⁾ Total length should also be recorded during the test as it is a good indicator of whether an adverse effect has occurred.

- Concentrations of test substance are not detected, or are present only at typical trace levels, in un-spiked food or control fish tissues relative to treated samples
- Mortality or other adverse effects/disease in both control and test group fish should be $\leq 10\%$ at the end of the test; if the test is extended for any reason, adverse effects in both groups are $\leq 5\%$ per month, and $\leq 30\%$ cumulatively. Significant differences in average growth between the test and the control groups of sampled fish could be an indication of a toxic effect of the test substance.

REFERENCE SUBSTANCES

If a laboratory has not performed the assay before or substantial changes (e.g. change of fish strain or supplier, different fish species, significant change of fish size, fish food or spiking method, etc.) have been made, it is advisable that a technical proficiency study is conducted, using a reference substance. The reference substance is primarily used to establish whether the food spiking technique is adequate to ensure maximum homogeneity and bioavailability of test substances. One example that has been used in the case of non-polar hydrophobic substances is hexachlorobenzene (HCB), but other substances with existing reliable data on uptake and biomagnification should be considered due to the hazardous property of HCB⁽¹⁾. If used, basic information on the reference substance should be presented in the test report, including name, purity, CAS number, structure, toxicity data (if available) as for test substances (*cf.* paragraphs 3 and 22).

DESCRIPTION OF THE METHOD

Apparatus

Materials and apparatus should be used as described in the aqueous exposure method (*cf.* paragraph 26). A flow-through or static renewal test system that provides a sufficient volume of dilution water to the test tanks should be used. The flow rates should be recorded.

Water

Test water should be used as described in the aqueous exposure method (*cf.* paragraphs 27-29). The test medium should be characterised as described and its quality should remain constant during the test. The natural particle content and total organic carbon should be as low as possible (≤ 5 mg/l particulate matter; ≤ 2 mg/l total organic carbon) before test start. TOC need only be measured before the test as part of the test water characterisation (*cf.* paragraph 53).

Diet

A commercially available fish food (floating and/or slow sinking pelletised diet) that is characterised in terms of at least protein and fat content is recommended. The food should have a uniform pellet size to increase the efficiency of the feed exposure, i.e. the fish will eat more of the food instead of eating the larger pieces and missing the smaller ones. The pellets should be appropriately sized for the size of the fish at the start of the test (e.g. pellet diameters roughly 0,6-0,85 mm for fish between 3 and 7 cm total length, and 0,85-1,2 mm for fish between 6 and 12 cm total length may be used). Pellet size may be adjusted depending on fish growth at the start of the depuration phase. An example of a suitable food composition, as commercially supplied, is given in Appendix 7. Test diets with total lipid content between 15 and 20 % (w/w) have commonly been used in the development of this method. Fish food with such a high lipid concentration may not be available in some regions. In such cases studies could be run with a lower lipid concentration in the food, and if necessary the feeding rate adjusted appropriately to maintain fish health (based on preliminary testing). The total lipid content of the test group and control group diets needs to be measured and recorded before the start of the test and at the end of the uptake phase. Details provided by the commercial feed supplier of analysis for nutrients, moisture, fibre and ash, and if possible minerals and pesticide residues (e.g. "standard" priority pollutants), should be presented in the study report.

⁽¹⁾ HCB is listed in Annexes A and C to the Stockholm Convention, and in Annexes I and III of Regulation (EC) No 850/2004 on persistent organic pollutants (OJ L 158, 30.4.2004, p. 7)

When spiking the food with test substance, all possible efforts should be made to ensure homogeneity throughout the test food. The concentration of test substance in the food for the test group should be selected taking into account the sensitivity of the analytical technique, the test substance's toxicity (NOEC if known) and relevant physicochemical data. If used, the reference substance should preferably be incorporated at a concentration around 10 % of that of the test substance (or in any case as low as is practicable), subject to analysis sensitivity (e.g. for hexachlorobenzene a concentration in the food of 1-100 µg/g has been found to be acceptable; cf. (47) for more information on assimilation efficiencies of HCB).

The test substance can be spiked to the fish food in several ways depending on its physical characteristics and solubility (see Appendix 7 for more details on spiking methods):

- If the substance is soluble and stable in triglycerides, the substance should be dissolved in a small amount of fish oil or edible vegetable oil before mixing with fish food. In this instance, care should be taken to avoid producing a ration that is too high in lipid, taking into account the natural lipid content of the spiked feed, by adding the minimum known quantity of oil required to achieve distribution and homogeneity of the test substance in the food, or;
- The food should be spiked using a suitable organic solvent, so long as homogeneity and bioavailability are not compromised (it is possible that (micro)crystals of the test substance may form in the food as a consequence of solvent evaporation and there is no easy way to prove this has not occurred; cf. (49)), or;
- Non-viscous liquids should be added directly to fish food but they should be well mixed to promote homogeneity and facilitate good assimilation. The technique for mixing should ensure homogeneity of the spiked feed.

In few cases, e.g. less hydrophobic test substances more likely to desorb from the food, it may be necessary to coat prepared food pellets with a small quantity of corn/fish oil (see paragraph 142). In such cases, control food should be treated similarly and the final prepared feed used for lipid measurement.

If used, the results of the reference substance should be comparable with literature study data carried out under similar conditions with a comparable feeding rate (cf. paragraph 45) and reference substance-specific parameters should meet the relevant criteria in paragraph 113 (3rd, 4th and 5th points).

If an oil or carrier solvent is used as a vehicle for the test substance, an equivalent amount of the same vehicle (excluding test substance) should be mixed with the control diet in order to maintain equivalency with the spiked diet. It is important that during both the uptake and depuration phases, diets of nutritional equivalency are fed to the test and control groups.

The spiked diet should be stored under conditions that maintain stability of the test substance within the feed mix (e.g. refrigeration) and these conditions reported.

Selection of fish species

Fish species as specified for the aqueous exposure may be used (cf. paragraph 32 and Appendix 3). Rainbow trout (*Oncorhynchus mykiss*), carp (*Cyprinus carpio*) and fathead minnow (*Pimephales promelas*) have been commonly used in dietary bioaccumulation studies with organic substances before the publication of this TM. The test species should have a feeding behaviour that results in rapid consumption of the administered food ration to ensure that any factor influencing the concentration of the test substance in food (e.g. leaching into the water and the possibility of aqueous exposure) is kept to a minimum. Fish within the recommended size/weight range (cf. Appendix 3) should be used. Fish should not be so small as to hamper ease of analyses on an individual basis. Species tested during a life-stage with rapid growth can complicate data interpretation, and high growth rates can influence the calculation of assimilation efficiency (¹).

(¹) For rapid growth during the uptake phase the true feeding rate will decrease below that set at the beginning of exposure.

Holding of fish

Acclimatisation, mortality and disease acceptance criteria are the same as for the aqueous exposure method prior to test conductance (*cf.* paragraphs 33-35).

PERFORMANCE OF THE TEST

Pre-study work and range-finding test

Pre-study analytical work is necessary to demonstrate recovery of the substance from spiked food/spiked fish tissue. A range-finding test to select a suitable concentration in the food is not always necessary. For the purposes of showing that no adverse effects are observed and evaluating the palatability of spiked diet, sensitivity of analytical method for fish tissue and food, and selection of suitable feeding rate and sampling intervals during depuration phase etc., preliminary feeding experiments may be undertaken but are not obligatory. A preliminary study may be valuable to estimate numbers of fish needed for sampling during the depuration phase. This can result in significant reduction in the number of fish used, especially for test substances that are particularly susceptible to metabolism.

Conditions of exposure

Uptake Phase duration

An uptake phase of 7-14 days is usually sufficient, during which one group of fish are fed the control diet and another group of fish the test diet daily at a fixed ration dependent on the species tested and the experimental conditions, e.g. between 1-2 % of body weight (wet weight) in the case of rainbow trout. The feeding rate should be selected such that fast growth and large increase of lipid content are avoided. If needed the uptake phase may be extended based on practical experience from previous studies or knowledge of the test substance's (or analogue's) uptake/depuration in fish. The start of the test is defined as the time of first feeding with spiked food. An experimental day runs from the time of feeding to shortly before the time of next feeding (e.g. one hour). Thus the first experimental day of uptake runs from the time of first feeding with spiked food and ends shortly before the second feeding with spiked food. In practice the uptake phase ends shortly before (e.g. one hour) the first feeding with unspiked test substance as the fish will continue to digest spiked food and absorb the test substance in the intervening 24 hours. It is important to ensure that a sufficiently high (non-toxic) body burden of the test substance is achieved with respect to the analytical method, so that at least an order of magnitude decline can be measured during the depuration phase. In special cases an extended uptake phase (up to 28 days) may be used with additional sampling to gain an insight into uptake kinetics. During uptake the concentration in the fish may not reach steady-state. Equations for estimating time to steady-state, as an indication of the likely duration needed to achieve appreciable fish concentrations, can be applied here as for the aqueous exposure test (*cf.* Appendix 5).

In some cases it may be known that uptake of substance in the fish over 7-14 days will be insufficient for the food concentration used to reach a high enough fish concentration to analyse at least an order of magnitude decline during depuration, either due to poor analytical sensitivity or to low assimilation efficiency. In such cases it may be advantageous to extend the initial feeding phase to longer than 14 days, or, especially for highly metabolisable substances, a higher dietary concentration should be considered. However, care should be taken to keep the body burden during uptake below the (estimated) chronic no effect concentration (NOEC) in fish tissue (*cf.* paragraph 138).

Duration of the depuration phase

Depuration typically lasts for up to 28 days, beginning once the test group fish are fed pure, untreated diet after the uptake phase. Depuration begins with the first feeding of “unspiked” food rather than straight after the last “spiked” food feeding as the fish will continue to digest the food and absorb the test substance in the intervening 24 hours, as noted in paragraph 126. Hence the first sample in the depuration phase is taken shortly before the second feeding with unspiked diet. This depuration period is designed to capture substances with a potential half-life of up to 14 days, which is consistent with that of bioaccumulative substances ⁽¹⁾, so 28 days comprises two half-lives of such substances. In cases of very highly bioaccumulating substances it may be advantageous to extend the depuration phase (if indicated by preliminary testing).

If a substance is depurated very slowly such that an exact half-life may not be determined in the depuration phase, the information may still be sufficient for assessment purposes to indicate a high level of bioaccumulation. Conversely, if a substance is depurated so fast that a reliable time zero concentration (concentration at the end of uptake/start of depuration, $C_{0,d}$) and k_2 cannot be derived, a conservative estimate of k_2 can be made (cf. Appendix 7).

If analyses of fish at earlier intervals (e.g. 7 or 14 days) show that the substance has depurated below quantification levels before the full 28-day period, then subsequent sampling may be discontinued and the test terminated.

In few cases no measurable uptake of the test substance may have occurred at the end of the uptake period (or with the second depuration sample). If it can be demonstrated that: i) the validity criteria in paragraph 113 are fulfilled; and ii) lack of uptake is not due to some other shortcoming of the test (e.g. uptake duration not long enough, deficiency in food spiking technique leading to poor bioavailability, lack of sensitivity of the analytical method, fish not consuming food, etc.); it may be possible to terminate the study without the need to re-run it with a longer uptake duration. If preliminary work has indicated that this may be the case, analysis of faeces, if possible, for undigested test substance may be advisable as part of a “mass balance” approach.

Numbers of test fish

Similar to the aqueous exposure test, fish of similar weight and length should be selected, with the smallest fish being no less than two-thirds of the weight of the largest (cf. paragraphs 40-42).

The total number of fish for the study should be selected based on the sampling schedule (a minimum of one sample at the end of the uptake phase and four to six samples during the depuration phase, but depending on the phases' durations), taking into account the sensitivity of the analytical technique, the concentration likely to be achieved at the end of the uptake phase (based on prior knowledge) and the depuration duration (if prior knowledge allows estimation). Five to ten fish should be sampled at each event, with growth parameters (weight and total length) being measured before chemical or lipid analysis.

Owing to the inherent variability in the size, growth rate, and physiology among fish and the likely variation in the quantity of administered diet that each fish consumes, at least five fish should be sampled at each interval from the test group and five from the control group in order to adequately establish the average concentration and its variability. The variability among the fish used is likely to contribute more to the overall uncontrolled variability in the test than the variability inherent in the analytical methodologies employed, and thus justifies the use of up to ten fish per sample point in some cases. However, if background test substance concentrations in control fish are not measurable at the start of depuration, chemical analysis of two-three control fish at the final sampling interval only may be sufficient so long as the remaining control fish at all sample points are still sampled for weight and total length (so that the same number are sampled from test and control groups for growth). Fish should be stored, weighed individually (even if it proves necessary for the sample results to be combined subsequently) and total length measured.

⁽¹⁾ In an aqueous exposure study, a 14-day half-life would correspond to a BCF of ca. 10 000 L/kg using fish of 1 g with a corresponding uptake rate of about 500 L/kg/d (according to the equation of Sijm *et al* (46)).

For a standard test with, for example, a 28-day depuration duration including five depuration samples, this means a total of 59-120 fish from test and 50-110 from control groups, assuming that the substance's analytical technique allows lipid content analysis to be carried out on the same fish. If lipid analysis cannot be conducted on the same fish as chemical analysis, and using control fish only for lipid analysis is also not feasible (*cf.* paragraph 56), an additional 15 fish would be required (three from the stock population at test start, three each from control and test groups at the start of depuration and three each from control and test groups at the end of the experiment). An example sampling schedule with fish numbers can be found in Appendix 4.

Loading

Similarly high water-to-fish ratios should be used as for the aqueous exposure method (*cf.* paragraphs 43 and 44). Although fish-to-water loading rates do not have an effect on exposure concentrations in this test, a loading rate of 0,1-1,0 g of fish (wet weight) per litre of water per day is recommended to maintain adequate dissolved oxygen concentrations and minimise test organism stress.

Test diet and Feeding

During the acclimatisation period, fish should be fed an appropriate diet as described above (paragraph 117). If the test is being conducted under flow-through conditions, the flow should be suspended while the fish are fed.

During the test, the diet for the test group should adhere to that described above (paragraphs 116-121). In addition to consideration of substance-specific factors, analytical sensitivity, expected concentration in the diet under environmental conditions and chronic toxicity levels/body burden, selection of the target spiking concentration should take into account palatability of the food (so that fish do not avoid eating). Nominal spiking concentration of the test substance should be documented in the report. Based on experience, spiking concentrations in the range of 1-1 000 µg/g provide a practical working range for test substances that do not exhibit a specific toxic mechanism. For substances acting via a non-specific mechanism, tissue residue levels should not exceed 5 µmol/g lipid since residues above this level are likely to pose chronic effects (19) (48) (50) ⁽¹⁾. For other substances care should be taken that no adverse effects occur from the accumulated exposure (*cf.* paragraph 127). This is especially true if more than one substance is being tested simultaneously (*cf.* paragraph 112).

The appropriate amount of the test substance can be spiked to the fish food in one of three ways, as described in paragraph 119 and Appendix 7. The methods and procedures for spiking the feed should be documented in the report. Untreated food is fed to the control fish, containing an equivalent quantity of unspiked oil vehicle if this has been used in the spiked feed for the uptake phase, or having been treated with "pure" solvent if a solvent vehicle was used for test group diet preparation. The treated and untreated diets should be measured analytically at least in triplicate for test substance concentration before the start and at the end of the uptake phase. After exposure to the treated feed (uptake phase), fish (both groups) are fed untreated food (depuration phase).

Fish are fed at a fixed ration (dependent on species; e.g. approximately 1-2 % of wet body weight per day in the case of rainbow trout). The feeding rate should be selected such that fast growth and large increase of lipid content are avoided. The exact feeding rate set during the experiment should be recorded. Initial feeding should be based on the scheduled weight measurements of the stock population just prior to the start of the test. The amount of feed should be adjusted based on the wet weights of sampled fish at each sampling event to account for growth during the experiment. Weights and lengths of fish in the test and control tanks can be estimated from the weights and total lengths of fish used at each sampling event; do not weigh or measure the fish remaining in the test and control tanks. It is important to maintain the same set feeding rate throughout the experiment.

⁽¹⁾ Since the actual internal concentrations can only be determined after the test has been performed, an estimate of the expected internal concentration is needed (e.g. based on the expected BMF and the concentration in the food; *cf.* Equation A5.8 in Appendix 5).

Feeding should be observed to ensure that the fish are visibly consuming all of the food presented in order to guarantee that the appropriate ingestion rates are used in the calculations. Preliminary feeding experiments or previous experience should be considered when selecting a feeding rate that will ensure that all food from once-daily feeding is consumed. In the event that food is consistently being left uneaten, it may be advisable to spread the dose over an extra feeding period in each experimental day (e.g. replace once-daily feeding with feeding half the amount twice daily). If this is necessary, the second feeding should occur at a set time and be timed so that the maximum period of time possible passes before fish sampling (e.g. time for second feeding is set within the first half of an experimental day).

Although fish generally rapidly consume the food, it is important to ensure that the substance remains adsorbed to the food. Efforts should be made to avoid the test substance becoming dispersed in water from the food, thereby exposing the fish to aqueous concentrations of the test substance in addition to the dietary route. This can be achieved by removing any uneaten food (and faeces) from the test and control tanks within one hour of feeding, but preferably within 30 minutes. In addition, a system where the water is continuously cleaned over an active carbon filter to absorb any 'dissolved' contaminant may be used. Flow-through systems may help to flush away food particles and dissolved substances rapidly ⁽¹⁾. In some cases, a slightly modified spiked food preparation technique can help to alleviate this problem (see paragraph 119).

Light and Temperature

As for the aqueous exposure method (*cf.* paragraph 48), a 12 to 16 hour photoperiod is recommended and temperature (± 2 °C) appropriate for the test species used (*cf.* Appendix 3). Type and characteristics of illumination should be known and documented.

Controls

One control group should be used, with fish fed the same ration as the test group but without the test substance present in the feed. If an oil or solvent vehicle has been used to spike the feed in the test group, the control group food should be treated in exactly the same way but with the absence of test substance so that the diets of the test group and control group are equivalent (*cf.* paragraphs 121 and 139).

Frequency of Water Quality Measurements

The conditions described in the aqueous exposure method apply here also, except that TOC need only be measured before the test as part of the test water characterisation (*cf.* paragraph 53).

Sampling and Analysis of Fish and Diet

Analysis of Diet Samples

Samples of the test and control diets should be analysed at least in triplicate for the test substance and for lipid content at least before the beginning and at the end of the uptake phase. The methods of analysis and procedures for ensuring homogeneity of the diet should be included in the report.

⁽¹⁾ The presence of the test substance in the test medium as a result of excretion by the fish or leaching from food may not be totally avoidable. Therefore one option is to measure the substance concentration in water at the end of the uptake period, especially if a semi-static set up is used, to help to establish whether any aqueous exposure has occurred.

Samples should be analysed for the test substance by the established and validated method. Pre-study work should be conducted to establish the limit of quantification, percent recovery, interferences and analytical variability in the intended sample matrix. If a radiolabelled material is being tested, similar considerations as those for the aqueous exposure method should be considered with feed analysis replacing water analysis (*cf.* paragraph 65).

Analysis of Fish

At each fish sampling event, 5-10 individuals will be sampled from exposure and control treatments (in some instances numbers of control fish can be reduced; *cf.* paragraph 134).

Sampling events should occur at the same time on each experimental day (relative to feeding time), and should be timed so that the likelihood of food remaining in the gut during the uptake phase and the early part of the depuration phase is minimised to prevent spurious contributions to total test substance concentrations (*i.e.* sampled fish should be removed at the end of an experimental day, keeping in mind that an experimental day starts at the time of feeding and ends at the time of the next feeding, approximately 24 hours later. Depuration begins with the first feeding of unspiked food; *cf.* paragraph 128). The first depuration phase sample (taken shortly before the second feeding with unspiked food) is important as extrapolation back one day from this measurement is used to estimate the time zero concentration ($C_{0,d}$, the concentration in the fish at the end of uptake/start of depuration). Optionally, the gastrointestinal tract of the fish can be removed and analysed separately at the end of uptake and at days 1 and 3 of depuration.

At each sampling event fish should be removed from both test vessels and treated in the same way as described in the aqueous method (*cf.* paragraphs 61-63).

Concentrations of test substance in whole fish (wet weight) are measured at least at the end of the uptake phase and during the depuration phase in both control and test groups. During the depuration phase, four to six sampling points are recommended (*e.g.* 1, 3, 7, 14 and 28 days). Optionally, an additional sampling point may be included after 1-3 days' uptake to estimate assimilation efficiency from the linear phase of uptake for the fish while still near the beginning of the exposure period. Two main deviations from the schedule exist: i) if an extended uptake phase is employed for the purposes of investigating uptake kinetics, there will be additional sampling points during the uptake phase and so additional fish will need to be included (*cf.* paragraph 126); ii) if the study has been terminated at the end of the uptake phase owing to no measurable uptake (*cf.* paragraph 131). Individual fish that are sampled should be weighed (and their total length measured) to allow growth rate constants to be determined. Concentrations of the substance in specific fish tissue (edible and non-edible portions) can also be measured at the end of uptake and selected depuration times. If a radiolabelled material is being tested, similar considerations as those for the aqueous exposure method should be considered with feed analysis replacing water analysis (*cf.* paragraph 65).

For the periodic use of a reference substance (*cf.* paragraph 25), it is preferable that concentrations are measured in the test group at the end of uptake and at all depuration times specified for the test substance (whole fish); concentrations need only be analysed in the control group at end of uptake (whole fish). In certain circumstances (for example if analysis techniques for test substance and reference substance are incompatible, such that additional fish would be needed to follow the sampling schedule) another approach may be used as follows to minimise the number of additional fish required. Concentrations of the reference substance are measured during depuration only on days 1, 3 and two further sampling points, selected such that reliable estimations of time zero concentration ($C_{0,d}$) and k_2 can be made for the reference substance.

If possible the lipid content of the individual fish should be determined on each sampling occasion, or at least at the start and end of the uptake phase and at the end of the depuration phase. (*cf.* paragraphs 56 and 67). Depending on the analytical method (refer to paragraph 67 and to Appendix 4), it may be possible to use the same fish for both lipid content and test substance concentration determination. This is preferred on the grounds of minimising fish numbers. However, should this not be possible, the same approach as described in the aqueous exposure method can be used (see paragraph 56 for these alternative lipid measurement options). The method used to quantify the lipid content should be documented in the report.

Quality of the analytical method

Experimental checks should be conducted to ensure the specificity, accuracy, precision and reproducibility of the substance-specific analytical technique, as well as recoveries of the test substance from both food and fish.

Fish growth measurement

At the start of the test a sample of fish from the stock population need to be weighed (and their total length measured). These fish should be sampled shortly before the first spiked feeding (e.g. one hour), and assigned to experimental day 0. The number of fish for this sample should be at least the same as that for the samples during the test. Some of these can be the same fish used for lipid analysis before the start of the uptake phase (cf. paragraph 153). At each sampling interval fish are first weighed and their length measured. In each individual fish the measured weight (and length) should be linked to the analysed chemical concentration (and lipid content, if applicable), for example using a unique identifier code for each sampled fish. The measurements of these sampled fish can be used to estimate the weight (and length) of fish remaining in the test and control tanks.

Experimental Evaluation

Observations of mortality should be performed and recorded daily. Additional observations for adverse effects should be performed, for example for abnormal behaviour or pigmentation, and recorded. Fish are considered dead if there is no respiratory movement and no reaction to a slight mechanical stimulus can be detected. Any dead or clearly moribund fish should be removed.

DATA AND REPORTING

Treatment of results

Test results are used to derive the depuration rate constant (k_d) as a function of the total wet weight of the fish. Growth rate constant, k_g , based on mean increase in fish weight is calculated and used to produce the growth-corrected depuration rate constant, k_{dg} , if appropriate. In addition, the assimilation efficiency (a ; absorption from the gut), the kinetic biomagnification factor (BMF_k) (if necessary growth corrected, BMF_{kg}), its lipid-corrected value (BMF_{KL} or BMF_{kgL} , if corrected for growth dilution) and feeding rate should be reported. Also, if an estimate of the time to steady-state in the uptake phase can be made (e.g. 95 % of steady-state or $t_{95} = 3.0/k_d$), an estimate of the steady-state BMF (BMF_{ss}) can be included (cf. paragraphs 105 and 106, and Appendix 5) if the t_{95} value indicates that steady-state conditions may have been reached. The same lipid correction should be applied to this BMF_{ss} as to the kinetically-derived BMF (BMF_k) to give a lipid-corrected value, BMF_{ssL} (note that no agreed procedure is available to correct a steady-state BMF for growth dilution). Formulae and example calculations are presented in Appendix 7. Approaches are available that make it feasible to estimate a kinetic bioconcentration factor (BCF_k) from data generated in the dietary study. This is discussed in Appendix 8.

Fish weight/length data

Individual fish wet weights and lengths for all time periods are tabulated separately for test and control groups for all sampling days during the uptake phase (stock population for start of uptake; control group and test group for end of uptake and, if conducted, the early phase (e.g. day 1-3 of uptake) and depuration phase (e.g. days 1, 2, 4, 7, 14, 28, for control and test group). Weight is the preferred measure of growth for growth dilution correction purposes. See below (paragraphs 162 and 163) and Appendix 5 for the method(s) used to correct data for growth dilution.

Test substance concentration in fish data

Individual fish test substance residue measurements (or pooled fish samples if individual fish measurements are not possible), expressed in terms of wet weight concentration (w/w), are tabulated for test and control fish for individual sample times. If lipid analysis has been conducted on each sampled fish then individual lipid-corrected concentrations, in terms of lipid concentration (w/w lipid), can be derived and tabulated.

- Test substance residue measurements in individual fish (or pooled fish samples if individual fish measurements are not possible, *cf.* paragraph 66) for the depuration period are converted to their natural logarithms and plotted versus time (day). If a visual inspection of the plot shows obvious outliers, a statistically valid outlier test may be applied to remove spurious data points as well as documented justification for their omission.
- A linear least squares correlation is calculated for the $\ln(\text{concentration})$ vs. depuration (day) data. The slope and intercept of the line are reported as the overall depuration rate constant (k_2) and natural logarithm of the derived time zero concentration ($C_{0,d}$) (*cf.* Appendix 5 and Appendix 7 for further details). Should this not be possible because concentrations fall below the limit of quantification for the second depuration sample, a conservative estimate of k_2 can be made (*cf.* Appendix 7).
- The variances in the slope and intercept of the line are calculated using standard statistical procedures and the 90 % (or 95 %) confidence intervals around these results evaluated and presented.
- The mean measured fish concentration for the final day of uptake (measured time zero concentration, $C_{0,m}$) is also calculated and compared with the derived value $C_{0,d}$. In case the derived value is lower than the measured value, the difference may suggest the presence of undigested spiked food in the gut. If the derived value is very much higher than the measured value, this may be an indication that the value derived from the depuration data linear regression is erroneous and should be re-evaluated (see Appendix 7).

Depuration rate and biomagnification factor

To calculate the biomagnification factor from the data, first the assimilation efficiency (absorption of test substance across the gut, a) should be obtained. To do this, equation A7.1 in Appendix 7 should be used, requiring the derived concentration in fish at time zero of the depuration phase ($C_{0,d}$), (overall) depuration rate constant (k_2), concentration in the food (C_{food}), food ingestion rate constant (I) and duration of the uptake period (t) to be known. The slope and intercept of the linear relationship between $\ln(\text{concentration})$ and depuration time are reported as the overall depuration rate constant ($k_2 = \text{slope}$) and time zero concentration ($C_{0,d} = e^{\text{intercept}}$), as above. The derived values should be checked for biological plausibility (e.g. assimilation efficiency as a fraction is not greater than 1). (I) is calculated by dividing the mass of food by the mass of fish fed each day (if fed at 2 % of body weight, (I) will be 0,02). However, the feeding rate used in the calculation may need to be adjusted for fish growth (this can be done using the known growth rate constant to estimate the fish weight at each time-point during the uptake phase; *cf.* Appendix 7). In cases where k_2 and $C_{0,d}$ cannot be derived because, for example, concentrations fell below the limit of detection for the second depuration sample, a conservative estimate of k_2 and an “upper bound” BMF_k can be made (*cf.* Appendix 7).

Once the assimilation efficiency (a) is obtained, the biomagnification factor can be calculated by multiplying a by the ingestion rate constant (I) and dividing by the (overall) depuration rate constant (k_2). The growth-corrected biomagnification factor is calculated in the same way but using the growth-corrected depuration rate constant (k_{2g} ; *cf.* paragraphs 162 and 163. An alternative estimate of the assimilation efficiency can be derived if tissue analysis was performed on fish sampled in the early, linear phase of the uptake phase; *cf.* paragraph 151 and Appendix 7. This value represents an independent estimate of assimilation efficiency for an essentially unexposed organism (*i.e.* the fish are near the beginning of the uptake phase). The assimilation efficiency estimated from depuration data is usually used to derive the BMF.

Lipid Correction and Growth-Dilution Correction

Fish growth during the depuration phase can lower measured chemical concentrations in the fish with the effect that the overall depuration rate constant, k_2 , is greater than would arise from removal processes (e.g. metabolism, egestion) alone (cf. paragraph 72). Lipid contents of test fish (which are strongly associated with the bioaccumulation of hydrophobic substances) and lipid contents of food can vary enough in practice such that their correction is necessary to present biomagnification factors in a meaningful way. The biomagnification factor should be corrected for growth dilution (as is the kinetic BCF in the aqueous exposure method) and corrected for the lipid content of the food relative to that of the fish (the lipid-correction factor). Equations and examples for these calculations can be found in Appendix 5 and Appendix 7, respectively.

To correct for growth dilution, the growth-corrected depuration rate constant (k_{2g}) should be calculated (see Appendix 5 for equations). This growth-corrected depuration rate constant (k_{2g}) is then used to calculate the growth-corrected biomagnification factor, as in paragraph 73. In some cases this approach is not possible. An alternative approach that circumvents the need for growth dilution correction involves using mass of test substance per fish (whole fish basis) depuration data rather than the usual mass of test substance per unit mass of fish (concentration) data. This can be easily achieved as tests according to this method should link recorded tissue concentrations to individual fish weights. The simple procedure for doing this is outlined in Appendix 5. Note that k_2 should still be estimated and reported even if this alternative approach is used.

To correct for the lipid content of the food and fish when lipid analysis has not been conducted on all sampled fish, the mean lipid fractions (w/w) in the fish and in the food are derived (¹). The lipid correction factor (L_c) is then calculated by dividing the fish mean lipid fraction by the mean food lipid fraction. The biomagnification factor, growth corrected or not as applicable, is divided by the lipid correction factor to calculate the lipid-corrected biomagnification factor.

If chemical and lipid analyses were conducted on the same fish at each sampling point, then the lipid-corrected tissue data for individual fish may be used to calculate a lipid-corrected BMF directly (cf. (37)). The plot of lipid-corrected concentration data gives $C_{0,d}$ on a lipid basis and k_2 . Mathematical analysis can then proceed using the same equations in Appendix 7, but assimilation efficiency (a) is calculated using the lipid-normalised food ingestion rate constant (I_{lipid}) and the dietary concentration on a lipid basis ($C_{food-lipid}$). Lipid corrected parameters are similarly then used to calculate BMF (note that growth rate constant correction should also be applied to the lipid fraction rather than the fish wet weight to calculate the lipid-corrected, growth corrected BMF_{kgL}).

Interpretation of results

Average growth in both test and control groups should in principle not be significantly different to exclude toxic effects. The growth rate constants or the growth curves of the two groups should be compared by an appropriate procedure (²).

Test report

After termination of the study, a final report is prepared containing the information on *Test Substance*, *Test Species* and *Test Conditions* as listed in paragraph 81 (as for the aqueous exposure method). In addition, the following information is required:

(¹) This approach is specific to the dietary study, distinct from the procedure followed in the aqueous exposure, hence the word “correction” has been used rather than “normalisation” to prevent confusion — see also footnote in paragraph 106.

(²) A t-test on growth rate constants can be performed, to test whether growth differs between control and test groups, or an F-test in case of analysis of variance. If needed, an F-test or likelihood ratio test can be used to assist in the choice of the appropriate growth model (OECD monograph 54, (32)).

Test Substance:

- Any information on stability of the test substance in prepared food;

Test Conditions:

- Substance nominal concentration in food, spiking technique, amount of (lipid) vehicle used in food spiking process (if used), test substance concentration measurements in spiked diet for each analysis (at least in triplicate before study start and at end of uptake) and mean values;
- If used, type and quality of carrier oil or solvent (grade, supplier, etc.) used for food spiking;
- Food type employed (proximate analysis ⁽¹⁾, grade or quality, supplier, etc.), feeding rate during uptake phase, amount of food administered and frequency (including any adjustments based on sampled fish weight);
- Time at which fish were collected and euthanised for chemical analysis for each sample point (e.g. one hour before the following day's feeding);

Results:

- Results from any preliminary study work;
- Information on any adverse effects observed;
- Complete description of all chemical analysis procedures employed including limits of detection and quantification, variability and recovery;
- Measured lipid concentrations in food (spiked and control diet), individual, mean values and standard deviations;
- Tabulated fish weight (and length) data linked to individual fish, both for control and exposure groups (for example using unique identifiers for each fish) and calculations, derived growth rate constant(s) and 95 % confidence interval(s);
- Tabulated test substance concentration data in fish, mean measured concentration at end of uptake ($C_{0,m}$), and derived (overall) depuration rate constant (k_2) and concentration in fish at start of depuration phase ($C_{0,d}$) together with the variances in these values (slope and intercept);
- Tabulated fish lipid contents data (listed against specific substance concentrations if applicable), mean values for test group and control at test start, end of uptake and end of depuration;
- Curves (including all measured data), showing the following (if applicable, concentrations may be expressed in relation to the whole body of the animal or specified tissues thereof):
 - growth (*i.e.* fish weight (and length) vs. time) or natural logarithm transformed weight vs. time;
 - the depuration of the test substance in the fish; and
 - natural logarithm transformed concentration (\ln concentration) vs. depuration time (including the derived depuration rate constant k_2 , and natural logarithm derived concentration in fish at start of depuration phase, $C_{0,d}$).
- If a visual inspection of a plot shows obvious outliers, a statistically valid outlier test may be applied to remove spurious data points as well as documented justification for their omission.

⁽¹⁾ A foodstuff analysis technique for protein, lipid, crude fibre and ash content; this information is usually available from the feed supplier.

- Calculated growth-corrected depuration rate constant and growth-corrected half-life.
- Calculated assimilation efficiency (α).
- “Raw” dietary BMF, lipid and growth-dilution corrected kinetic BMF (“raw” and lipid-corrected based on whole fish wet weight), tissue-specific BMF if applicable.
- Any information concerning radiolabelled test substance metabolites and their accumulation.
- Anything unusual about the test, any deviation from these procedures, and any other relevant information.
- A summary table of relevant measured and calculated data, as hereafter:

Substance Depuration Rate constants and Biomagnification Factors (BMF _k)	
k_g (growth rate constant; day ⁻¹):	Insert Value (95 % CI) ⁽¹⁾
k_2 (overall depuration rate constant, day ⁻¹):	Insert Value (95 % CI)
k_{2g} (growth-corrected depuration rate constant; day ⁻¹):	Insert Value (95 % CI) ⁽¹⁾
$C_{0,m}$ (measured time zero concentration, the concentration in fish at end of uptake) (µg/g):	Insert Value ± SD ⁽²⁾
$C_{0,d}$ (derived time zero concentration of depuration phase; µg/g):	Insert Value ± SD ⁽²⁾
I (set feed ingestion rate; g food/g fish/day):	Insert Value
I_g (effective feeding rate, adjusted for growth; g food/g fish/day) ⁽²⁾ :	Insert Value ± SD ⁽²⁾
C_{food} (chemical concentration in the food; µg/g):	Insert Value ± SD ⁽²⁾
α (substance assimilation efficiency):	Insert Value ± SD ⁽²⁾
BMF _k (kinetic dietary BMF):	Insert Value (95 % CI) ⁽¹⁾
BMF _{kg} (growth-corrected kinetic dietary BMF):	Insert Value (95 % CI) ⁽¹⁾
$t_{1/2g}$ (growth-corrected half-life in days):	Insert Value ± SD ⁽²⁾
L_c (lipid correction factor):	Insert Value
BMF _{kgL} (lipid-corrected growth-corrected kinetic BMF):	Insert Value
BMF _{ss-L} (indicative lipid-corrected steady-state BMF) ⁽²⁾ :	Insert Value ± SD ⁽²⁾

⁽¹⁾ CI: confidence interval (where possible to estimate).

⁽²⁾ SD: Standard deviation (where possible to estimate).

LITERATURE:

- (1) Chapter C.13 of this Annex, *Bioconcentration: Flow-through Fish Test*.
- (2) Chapter A.6 of this Annex, *Water Solubility*
- (3) Li A, Doucette W.J. (1993), The effect of cosolutes on the aqueous solubilities and octanol/water partition coefficients of selected polychlorinated biphenyl congeners. *Environ Toxicol Chem* 12: 2031-2035
- (4) Chapter A.8 of this Annex, *Partition Coefficient (n-octanol/water): Shake Flask Method*.
- (5) Chapter A.24 of this Annex, *Partition Coefficient (n-octanol/water), HPLC Method*.
- (6) Chapter A.23 of this Annex, *Partition Coefficient (1-Octanol/Water): Slow-Stirring Method*.
- (7) Chapter C.7 of this Annex, *Hydrolysis as a Function of pH*.
- (8) OECD (1997), OECD Environmental Health and Safety Publications Series on Testing and Assessment Number 7: Guidance Document on Direct Phototransformation of Chemicals in Water OCDE/GD(97)21. Organisation for Economic Co-operation and Development (OECD), Paris, France.
- (9) Chapter A.5 of this Annex, *Surface Tension of Aqueous Solutions*.
- (10) Chapter A.4 of this Annex, *Vapour Pressure*.
- (11) Chapter C.4 of this Annex, *Ready Biodegradability*.
- (12) Chapter C.29 of this Annex, *Ready Biodegradability — CO₂ in sealed vessels*
- (13) Connell D.W. (1988), Bioaccumulation behaviour of persistent chemicals with aquatic organisms. *Rev. Environ. Contam. Toxicol.* 102: 117-156.
- (14) Bintein S., Devillers J. and Karcher W. (1993), Nonlinear dependence of fish bioconcentration on *n*-octanol/water partition coefficient. *SAR QSAR Environ. Res.* 1: 29-39.
- (15) OECD (2011), QSAR Toolbox 2.1. February 2011. Available from: http://www.oecd.org/document/54/0,3746,en_2649_34379_42923638_1_1_1_1,00.html.
- (16) Brown R.S., Akhtar P., Åkerman J., Hampel L., Kozin I.S., Villerius L.A. and Klammer H.J.C. (2001), Partition controlled delivery of hydrophobic substances in toxicity tests using poly(dimethylsiloxane) (PDMS) films. *Environ. Sci. Technol.* 35: 4097-4102.
- (17) Fernandez J.D., Denny J.S. and Tietge J.E. (1998), A simple apparatus for administering 2,3,7,8-tetrachloro-dibenzo-*p*-dioxin to commercially available pelleted fish food. *Environ. Toxicol. Chem.* 17: 2058-2062.
- (18) Nichols J.W., Fitzsimmons P.N., Whiteman F.W. and Dawson T.D. (2004), A physiologically based toxicokinetic model for dietary uptake of hydrophobic organic compounds by fish: I. Feeding studies with 2,2', 5,5'-tetrachlorobiphenyl. *Toxicol. Sci.* 77: 206-218.
- (19) Parkerton T.F., Arnot J.A., Weisbrod A.V., Russom C., Hoke R.A., Woodburn K., Traas T., Bonnell M., Burkhard L.P. and Lampi M.A. (2008), Guidance for evaluating *in vivo* fish bioaccumulation data. *Integr. Environ. Assess. Manag.* 4: 139-155.
- (20) Verbruggen E.M.J., Beek M., Pijnenburg J. and Traas T.P. (2008). Ecotoxicological environmental risk limits for total petroleum hydrocarbons on the basis of internal lipid concentrations. *Environ. Toxicol. Chem.* 27: 2436-2448.
- (21) Schlechtriem C., Fliedner A. and Schäfers C. (2012), Determination of lipid content in fish samples from bioaccumulation studies: Contributions to the revision of OECD Test Guideline 305. *Environmental Sciences Europe* 2012, 24:13. published: 3 April 2012.

- (22) Chapter C.47 of this Annex, *Fish, Early-Life Stage Toxicity Test*.
- (23) Chapter C.15 of this Annex, *Fish, Short-term Toxicity Test on Embryo and Sac-Fry Stages*.
- (24) Chapter C.14 of this Annex, *Fish, Juvenile Growth Test*.
- (25) OECD (2000), OECD Environmental Health and Safety Publications Series on Testing and Assessment No. 23: Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures ENV/JM/MONO(2000)6. Organisation for Economic Co-operation and Development (OECD), Paris, France.
- (26) US-EPA (1994), Great Lake water quality initiative technical support document for the procedure to determine bioaccumulation factors 822-R-94-002. US EPA, Office of Water, Office of Science and Technology, Washington, DC, USA.
- (27) US-FDA (1999), Pesticide analytical manual (PAM). Vol.1. US Food and Drug Administration, Rockville, MD, USA.
- (28) US-EPA (1974), Section 5, A (1) Analysis of Human or Animal Adipose Tissue, in Analysis of Pesticide Residues in Human and Environmental Samples, Thompson, J.F., Editor. US-EPA, Research Triangle Park, NC, USA
- (29) Bligh E.G. and Dyer W.J. (1959), A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37: 911-917.
- (30) Gardner W.S., Frez W.A., Cichocki E.A. and Parrish C.C. (1985), Micromethod for lipids in aquatic invertebrates. *Limnol. Oceanogr.* 30: 1099-1105.
- (31) Smedes F. (1999), Determination of total lipid using non-chlorinated solvents. *Analyst.* 124: 1711-1718.
- (32) OECD (2006), OECD Environmental Health and Safety Publications Series on Testing and Assessment No. 54: Current approaches in the statistical analysis of ecotoxicity data: a guidance to application. ENV/JM/MONO(2006)18. Organisation for Economic Co-operation and Development (OECD), Paris, France.
- (33) Springer T.A., Guiney P.D., Krueger H.O. and Jaber M.J. (2008), Assessment of an approach to estimating aquatic bioconcentration factors using reduced sampling. *Environ. Toxicol. Chem.* 27: 2271-2280.
- (34) Springer T.A. (2009), Statistical Research Related to Update of OECD Guideline 305. Wildlife International, Ltd, Easton, MD, USA.
- (35) Arnot J.A., Meylan W., Tunkel J., Howard P.H., Mackay D., Bonnell M. and Boethling R.S. (2009), A quantitative structure-activity relationship for predicting metabolic biotransformation rates for organic chemicals in fish. *Environ. Toxicol. Chem.* 28: 1168-1177.
- (36) Parkerton T., Letinski D., Febbo E., Davi R., Dzambia C., Connelly M., Christensen K. and Peterson D. (2001), A practical testing approach for assessing bioaccumulation potential of poorly water soluble organic chemicals (presentation). in SETAC Europe 12th Annual Meeting: Madrid, Spain.
- (37) Fisk A.T., Cymbalisky C.D., Bergman Å. and Muir D.C.G. (1996), Dietary accumulation of C₁₂- and C₁₆-chlorinated alkanes by juvenile rainbow trout (*Oncorhynchus mykiss*). *Environ. Toxicol. Chem.* 15: 1775-1782.
- (38) Anonymous (2004), Fish, dietary bioaccumulation study — Basic protocol, document submitted to the TC-NES WG on PBT.
- (39) Anonymous (2004), Background document to the fish dietary study protocol, document submitted to the TC-NES WG on PBT.
- (40) Bruggeman W.A., Opperhuizen A., Wijnnga A. and Hutzinger O. (1984), Bioaccumulation of superlipophilic chemicals in fish, *Toxicol. Environ. Chem.* 7: 173-189.
- (41) Muir D.C.G., Marshall W.K. and Webster G.R.B. (1985), Bioconcentration of PCDDs by fish: effects of molecular structure and water chemistry. *Chemosphere.* 14: 829-833.

- (42) Thomann R.V. (1989), Bioaccumulation model of organic chemical distribution in aquatic food chains. Environ. Sci. Technol. 23: 699-707.
 - (43) Nichols J.W., Fitzsimmons P.N. and Whiteman F.W. (2004), A physiologically based toxicokinetic model for dietary uptake of hydrophobic organic compounds by fish: II. Stimulation of chronic exposure scenarios. Toxicol. Sci. 77: 219-229.
 - (44) Gobas F.A.P.C., de Wolf W., Burkhard L.P., Verbruggen E. and Plotzke K. (2009), Revisiting bioaccumulation criteria for POPs and PBT assessments. Integr. Environ. Assess. Manag. 5: 624-637.
 - (45) Sijm D.T.H.M. and van der Linde A. (1995), Size-dependent bioconcentration kinetics of hydrophobic organic chemicals in fish based on diffusive mass transfer and allometric relationships. Environ. Sci. Technol. 29: 2769-2777.
 - (46) Sijm D.T.H.M., Verberne M.E., de Jonge W.J., Pärt P. and Opperhuizen A. (1995), Allometry in the uptake of hydrophobic chemicals determined *in vivo* and in isolated perfused gills. Toxicol. Appl. Pharmacol. 131: 130-135.
 - (47) Fisk A.T., Norstrom R.J., Cymbalisty C.D. and Muir D.G.G. (1998), Dietary accumulation and depuration of hydrophobic organochlorines: Bioaccumulation parameters and their relationship with the octanol/water partition coefficient. Environ. Toxicol. Chem. 17: 951-961.
 - (48) McGrath J.A., Parkerton T.F. and Di Toro D.M. (2004), Application of the narcosis target lipid model to algal toxicity and deriving predicted-no-effect concentrations. Environ. Toxicol. Chem. 23: 2503-2517.
 - (49) Poppendieck D.G. (2002), Polycyclic Aromatic Hydrocarbon Desorption Mechanisms from Manufactured Gas Plant Site Samples. Dissertation. Department of Civil, Architectural and Environmental Engineering, University of Texas, Austin, TX, USA.
 - (50) McCarty L.S. and Mackay D. (1993), Enhancing ecotoxicological modelling and assessment: body residues and modes of toxic action. Environ. Sci. Technol. 27: 1718-1728.
 - (51) OECD (2012), OECD Environmental Health and Safety Publications Series on Testing and Assessment No. 175: Part I — Validation Report of a ring test for the OECD TG 305 dietary exposure bioaccumulation fish test. Part II — Additional Report including comparative analysis of trout and carp results ENV/JM/MONO (2012)20. Organisation for Economic Co-operation and Development (OECD), Paris, France.
-

Appendix 1

DEFINITIONS AND UNITS:

The assimilation efficiency (α) is a measure of the relative amount of substance absorbed from the gut into the organism (α is unitless, but it is often expressed as a percentage rather than a fraction).

Bioaccumulation is generally referred to as a process in which the substance concentration in an organism achieves a level that exceeds that in the respiratory medium (e.g. water for a fish or air for a mammal), the diet, or both (1).

Bioconcentration is the increase in concentration of the test substance in or on an organism (or specified tissues thereof) relative to the concentration of test substance in the surrounding medium.

The bioconcentration factor (BCF or K_b) at any time during the uptake phase of this accumulation test is the concentration of test substance in/on the fish or specified tissues thereof (C_f as mg/kg) divided by the concentration of the substance in the surrounding medium (C_w as mg/l). BCF is expressed in $\text{l}\cdot\text{kg}^{-1}$. Please note that corrections for growth and/or a standard lipid content are not accounted for.

Biomagnification is the increase in concentration of the test substance in or on an organism (or specified tissues thereof) relative to the concentration of test substance in the food.

The biomagnification factor (BMF) is the concentration of a substance in a predator relative to the concentration in the predator's prey (or food) at steady-state. In the method described in this test method, exposure via the aqueous phase is carefully avoided and thus a BMF value from this test method cannot directly be compared to a BMF value from a field study (in which both water and dietary exposure may be combined).

The dietary biomagnification factor (dietary BMF) is the term used in this test method to describe the result of dietary exposure test, in which exposure via the aqueous phase is carefully avoided and thus the dietary BMF from this test method cannot directly be compared to a BMF value from a field study (in which both water and dietary exposure may be combined).

The depuration or post-exposure (loss) phase is the time, following the transfer of the test fish from a medium containing test substance to a medium free of that substance, during which the depuration (or the net loss) of the substance from the test fish (or specified tissue thereof) is studied.

The depuration (loss) rate constant (k_2) is the numerical value defining the rate of reduction in the concentration of the test substance in the test fish (or specified tissues thereof) following the transfer of the test fish from a medium containing the test substance to a medium free of that substance (k_2 is expressed in day^{-1}).

Dissolved organic carbon (DOC) is a measure of the concentration of carbon originating from dissolved organic sources in the test media.

The exposure or uptake phase is the time during which the fish are exposed to the test substance.

The food ingestion rate (I) is the average amount of food eaten by each fish each day, relative to the estimated average fish whole body weight (expressed in terms of g food/g fish/day).

The kinetic bioconcentration factor (BCF_k) is the ratio of the uptake rate constant, k_1 , to the depuration rate constant, k_2 (i.e. k_1/k_2 — see corresponding definitions in this Appendix). In principle the value should be comparable to the BCF_{ss} (see definition above), but deviations may occur if steady-state was uncertain or if corrections for growth have been applied to the kinetic BCF.

The lipid normalised kinetic bioconcentration factor (BCF_{kl}) is normalised to a fish with a 5 % lipid content.

The lipid normalised, growth corrected kinetic bioconcentration factor (BCF_{kgl}) is normalised to a fish with a 5 % lipid content and corrected for growth during the study period as described in Appendix 5.

The lipid normalised steady-state bioconcentration factor (BCF_{ssl}) is normalised to a fish with 5 % lipid content.

A multi-constituent substance is defined for the purpose of REACH as a substance which has more than one main constituent present in a concentration between 10 % and 80 % (w/w).

The octanol-water partition coefficient (K_{ow}) is the ratio of a substance's solubility in *n*-octanol and water at equilibrium (Methods A.8 (2), A.24 (3), A.23 (4)); also expressed as P_{ow} . The logarithm of K_{ow} is used as an indication of a substance's potential for bioconcentration by aquatic organisms.

Particulate organic carbon (POC) is a measure of the concentration of carbon originating from suspended organic sources in the test media.

Solid-phase microextraction (SPME) is a solvent-free analytical technique developed for dilute systems. In this method, a polymer coated fibre is exposed to the gas or liquid phase containing the analyte of interest. Generally, a minimum analysis time is imposed so that equilibrium conditions are established between the solid and fluid phases, with respect to the measured species. Subsequently the concentration of the analyte of interest can be determined directly from the fibre or after extracting it from the fibre into a solvent, depending on the determination technique.

A steady-state is reached in the plot of test substance in fish (C_f) against time when the curve becomes parallel to the time axis and three successive analyses of C_f made on samples taken at intervals of at least two days are within ± 20 % of each other, and there is no significant increase of C_f in time between the first and last successive analysis. When pooled samples are analysed at least four successive analyses are required. For test substances which are taken up slowly the intervals would more appropriately be seven days.

The steady-state bioconcentration factor (BCF_{ss}) does not change significantly over a prolonged period of time, the concentration of the test substance in the surrounding medium being constant during this period of time (cf. Definition of steady-state).

Total organic carbon (TOC) is a measure of the concentration of carbon originating from all organic sources in the test media, including particulate and dissolved sources.

The uptake rate constant (k_1) is the numerical value defining the rate of increase in the concentration of test substance in/on test fish (or specified tissues thereof) when the fish are exposed to that substance (k_1 is expressed in $l\ kg^{-1}\ day^{-1}$).

Substances of Unknown or Variable composition, Complex reaction products and Biological materials are known as UVCB

Chemical is a substance or a mixture.

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

LITERATURE:

- (1) Gobas F.A.P.C., de Wolf W., Burkhard L.P., Verbruggen E. and Plotzke K. (2009), Revisiting bioaccumulation criteria for POPs and PBT assessments. Integr. Environ. Assess. Manag. 5: 624-637.

- (2) Chapter A.8 of this Annex, *Partition Coefficient (n-octanol/water): Shake Flask Method*
 - (3) Chapter A.24 of this Annex, *Partition Coefficient (n-octanol/water), HPLC Method*.
 - (4) Chapter A.23 of this Annex, *Partition Coefficient (1-Octanol/Water): Slow-Stirring Method*.
-

Appendix 2

SOME CHEMICAL CHARACTERISTICS OF AN ACCEPTABLE DILUTION WATER

Component	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

FISH SPECIES RECOMMENDED FOR TESTING

Recommended species	Recommended range of test temperature (°C)	Recommended total length of test animal (cm) ⁽²⁾
<i>Danio rerio</i> ⁽¹⁾ (Teleostei, Cyprinidae) (Hamilton-Buchanan) Zebra-fish	20 – 25	3,0 ± 0,5
<i>Pimephales promelas</i> (Teleostei, Cyprinidae) (Rafinesque) Fathead minnow	20 – 25	5,0 ± 2,0
<i>Cyprinus carpio</i> (Teleostei, Cyprinidae) (Linnaeus) Common carp	20 – 25	8,0 ± 4,0 ⁽³⁾
<i>Oryzias latipes</i> (Teleostei, Poeciliidae) (Temminck and Schlegel) Ricefish	20 – 25	4,0 ± 1,0
<i>Poecilia reticulata</i> (Teleostei, Poeciliidae) (Peters) Guppy	20 – 25	3,0 ± 1,0
<i>Lepomis macrochirus</i> (Teleostei Centrarchidae) (Rafinesque) Bluegill	20 – 25	5,0 ± 2,0
<i>Oncorhynchus mykiss</i> (Teleostei Salmonidae) (Walbaum) Rainbow trout	13 – 17	8,0 ± 4,0
<i>Gasterosteus aculeatus</i> (Teleostei, (Gasterosteidae) (Linnaeus) Three-spined stickleback	18 — 20	3,0 ± 1,0

⁽¹⁾ Meyer *et al.* (1)

⁽²⁾ It should be noted that in the test itself weight is preferred as the measure for size and growth rate constant derivations. It is however recognised that length is a more practical measure if fish have to be selected by sight at the beginning of an experiment (i.e. from the stock population).

⁽³⁾ This length range is indicated in the Testing Methods for New Chemical Substances etc. based on the Japan's Chemical Substances Control Law (CSCL).

Various estuarine and marine species have less widely been used, for example:

Spot	(<i>Leiostomus xanthurus</i>)
Sheepshead minnow	(<i>Cyprinodon variegatus</i>)
Silverside	(<i>Menidia beryllina</i>)
Shiner perch	(<i>Cymatogaster aggregata</i>)
English sole	(<i>Parophrys vetulus</i>)
Staghorn sculpin	(<i>Leptocottus armatus</i>)
Three-spined stickleback	(<i>Gasterosteus aculeatus</i>)
Sea bass	(<i>Dicentrarchus labrax</i>)
Bleak	(<i>Alburnus alburnus</i>)

The freshwater fish listed in the table above are easy to rear and/or are widely available throughout the year, whereas the availability of marine and estuarine species is partially confined to the respective countries. They are capable of being bred and cultivated either in fish farms or in the laboratory, under disease- and parasite-controlled conditions, so that the test animal will be healthy and of known parentage. These fish are available in many parts of the world.

LITERATURE:

- (1) Meyer A., Biermann C.H. and Orti G. (1993), The phylogenetic position of the zebrafish (*Danio rerio*), a model system in developmental biology: An invitation to the comparative method Proc. R. Soc. Lond. B. 252: 231-236.
-

Appendix 4

SAMPLING SCHEDULES FOR AQUEOUS AND DIETARY EXPOSURE TESTS

1. Theoretical example of a sampling schedule for a full aqueous exposure bioconcentration test of a substance with $\log K_{ow} = 4$.

Fish Sampling	Sample time schedule		No. of water samples (¹)	No. of fish per sample (¹)
	Minimal required frequency (days) (²)	Additional sampling (days) (²)		
Uptake phase				
1	– 1		2 (³)	4 (⁴)
	0		(2)	(3 (⁶))
2	0,3		2	4
		0,4	(2)	(4)
3	0,6		2	4
		0,9	(2)	(4)
4	1,2		2	4
		1,7	(2)	(4)
5	2,4		2	4
		3,3	(2)	(4)
6	4,7		2	4 – 8 (⁵)
				(3 (⁶))
Depuration phase				Transfer fish to water free of test substance
7	5,0		2	4
		5,3		(4)
8	5,9		2	4
		7,0		(4)
9	9,3		2	4
		11,2		(4)

Fish Sampling	Sample time schedule		No. of water samples (¹)	No. of fish per sample (¹)
	Minimal required frequency (days) (²)	Additional sampling (days) (²)		
Uptake phase				
10	14,0		2	4 – 8 (⁵)
		17,5		(4+3 (⁶))
TOTAL				40 – 72 (48 – 80) (⁵)

(¹) Values in brackets are numbers of samples (water, fish) to be taken if additional sampling is carried out.

(²) Pre-test estimate of k_2 for log K_{OW} of 4,0 is 0,652 days⁻¹. The total duration of the experiment is set to $3 \times t_{ss} = 3 \times 4,6$ days, i.e. 14 days. For the estimation of t_{ss} refer to Appendix 5.

(³) Sample water after a minimum of 3 “chamber-volumes” has been delivered.

(⁴) These fish are sampled from the stock population.

(⁵) If greater precision or metabolism studies are necessary that require more fish, these should be sampled particularly at the end of the uptake and depuration phases (cf. paragraph 40).

(⁶) At least 3 additional fish may be required for lipid content analysis if it is not possible to use the same fish sampled for substance concentrations at the start of the test, the end of the uptake phase and the end of the depuration phase. Note it should be possible in many cases to use the 3 control fish alone (cf. paragraph 56).

2. Theoretical example of sampling schedule for dietary bioaccumulation test of substance following 10 day uptake and 42 day depuration phases.

Sampling event	Sample time schedule		No. food samples	No. fish per sample	
	Day of phase	Additional fish samples?		Test Group	Control Group
Uptake phase					
1	0	Possible (¹) (²)	3 — test group	0	5 – 10
			3 — control group (¹)		(8 – 13) (²)
1A (³)	1-3			5 – 10	5 – 10
2	10	Yes (⁴)	3 — test group	10 – 15 (⁴)	5 – 10
			3 — control group (¹)	(13 – 18) (⁵)	(8 – 13) (⁵)
Depuration phase					
3	1	Yes (⁴)		10 – 15 (⁴)	5 – 10
4	2			5 – 10	5 – 10
5	4			5 – 10	5 – 10

Sampling event	Sample time schedule		No. food samples	No. fish per sample	
	Day of phase	Additional fish samples?		Test Group	Control Group
Uptake phase					
6	7	Yes ⁽⁴⁾		10 – 15 ⁽⁴⁾	5 – 10
7	14			5 – 10	5 – 10
8	28			5 – 10	5 – 10
9	42	Yes ⁽⁴⁾		10 – 15 ⁽⁴⁾ (13 – 18) ⁽⁵⁾	5 – 10 (8 – 13) ⁽⁵⁾
TOTAL				59 – 120 (63 – 126) ⁽⁴⁾ ⁽⁵⁾	50 – 110 (56 – 116) ⁽⁴⁾ ⁽⁵⁾

⁽¹⁾ 3 samples of feed from both control and test groups analysed for test substance concentrations and for lipid content.

⁽²⁾ Fish are sampled from the stock population as near to the start of the study as possible; at least 3 fish from the stock population at test start should be sampled for lipid content.

⁽³⁾ (Optional) sampling early in the uptake phase provides data to calculate dietary assimilation of test substance that can be compared with the assimilation efficiency calculated from the depuration phase data.

⁽⁴⁾ 5 extra fish may be sampled for tissue-specific analysis.

⁽⁵⁾ At least 3 additional fish may be required for lipid content analysis if it is not possible to use the same fish sampled for substance concentrations at the start of the test, the end of the uptake phase and the end of the depuration phase. Note it should be possible in many cases to use the 3 control fish alone (*cf.* paragraphs 56 and 153).

Note on phase and sampling timings: The uptake phase begins with the first feeding of spiked diet. An experimental day runs from one feeding until shortly before the next, 24 hours later. The first sampling event (1 in the table) should be taken shortly before the first feeding (e.g. one hour). Sampling during a study should ideally be carried out shortly before the following day's feeding (*i.e.* about 23 hours after the sample day's feeding). The uptake phase ends shortly before the first feeding with unspiked diet, when the depuration phase begins (test group fish are likely to be still digesting spiked feed in the intervening 24 hours after the last spiked diet feeding). This means that the end of uptake sample should be taken shortly before the first feeding with unspiked diet and the first depuration phase sample should be taken about 23 hours after the first feeding with unspiked feed.

Appendix 5

GENERAL CALCULATIONS

1. Introduction
2. Prediction of the duration of the uptake phase
3. Prediction of the duration of the depuration phase
4. Sequential method: determination of depuration (loss) rate constant k_2
5. Sequential method: determination of uptake rate constant k_1 (aqueous exposure method only)
6. Simultaneous method for calculation of uptake and depuration (loss) rate constants (aqueous exposure method only)
7. Growth dilution correction for kinetic BCF and BMF
8. Lipid normalisation to 5 % lipid content (aqueous exposure method only)

1. INTRODUCTION

The general fish aquatic bioaccumulation model can be described in terms of uptake and loss processes, ignoring uptake with food. The differential equation (dC_f/dt) describing the rate of change in fish concentration ($\text{mg}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$) is given by (1):

$$\frac{dC_f}{dt} = k_1 \times C_w - (k_2 + k_g + k_m + k_e) \times C_f \quad [\text{Equation A5.1}]$$

Where

- k_1 = First order rate constant for uptake into fish ($\text{l}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$).
- k_2 = First order rate constant for depuration from fish (day^{-1}).
- k_g = First order rate constant for fish growth ('growth dilution') (day^{-1})
- k_m = First order rate constant for metabolic transformation (day^{-1})
- k_e = First order rate constant for faecal egestion (day^{-1})
- C_w = Concentration in water ($\text{mg}\cdot\text{l}^{-1}$).
- C_f = Concentration in fish ($\text{mg}\cdot\text{kg}^{-1}$ wet weight).

For bioaccumulating substances, it can be expected that a time-weighted average (TWA) is the most relevant exposure concentration in water (C_w) within the allowed range of fluctuation (cf. paragraph 24). It is recommended to calculate a TWA water concentration, according to the procedure in Appendix 6 of TM C.20 (2). It should be noted that the ln-transformation of the water concentration is suitable when exponential decay between renewal periods is expected, e.g. in a semi-static test design. In a flow through system, ln-transformation of exposure concentrations may not be needed. If TWA water concentrations are derived, they should be reported and used in subsequent calculations.

In a standard fish BCF test uptake and depuration can be described in terms of two first order kinetic processes.

$$\text{Rate of uptake} = k_1 \times C_w \quad [\text{Equation A5.2}]$$

$$\text{Overall loss rate} = (k_2 + k_g + k_m + k_d) \times C_f \quad [\text{Equation A5.3}]$$

At steady-state, assuming growth and metabolism are negligible (*i.e.* the values for k_g and k_m cannot be distinguished from zero), the rate of uptake equals the rate of depuration, and so combining Equation A5.2 and Equation A5.3 gives the following relationship:

$$BCF = \frac{C_{f-ss}}{C_{w-ss}} = \frac{k_1}{k_2} \quad [\text{Equation A5.4}]$$

Where

C_{f-ss} = Concentration in fish at steady-state (mg kg^{-1} wet weight).

C_{w-ss} = Concentration in water at steady-state (mg l^{-1}).

The ratio of k_1/k_2 is known as the kinetic BCF (BCF_k) and should be equal to the steady-state BCF (BCF_{ss}) obtained from the ratio of the steady-state concentration in fish to that in water, but deviations may occur if steady-state was uncertain or if corrections for growth have been applied to the kinetic BCF. However, as k_1 and k_2 are constants, steady-state does not need to be reached to derive a BCF_k .

Based on these first order equations, this Appendix 5 includes the general calculations necessary for both aqueous and dietary exposure bioaccumulation methods. However, sections 5, 6 and 8 are only relevant for the aqueous exposure method but are included here as they are “general” techniques. The sequential (sections 4 and 5) and simultaneous (section 6) methods allow the calculation of uptake and depuration constants which are used to derive kinetic BCFs. The sequential method for determining k_2 (section 4) is important for the dietary method as it is needed to calculate both assimilation efficiency and BMF. Appendix 7 details the calculations that are specific to the dietary method.

2. PREDICTION OF THE DURATION OF THE UPTAKE PHASE

Before performing the test, an estimate of k_2 and hence some percentage of the time needed to reach steady-state may be obtained from empirical relationships between k_2 and the *n*-octanol/water partition coefficient (K_{ow}) or k_1 and BCF. It should be realised, however, that the equations in this section only apply when uptake and depuration follow first-order kinetics. If this is clearly not the case it is advised to seek advice from a biostatistician and/or pharmacokineticist, if predictions of the uptake phase are desirable.

An estimate of k_2 (day^{-1}) may be obtained by several methods. For example, the following empirical relationships could be used in the first instance ⁽¹⁾:

$$\log k_2 = 1,47 - 0,414 \log K_{ow} \quad (r^2=0,95) [(3); \text{Equation A5.5}]$$

or

$$k_2 = \frac{k_1}{BCF} \quad [\text{Equation A5.6}]$$

$$\text{Where } k_1 = 520 \times W^{-0,32} \text{ (for substances with a } \log K_{ow} > 3) \quad (r^2=0,85) [(4); \text{Equation A5.7}]$$

$$\text{And } BCF = 10^{(0,910 \cdot \log K_{ow} - 1,975 \cdot \log(6,8 \cdot 10^{-7} K_{ow} + 1) - 0,786)} \quad (r^2=0,90) [(5); \text{Equation A5.8}]$$

⁽¹⁾ As with every empirical relationship, it should be verified that the test substance falls within the applicability domain of the relationship

W = mean treated fish weight (grams wet weight) at the end of uptake/start of depuration ⁽¹⁾

For other related relationships *see* (6). It may be advantageous to employ more complicated models in the estimation of k_2 if, for example, it is likely that significant metabolism may occur (7) (8). However as the complexity of the model increases, greater care should be taken with the interpretation of the predictions. For example the presence of nitro groups might indicate fast metabolism, but this is not always the case. Therefore the user should weigh up the predictive method results against chemical structure and any other relevant information (for example preliminary studies) in the scheduling of a study.

The time to reach a certain percentage of steady-state may be obtained, by applying the k_2 -estimate, from the general kinetic equation describing uptake and depuration (first-order kinetics), assuming growth and metabolism is negligible. If substantial growth occurs during the study, the estimations described below will not be reliable. In such cases, it is better to use the growth corrected k_{2g} as described later (see Section 7 of this Appendix):

$$\frac{dC_f}{dt} = k_1 C_w - k_2 C_f \quad [\text{Equation A5.9}]$$

or, if C_w is constant:

$$C_f = \frac{k_1}{k_2} \cdot C_w (1 - e^{-k_2 t}) \quad [\text{Equation A5.10}]$$

When steady-state is approached ($t \rightarrow \infty$), Equation A5.10 may be reduced (*cf.* (9) (10)) to:

$$C_f = \frac{k_1}{k_2} \cdot C_w \quad [\text{Equation A5.11}]$$

or

$$\frac{C_f}{C_w} = \frac{k_1}{k_2} = BCF \quad [\text{Equation A5.12}]$$

Then $BCF \times C_w$ is an approximation to the concentration in the fish at steady-state (C_{f-ss}). [Note: the same approach can be used when estimating a steady-state BMF with the dietary test. In this case, BCF is replaced with BMF and C_w with C_{food} , concentration in the food, in the equations above]

Equation A5.10 may be transcribed to:

$$C_f = C_{f-ss} (1 - e^{-k_2 t}) \quad [\text{Equation A5.13}]$$

or

$$\frac{C_f}{C_{f-ss}} = 1 - e^{-k_2 t} \quad [\text{Equation A5.14}]$$

Applying Equation A5.14, the time to reach a certain percentage of steady-state may be predicted when k_2 is pre-estimated using Equation A5.5 or Equation A5.6.

As a guideline, the statistically optimal duration of the uptake phase for the production of statistically acceptable data (BCF_k) is that period which is required for the curve of the logarithm of the concentration of the test substance in fish plotted against linear time to reach at least 50 % of steady-state (*i.e.* $0.69/k_2$), but not more than 95 % of steady-state (*i.e.* $3.0/k_2$) (11). In case accumulation reaches beyond 95 % of steady-state, calculation of a BCF_{ss} becomes feasible.

⁽¹⁾ The weight of fish at the end of the uptake phase can be estimated from previous study data or knowledge of the test species' likely increase in size from a typical test starting weight over a typical uptake duration (e.g. 28 days).

The time to reach 80 percent of steady-state is (using Equation A5.14):

$$0,80 = 1 - e^{-k_2 t} \quad [\text{Equation A5.15}]$$

or

$$t_{80} = \frac{-\ln(0,20)}{k_2} = \frac{1,6}{k_2} \quad [\text{Equation A5.16}]$$

Similarly the time to reach 95 percent of steady-state is:

$$t_{95} = \frac{-\ln(0,05)}{k_2} = \frac{3,0}{k_2} \quad [\text{Equation A5.17}]$$

For example, the duration of the uptake phase (*i.e.* time to reach a certain percentage of steady-state, e.g. t_{80} or t_{95}) for a test substance with $\log K_{OW} = 4$ would be (using Equation A5.5, Equation A5.16 and Equation A5.17):

$$\log k_2 = 1,47 - 0,414 \cdot 4$$

$$k_2 = 0,652 \text{ day}^{-1}$$

$$t_{80} = \frac{1,6}{0,652} = 2,45 \text{ days (59 hours)}$$

$$\text{or } t_{95} = \frac{3,0}{0,652} = 4,60 \text{ days (110 hours)}$$

Alternatively, the expression:

$$t_{ess} = 6,54 \cdot 10^{-3} \cdot K_{OW} + 55,31 \text{ (hours)} \quad [\text{Equation A5.18}]$$

may be used to calculate the time for effective steady-state (t_{ess}) to be reached (12). For a test substance with $\log K_{OW} = 4$ this results in:

$$t_{ess} = 6,54 \cdot 10^{-3} \cdot 10^4 + 55,31 = 121 \text{ hours}$$

3. PREDICTION OF THE DURATION OF THE DEPURATION PHASE

A prediction of the time needed to reduce the body burden to a certain percentage of the initial concentration may also be obtained from the general equation describing uptake and depuration (assuming first order kinetics, *cf.* Equation A5.9 (1) (13).

For the depuration phase, C_w (or C_{food} for the dietary test) is assumed to be zero. The equation may then be reduced to:

$$\frac{dC_f}{dt} = k_2 C_f \quad [\text{Equation A5.19}]$$

or

$$C_f = C_{f,0} \cdot e^{-k_2 t} \quad [\text{Equation A5.20}]$$

where $C_{f,0}$ is the concentration at the start of the depuration period.

50 percent depuration will then be reached at the time (t_{50}):

$$\frac{C_f}{C_{f,0}} = \frac{1}{2} = e^{-k_2 t_{50}}$$

or

$$t_{50} = \frac{-\ln(0,50)}{k_2} = \frac{0,693}{k_2}$$

Similarly 95 percent depuration will be reached at:

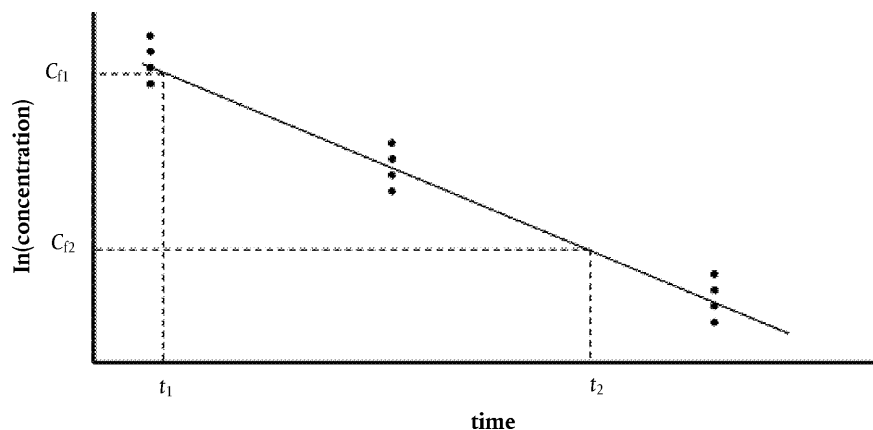
$$t_{95} = \frac{-\ln(0,05)}{k_2} = \frac{3,0}{k_2}$$

If 80 % uptake is used for the first period ($1,6/k_2$) and 95 % loss in the depuration phase ($3,0/k_2$), then depuration phase is approximately twice the duration of the uptake phase.

Note that the estimations are based on the assumption that uptake and depuration patterns will follow first order kinetics. If first-order kinetics is obviously not obeyed, these estimations are not valid.

4. SEQUENTIAL METHOD: DETERMINATION OF DEPURATION (LOSS) RATE CONSTANT K_2

Most bioconcentration data have been assumed to be 'reasonably' well described by a simple two-compartment/two-parameter model, as indicated by the rectilinear curve which approximates to the points for concentrations in fish (on an ln scale), during the depuration phase.



Note that deviations from a straight line may indicate a more complex depuration pattern than first order kinetics. The graphical method may be applied for resolving types of depuration deviating from first order kinetics.

To calculate k_2 for multiple time (sampling) points, perform a linear regression of $\ln(\text{concentration})$ versus time. The slope of the regression line is an estimate of the depuration rate constant k_2 ⁽¹⁾. From the intercept the average concentration in the fish at the start of the depuration phase ($C_{0,d}$; which equals the average concentration in the fish at the end of the uptake phase) can easily be calculated (including error margins) ⁽¹⁾:

$$C_{0,d} = e^{\text{intercept}}$$

[Equation A5.21]

⁽¹⁾ In most programs that allow a linear regression, also standard errors and confidence interval (CI) of the estimates are given, e.g. in Microsoft Excel using the Data Analysis tool pack.

To calculate k_2 when only two time (sampling) points are available (as in the minimised design), substitute the two average concentrations into the following equation

$$k_2 = \frac{\ln(C_{f1}) - \ln(C_{f2})}{t_2 - t_1} \quad [\text{Equation A5.22}]$$

Where $\ln(C_{f1})$ and $\ln(C_{f2})$ are the natural logarithms of the concentrations at times t_1 and t_2 , respectively, and t_2 and t_1 are the times when the two samples were collected relative to the start of depuration ⁽¹⁾.

5. SEQUENTIAL METHOD: DETERMINATION OF UPTAKE RATE CONSTANT k_1 (AQUEOUS EXPOSURE METHOD ONLY)

To find a value for k_1 given a set of sequential time concentration data for the uptake phase, use a computer program to fit the following model:

$$C_f(t) = C_w(t) \cdot \frac{k_1}{k_2} \cdot (1 - e^{-k_2 t}) \quad [\text{Equation A5.23}]$$

Where k_2 is given by the previous calculation, $C_f(t)$ and $C_w(t)$ are the concentrations in fish and water, respectively, at time t .

To calculate k_1 when only two time (sampling) points are available (as in the minimised design), use the following formula:

$$k_1 = \frac{C_f \cdot k_2}{C_w(1 - e^{-k_2 t})} \quad [\text{Equation A5.24}]$$

Where k_2 is given by the previous calculation, C_f is the concentration in fish at the start of the depuration phase, and C_w is the average concentration in the water during the uptake phase ⁽²⁾.

Visual inspection of the k_1 and k_2 slopes when plotted against the measured sample point data can be used to assess goodness of fit. If it turns out that the sequential method has given a poor estimate for k_1 then the simultaneous approach to calculate k_1 and k_2 should be applied (see next section 6). Again, the resulting slopes should be compared against the plotted measured data for visual inspection of goodness of fit. If the goodness of fit is still poor this may be an indication that first order kinetics do not apply and other more complex models should be employed.

6. SIMULTANEOUS METHOD FOR CALCULATION OF UPTAKE AND DEPURATION (LOSS) RATE CONSTANTS (AQUEOUS EXPOSURE METHOD ONLY)

Computer programs can be used to find values for k_1 and k_2 given a set of sequential time concentration data and the model:

$$C_f = C_w \cdot \frac{k_1}{k_2} \cdot (1 - e^{-k_2 t}) \quad 0 < t < t_c \quad [\text{Equation A5.25}]$$

$$C_f = C_w \cdot \frac{k_1}{k_2} \cdot (e^{-k_2(t-t_c)} - e^{-k_2 t}) \quad t > t_c \quad [\text{Equation A5.26}]$$

where

t_c = time at the end of the uptake phase.

⁽¹⁾ In contrast with the linear regression method, using this formula will not yield a standard error for k_2 .

⁽²⁾ In contrast with a linear fitting procedure, this method will usually not yield a standard error or confidence interval for the estimated k_1 .

This approach directly provides standard errors for the estimates of k_1 and k_2 . When k_1/k_2 is substituted by BCF (cf. Equation A5.4) in Equation A5.25 and Equation A5.26, the standard error and 95 % CI of the BCF can be estimated as well. This is especially useful when comparing different estimates due to data transformation. The dependent variable (fish concentration) can be fitted with or without \ln transformation, and the resulting BCF uncertainty can be evaluated.

As a strong correlation exists between the two parameters k_1 and k_2 if estimated simultaneously, it may be advisable first to calculate k_2 from the depuration data only (see above); k_2 in most cases can be estimated from the depuration curve with relatively high precision. k_1 can be subsequently calculated from the uptake data using non-linear regression ⁽¹⁾. It is advised to use the same data transformation when fitting sequentially.

Visual inspection of the resulting slopes when plotted against the measured sample point data can be used to assess goodness of fit. If it turns out that this method has given a poor estimate for k_1 then the simultaneous approach to calculate k_1 and k_2 can be applied. Again, the fitted model should be compared against the plotted measured data for visual inspection of goodness of fit and the resulting parameter estimates for k_1 , k_2 and resulting BCF and their standard errors and/or confidence intervals should be compared between different types of fit.

If the goodness of fit is poor this may be an indication that first order kinetics does not apply and other more complex models should be employed. One of the most common complications is fish growth during the test.

7. GROWTH DILUTION CORRECTION FOR KINETIC BCF AND BMF

This section describes a standard method for correction due to fish growth during the test (so called 'growth dilution') which is only valid when first order kinetics applies. In case there are indications that first order kinetics do not apply, it is advised to seek advice from a biostatistician for a proper correction of growth dilution or to use the mass based approach described below.

In some cases this method for correcting growth dilution is subject to a lack of precision or sometimes does not work (for example for very slowly depurating substances tested in fast growing fish the derived depuration rate constant corrected for growth dilution, k_{2g} , may be very small and so the error in the two rate constants used to derive it become critical, and in some cases k_g estimates may be larger than k_2). In such cases an alternative approach (i.e. mass approach), which also works when first order growth kinetics have not been obeyed, can be used which avoids the need for the correction. This approach is outlined at the end of this section.

Growth rate constant subtraction method for growth correction

For the standard method all individual weight and length data are converted to natural logarithms and \ln (weight) or $\ln(1/\text{weight})$ is plotted vs. time (day), separately for treatment and control groups. The same process is carried out for the data from the uptake and depuration phases separately. Generally for growth dilution correction it is more appropriate to use the weight data from the whole study to derive the growth rate constant (k_g), but statistically significant differences between the growth rate constants derived for the uptake phase and depuration phase may indicate that the depuration phase rate constant should be used. Overall growth rates from aqueous studies for test and control groups can be used to check for any treatment related effects.

⁽¹⁾ It should be realised that the uncertainty in the k_2 estimate is not used properly in the bioaccumulation model when this is essentially regarded as constant when fitting k_1 in the sequential fit method. The resulting BCF uncertainty will therefore be different between the simultaneous and sequential fitting methods.

A linear least squares correlation is calculated for the $\ln(\text{fish weight})$ vs. day (and for $\ln(1/\text{weight})$ vs. day) for each group (test(s) and control groups, individual data, not daily mean values) for the whole study, uptake and depuration phases using standard statistical procedures. The variances in the slopes of the lines are calculated and used to evaluate the statistical significance ($p = 0,05$) of the difference in the slopes (growth rate constants) using the student t -test (or ANOVA if more than one concentration is tested). Weight data are generally preferred for growth correction purposes. Length data, treated in the same way, may be useful to compare control and test groups for treatment related effects. If there is no statistically significant difference in the weight data analysis, the test and control data may be pooled and an overall fish growth rate constant for the study (k_g) calculated as the overall slope of the linear correlation. If statistically significant differences are observed, growth rate constants for each fish group, and/or study phase, are reported separately. The rate constant from each treated group should then be used for growth dilution correction purposes of that group. If statistical differences between the uptake and depuration phase rate constants were noted, depuration phase derived rate constants should be used.

The calculated growth rate constant (k_g expressed as day^{-1}) can be subtracted from the overall depuration rate constant (k_2) to give the depuration rate constant, k_{2g} .

$$k_{2g} = k_2 - k_g \quad [\text{Equation A5.27}]$$

The uptake rate constant is divided by the growth-corrected depuration rate constant to give the growth-corrected kinetic BCF, denoted BCF_{kg} (or BMF_{kg}).

$$\text{BCF}_{kg} = \frac{k_1}{k_{2g}} \quad [\text{Equation A5.28}]$$

The growth rate constant derived for a dietary study is used in Equation A7.5 to calculate the growth corrected BMF_{kg} (cf. Appendix 7).

Mass based method for growth correction

An alternative to the above "growth rate constant subtraction method" that avoids the need to correct for growth can be used as follows. The principle is to use depuration data on a mass basis per whole fish rather than on a concentration basis.

Convert depuration phase tissue concentrations (mass of test substance/unit mass of fish) into mass of test substance/fish: match concentrations and individual fish weights in tabular form (e.g. using a computer spreadsheet) and multiply each concentration by the total fish weight for that measurement to give a set of mass test substance/fish for all depuration phase samples.

Plot the resulting natural logarithm of substance mass data against time for the experiment (depuration phase) as would be done normally.

For the aqueous exposure method, derive the uptake rate constant routinely (see sections 4 and 6) note that the "normal" k_2 value should be used in the curve fitting equations for k_1) and derive the depuration rate constant from the above data. Because the resulting value for the depuration rate constant is independent of growth as it has been derived on a mass basis per whole fish, it should be denoted as k_{2g} and not k_2 .

8. LIPID NORMALISATION TO 5 % LIPID CONTENT (AQUEOUS EXPOSURE METHOD ONLY)

BCF results (kinetic and steady-state) from aqueous exposure tests should also be reported relative to a default fish lipid content of 5 % wet weight, unless it can be argued that the test substance does not primarily accumulate in lipid (e.g. some perfluorinated substances may bind to proteins). Fish concentration data, or the BCF, need to be converted to a 5 % lipid content wet weight basis. If the same fish were used for measuring substance concentrations and lipid contents at all sampling points, this requires each individual measured concentration in the fish to be corrected for that fish's lipid content.

$$C_{fL} = \frac{0.05}{L} \cdot C_f \quad [\text{Equation A5.29}]$$

where

C_{fL} = lipid-normalised concentration in fish (mg kg⁻¹ wet weight)

L = lipid fraction (based on wet weight)

C_f = concentration of test substance in fish (mg kg⁻¹ wet weight)

If lipid analysis was not conducted on all sampled fish, a mean lipid value is used to normalise the BCF. For the steady-state BCF, the mean value recorded at the end of the uptake phase in the treatment group should be used. For the normalisation of a kinetic BCF there may be some cases where a different approach is warranted, for example if the lipid content changed markedly during the uptake or depuration phase. However a feeding rate that minimises dramatic changes in lipid content should be used anyway routinely.

$$BCF_{KL} = \frac{0.05}{L_n} \cdot BCF_K \quad [\text{Equation A5.30}]$$

where

BCF_{KL} = lipid-normalised kinetic BCF (L kg⁻¹)

L_n = mean lipid fraction (based on wet weight)

BCF_K = kinetic BCF (L kg⁻¹)

LITERATURE:

- (1) Arnot J.A. and Gobas F.A.P.C. (2004), A food web bioaccumulation model for organic chemicals in aquatic ecosystems, *Environ. Toxicol. Chem.* 23: 2343–2355.
- (2) Chapter C.20 of this Annex, *Daphnia magna* Reproduction Test.
- (3) Spacie A. and Hamelink J.L. (1982), Alternative models for describing the bioconcentration of organics in fish. *Environ. Toxicol. Chem.* 1: 309-320.
- (4) Sijm D.T.H.M., Verberne M.E., de Jonge W.J., Pärt P. and Opperhuizen A. (1995), Allometry in the uptake of hydrophobic chemicals determined *in vivo* and in isolated perfused gills. *Toxicol. Appl. Pharmacol.* 131: 130-135.
- (5) Bintein S., Devillers J. and Karcher W. (1993), Nonlinear dependence of fish bioconcentration on *n*-octanol/water partition coefficient. *SAR QSAR Environ. Res.* 1: 29-39.

- (6) Kristensen P. (1991), Bioconcentration in fish: comparison of BCF's derived from OECD and ASTM testing methods; influence of particulate matter to the bioavailability of chemicals. Danish Water Quality Institute, Hørsholm, Denmark.
 - (7) Arnot J.A., Meylan W., Tunkel J., Howard P.H., Mackay D., Bonnell M. and Boethling R.S. (2009), A quantitative structure-activity relationship for predicting metabolic biotransformation rates for organic chemicals in fish. *Environ. Toxicol. Chem.* 28: 1168-1177.
 - (8) OECD (2011), QSAR Toolbox 2.1. February 2011. Available from: http://www.oecd.org/document/54/0,3746,en_2649_34379_42923638_1_1_1_1,00.html.
 - (9) Branson D.R., Blau G.E., Alexander H.C. and Neely W.B. (1975). Bioconcentration of 2,2',4,4' tetrachlorobiphenyl in rainbow trout as measured by an accelerated test. *T. Am. Fish. Soc.* 104: 785-792.
 - (10) Ernst W. (1985), Accumulation in aquatic organisms, in *Appraisal of tests to predict the environmental behaviour of chemicals*, Sheeman, P., *et al.*, Editors. John Wiley & Sons Ltd, New York, NY, USA: 243-255.
 - (11) Reilly P.M., Bajramovic R., Blau G.E., Branson D.R. and Sauerhoff M.W. (1977), Guidelines for the optimal design of experiments to estimate parameters in first order kinetic models. *Can. J. Chem. Eng.* 55: 614-622.
 - (12) Hawker D.W. and Connell D.W. (1988), Influence of partition coefficient of lipophilic compounds on bioconcentration kinetics with fish. *Wat. Res.* 22: 701-707.
 - (13) Konemann H. and van Leeuwen K. (1980), Toxicokinetics in fish: Accumulation and elimination of six chlorobenzenes by guppies. *Chemosphere.* 9: 3-19.
-

Appendix 6

EQUATION SECTION FOR AQUEOUS EXPOSURE TEST: MINIMISED TEST DESIGN

The rationale for this approach is that the bioconcentration factor in a full test can either be determined as a steady-state bioconcentration factor (BCF_{ss}) by calculating the ratio of the concentration of the test substance in the fish's tissue to the concentration of the test substance in the water, or by calculating the kinetic bioconcentration factor (BCF_k) as the ratio of the uptake rate constant k_1 to the depuration rate constant k_2 . The BCF_k is valid even if a steady-state concentration of a substance is not achieved during uptake, provided that uptake and depuration act approximately according to first order kinetic processes.

If a measurement of the concentration of the substance in tissues (C_{f1}) is made at the time that exposure ends (t_1) and the concentration in tissue (C_{f2}) is measured again after a period of time has elapsed (t_2), the depuration rate constant (k_2) can be estimated using Equation A5.22 from Appendix 5.

The uptake rate constant, k_1 , can then be determined algebraically using Equation A5.23 from Appendix 5 (where C_f equals C_{f1} and t equals t_1) (1). The kinetic bioconcentration factor for the minimised design (designated as BCF_{km} to distinguish it from kinetic bioconcentration factors determined using other methods) is thus:

$$BCF_{km} = \frac{k_1}{k_2} \quad \text{[Equation A6.1]}$$

Concentrations or results should be corrected for growth dilution and normalised to a fish lipid content of 5 %, as is described in Appendix 5.

The minimised BCF_{ss} is the BCF calculated at the end of the uptake phase, assuming that steady-state has been reached. This can only be assumed, as the number of sampling points is not sufficient for proving this.

$$\text{minimised } BCF_{ss} = \frac{C_{f - minSS}}{C_{w - minSS}} \quad \text{[Equation A6.2]}$$

Where

$C_{f-minSS}$ = Concentration in fish at assumed steady-state at end of uptake (mg kg^{-1} wet weight).

$C_{w-minSS}$ = Concentration in water at assumed steady-state at end of uptake (mg l^{-1}).

LITERATURE:

- (1) Springer T.A., Guiney P.D., Krueger H.O. and Jaber M.J. (2008), Assessment of an approach to estimating aquatic bioconcentration factors using reduced sampling. *Environ. Toxicol. Chem.* 27: 2271-2280.

Appendix 7

EQUATION SECTION FOR DIETARY EXPOSURE TEST

1. Example of constituent quantities of a suitable commercial fish food
2. Food spiking technique examples
3. Calculation of assimilation efficiency and biomagnification factor
4. Lipid correction
5. Evaluation of differences between measured time zero concentration (C_{0,m}) and derived time zero concentration (C_{0,d})
6. Guidance for very fast depurating test substances

1. EXAMPLE OF CONSTITUENT QUANTITIES OF A SUITABLE COMMERCIAL FISH FOOD

Major constituent	Fish meal
Crude Protein	≤ 55,0 %
Crude fat	≤ 15,0 % ⁽¹⁾
Crude Fibre	≥ 2,0 %
Moisture	≥ 12 %
Ash	≥ 8 %

⁽¹⁾ In some regions it may only be possible to obtain fish food with a lipid concentration that falls far short of this upper limit. In such cases studies should be run with the lower lipid concentration in the food as supplied, and the feeding rate adjusted appropriately to maintain fish health. Diet lipids should not be artificially increased by the addition of excess oil.

2. FOOD SPIKING TECHNIQUE EXAMPLES

General Points

Control diets should be prepared in exactly the same way as the spiked diet, but with an absence of test substance.

To check the concentration of the treated diet, triplicate samples of the dosed food should be extracted with a suitable extraction method and the test substance concentration or radioactivity in the extracts measured. High analytical recoveries (> 85 %) with low variation between samples (three sample concentrations for the substance taken at test start should not vary more than ± 15 % from the mean) should be demonstrated.

During the dietary test, three diet samples for analysis should be collected on day 0 and at the end of the uptake phase for the determination of the test substance content in the diet.

Fish food preparation with a liquid test material (neat)

A target, nominal test concentration in the treated fish food is set, for example 500 µg test substance/g food. The appropriate quantity (by molar mass or specific radioactivity) of neat test substance is added to a known mass of fish food in a glass jar or rotary evaporator bulb. The mass of fish food should be sufficient for the duration of the uptake phase (taking into account the need for increasing quantities at each feed owing to fish growth). The fish feed/test substance should be mixed overnight by slow tumbling (e.g. using a roto-rack mixer or by rotation if a rotary evaporator bulb is used). The spiked diet should be stored under conditions that maintain stability of the test substance within the feed mix (e.g. refrigeration) until use.

Fish food preparation with a corn or fish oil vehicle

Solid test substances should be ground in a mortar to a fine powder. Liquid test substances can be added directly to the corn or fish oil. The test substance is dissolved in a known quantity of corn or fish oil (e.g. 5-15 ml). The dosed oil is quantitatively transferred into a rotary evaporation bulb of suitable size. The flask used to prepare the dosed oil should be flushed with two small aliquots of oil and these added to the bulb to make sure all dissolved test substance is transferred. To ensure complete dissolution/dispersion in the oil (or if more than one test substance is being used in the study), a micro-stirrer is added, the flask stoppered and the mixture stirred rapidly overnight. An appropriate quantity of fish diet (usually in pellet form) for the test is added to the bulb, and the bulb's contents are mixed homogeneously by continuously turning the glass bulb for at least 30 minutes, but preferably overnight. Thereafter, the spiked food is stored appropriately (e.g. refrigerated) to ensure test substance stability in the food until use.

Fish food preparation with an organic solvent

An appropriate quantity of test substance (by molar mass or specific radioactivity) sufficient to achieve the target dose is dissolved in a suitable organic solvent (e.g. cyclohexane or acetone; 10-40 ml, but a greater volume if necessary depending on the quantity of food to spike). Either an aliquot, or all (added portion wise), of this solution is mixed with the appropriate mass of fish food sufficient for the test to achieve the required nominal dose level. The food/test substance can be mixed in a stainless steel mixing bowl and the freshly-dosed fish food left in the bowl in a laboratory hood for two days (stirred occasionally) to allow the excess solvent to evaporate, or mixed in a rotary evaporator bulb with continuous rotation. The excess solvent can be "blown" off under a stream of air or nitrogen if necessary. Care should be taken to ensure that the test substance does not crystallise as the solvent is removed. The spiked diet should be stored under conditions (e.g. refrigeration) that maintain stability of the test substance within the feed mix until use.

3. CALCULATION OF ASSIMILATION EFFICIENCY AND BIOMAGNIFICATION FACTOR

To calculate the assimilation efficiency, the overall depuration rate constant should first be estimated according to section 4 of Appendix 5 (using the "sequential method", i.e. standard linear regression) using mean sample concentrations from the depuration phase. The feeding rate constant, I , and uptake duration, t , are known parameters of the study. C_{food} , the mean measured concentration of the test substance in the food is a measured variable in the study. $C_{0,d}$, the test substance concentration in the fish at the end of the uptake phase, is usually derived from the intercept of a plot of $\ln(\text{concentration})$ vs. depuration day.

The substance assimilation efficiency (a , absorption of test substance across the gut) is calculated as:

$$a = \frac{C_{0,d} \cdot k_2}{I \cdot C_{\text{food}}} \cdot \frac{1}{1 - e^{-k_2 t}} \quad [\text{Equation A7.1}]$$

where:

$C_{0,d}$ = derived concentration in fish at time zero of the depuration phase (mg kg^{-1});

k_2 = overall (not growth-corrected) depuration rate constant (day^{-1}), calculated according to equations in Appendix 5, Section 3;

I = food ingestion rate constant ($\text{g food g}^{-1} \text{ fish day}^{-1}$);

C_{food} = concentration in food ($\text{mg kg}^{-1} \text{ food}$);

t = duration of the feeding period (day)

However, the feeding rate, I , used in the calculation may need to be adjusted for fish growth to give an accurate assimilation efficiency, a . In a test where fish grow significantly during the uptake phase (in which no correction of feed quantities is made to maintain the set feeding rate), the effective feeding rate as the uptake phase progresses will be lower than that set, resulting in a higher 'real' assimilation efficiency. (Note this is not important for the overall calculation of BMF as the I terms effectively cancel out between Equation A7.1 and Equation A7.4). The mean feeding rate corrected for growth dilution, I_g , can be derived in several ways, but a straightforward and rigorous one is to use the known growth rate constant (k_g) to estimate the test fish weights at time points during the uptake phase, i.e.:

$$W_f(t) = W_{f,0} \times e^{k_g t} \quad [\text{Equation A7.2}]$$

where

$W_f(t)$ = mean fish weight at uptake day t

$W_{f,0}$ = mean fish weight at the start of the experiment

In this way (at least) the mean fish weight on the last day of exposure ($W_{f,\text{end-of-uptake}}$) can be estimated. As the feeding rate was set based on $W_{f,0}$, the effective feeding rate for each day of uptake can be calculated using these two weight values. The growth-corrected feeding rate, I_g ($\text{g food g}^{-1} \text{ fish day}^{-1}$), to use instead of I in cases of rapid growth during the uptake phase, can then be calculated as

$$I_g = \frac{I \times W_{f,0}}{W_{f,\text{end-of-uptake}}} \quad [\text{Equation A7.3}]$$

Once the assimilation efficiency has been obtained, the BMF can be calculated by multiplying it with the feeding rate constant I (or I_g , if used to calculate a) and dividing the product by the overall depuration rate constant k_2 :

$$\text{BMF} = \frac{I \times a}{k_2} \quad [\text{Equation A7.4}]$$

The growth-corrected biomagnification factor should also be calculated in the same way, using the growth corrected depuration rate constant (as derived according to section 7 in Appendix 5). Again, if I_g has been used to calculate a , it should also be used here instead of I :

$$\text{BMF} = \frac{I \times a}{k_{2g}} \quad [\text{Equation A7.5}]$$

where:

α = assimilation efficiency (absorption of test substance across the gut);

k_2 = overall (not growth-corrected) depuration rate constant (day^{-1}), calculated according to equations in Appendix 5, Section 3;

k_{2g} = growth-corrected depuration rate constant (day^{-1});

I = food ingestion rate constant ($\text{g food g}^{-1} \text{ fish day}^{-1}$);

The growth-corrected half-life ($t_{1/2}$) is calculated as follows.

$$t_{1/2} = \frac{0,693}{k_{2g}} \quad [\text{Equation A7.6}]$$

The substance assimilation efficiency from the diet can also be estimated if tissue residues are determined during the linear phase of the uptake phase (between days 1 and 3). In this case the substance assimilation efficiency (α) can be determined as follows

$$\alpha = \frac{C_{\text{fish}}(t)}{I \times C_{\text{food}} \times t} \quad [\text{Equation A7.7}]$$

Where

$C_{\text{fish}}(t)$ = the concentration of test substance in the fish at time t (mg kg^{-1} wet weight).

4. LIPID CORRECTION

If lipid content was measured on the same fish as chemical analysis for all sampling intervals, then individual concentrations should be corrected on a lipid basis and the $\ln(\text{concentration, lipid corrected})$ plotted against depuration (day) to give $C_{0,d}$ and k_2 . Assimilation efficiency (Equation A7.1) can then be calculated on a lipid basis, using C_{food} on a lipid basis (i.e. C_{food} is multiplied by the mean lipid fraction of the food). Subsequent calculation using Equation A7.4 and Equation A7.5 will give the lipid-corrected (and growth-dilution corrected) BMF directly.

Otherwise, the mean lipid fraction (w/w) in the fish and in the food are derived for both treatment and control groups (for food and control group fish this is usually from data measured at exposure start and end; for treatment group fish this is usually from data measured at end of exposure only). In some studies, fish lipid content may increase markedly; in such cases it is more appropriate to use a mean test fish lipid concentration calculated from the measured values at the end of exposure and end of depuration. In general, data from the treatment group only should be used to derive both of the lipid fractions.

The lipid-correction factor (L_c) is calculated as:

$$L_c = \frac{L_{\text{fish}}}{L_{\text{food}}} \quad [\text{Equation A7.8}]$$

where L_{fish} and L_{food} are the mean lipid fractions in fish and food, respectively.

The lipid-correction factor is used to calculate the lipid-corrected biomagnification factor (BMF_L):

$$BMF_L = \frac{BMF}{L_c} \quad [\text{Equation A7.9}]$$

5. EVALUATION OF DIFFERENCES BETWEEN MEASURED TIME ZERO CONCENTRATION ($C_{0,m}$) AND DERIVED TIME ZERO CONCENTRATION ($C_{0,d}$)

The measured time zero concentration ($C_{0,m}$) and derived time zero concentration ($C_{0,d}$) should be compared. If they are very similar, then this supports the first order model used to derive the depuration parameters.

In some studies there may be a marked difference between the derived time zero value, $C_{0,d}$, and the mean measured time zero concentration, $C_{0,m}$ (see last bullet point of paragraph 159 of this test method). If $C_{0,d}$ is very much lower than $C_{0,m}$ ($C_{0,d} \ll C_{0,m}$), the difference may suggest the presence of undigested spiked food in the gut. This may be tested experimentally by conducting separate analysis on the excised gut if additional (whole fish) samples were taken and stored at the end of the uptake phase. Otherwise, if a statistically valid outlier test applied to the depuration phase linear regression indicates that the first sample point of depuration is erroneously elevated, carrying out the linear regression to derive k_2 but omitting the first depuration concentration point may be appropriate. In such cases, if the uncertainty in the linear regression is greatly decreased, and it is clear that approximately first order depuration kinetics were obeyed, it may be appropriate to use the resulting $C_{0,d}$ and k_2 values in the assimilation efficiency calculation. This should be fully justified in the report. It is also possible that non-first order kinetics were operating in the depuration phase. If this is likely (i.e. the natural logarithm transformed data appear to follow a curve compared with the straight-line linear regression plot), then the calculations of k_2 and $C_{0,d}$ are unlikely to be valid and the advice of a biostatistician should be sought.

If $C_{0,d}$ is very much higher than the measured value ($C_{0,d} \gg C_{0,m}$) this may indicate: that the substance was depurated very fast (i.e. sampling points approached the limit of quantification of the analytical method very early in the depuration phase, cf. Section 6 below); that there was a deviation from first order depuration kinetics; that the linear regression to derive k_2 and $C_{0,d}$ is flawed; or that a problem with the measured concentrations in the study occurred at some sampling time points. In such cases the linear regression plot should be scrutinised for evidence of samples at or near the limit of quantification, for outliers and for obvious curvature (suggestive of non-first order kinetics), and highlighted in the report. Any subsequent re-evaluation of the linear regression to improve estimated values should be described and justified. If marked deviation from first order kinetics is observed, then the calculations of k_2 and $C_{0,d}$ are unlikely to be valid and the advice of a biostatistician should be sought.

6. GUIDANCE FOR VERY FAST DEPURATING TEST SUBSTANCES

As discussed in paragraph 129 of the test method, some substances may depurate so fast that a reliable time zero concentration, $C_{0,d}$, and k_2 cannot be derived because in samples very early in the depuration phase (i.e. from the second depuration sample onwards) the substance is effectively no longer measured (concentrations reported at the limit of quantification). This situation was observed in the ring test carried out in support of this test method with benzo[a]pyrene, and has been documented in the validation report for the method. In such cases linear regression cannot be carried out reliably, and is likely to give an unrealistically high estimate of $C_{0,d}$, resulting in an apparent assimilation efficiency much greater than 1. It is possible to calculate a conservative estimate of k_2 and an "upper bound" BMF in these instances.

Using those data points of the depuration phase where a concentration was measured, up to and including the first “non-detect” concentration (concentration set at limit of quantification), a linear regression (using natural logarithm transformed concentration data against time) will give an estimate of k_2 . For these sorts of cases this is likely only to involve two data points (e.g. sample days 1 and 2 of depuration) and then k_2 can be estimated using Equation A5.22 in Appendix 5. This k_2 estimate can be used to estimate an assimilation efficiency according to equation A7.1, substituting the $C_{0,d}$ value in the equation with the measured time zero concentration ($C_{0,m}$) in cases where $C_{0,d}$ is clearly estimated to be much higher than could have been achievable in the test. If $C_{0,m}$ was not measureable, then the limit of detection in fish tissue should be used. If, in some cases, this gives a value of $\alpha > 1$, then the assimilation efficiency is assumed to 1 as a “worst case”.

The maximum BMF_k can then be estimated using Equation A7.4, and should be quoted as a “much less than” (<<) value. For example, for a study carried out with a feeding rate of 3 % and a depuration half-life less than 3 days, and a “worst case” α of 1, the BMF_k is likely to be below about 0,13. Given the purpose of this estimation and the fact that values will be conservative in nature, it is not necessary to correct them for growth dilution or fish and food lipid content.

Appendix 8

APPROACHES TO ESTIMATE TENTATIVE BCFs FROM DATA COLLECTED IN THE DIETARY EXPOSURE STUDY

The dietary method is included in this test method for the bioaccumulation testing of substances that cannot in practice be tested using the aqueous exposure method. The aqueous exposure method gives a bioconcentration factor, whereas the dietary method leads directly to information on feeding biomagnification potential. In many chemical safety regimes information on aquatic bioconcentration is required (for example in risk assessment and the Globally Harmonization System of Classification). Hence there is a need to use the data generated in a dietary study to estimate a bioconcentration factor that is comparable to tests conducted according to the aqueous exposure method ⁽¹⁾. This section explores approaches that may be followed to do this, while recognising the shortcomings that are inherent in the estimations.

The dietary study measures depuration to give a depuration rate constant, k_2 . If an uptake rate constant can be estimated with the available data for the situation where the fish had been exposed to the test substance via the water, then a kinetic BCF could be estimated.

The estimation of an uptake rate constant for water exposure of a test substance is reliant on many assumptions, all of which will contribute to the estimate's uncertainty. Furthermore, this approach to estimating a BCF assumes that the overall rate of depuration (including contributory factors like distribution in the body and individual depuration processes) is independent of the exposure technique used to produce a test substance body burden.

The main assumptions inherent in the estimation approach can be summarised as follows.

Depuration following dietary uptake is the same as depuration following aqueous exposure for a given substance

Uptake from water would follow first order kinetics

Depending on the method used to estimate uptake:

- uptake can be correlated with fish weight alone
- uptake can be correlated with the substance's octanol-water partition coefficient alone
- uptake can be correlated with a combination of fish weight and the substance's octanol-water partition coefficient
- factors that can affect uptake in an aqueous exposure study in practice such as substance bioavailability, adsorption to apparatus, molecular size etc. have little effect
- and, crucially:

The database ("training set") used to develop the uptake estimation method is representative of the substance under consideration

Several publications in the open literature have derived equations relating uptake from water in fish via the gills to a substance's octanol-water partition coefficient, fish weight (1) (2) (3) (4), volume and/or lipid content, membrane permeation/diffusion (5) (6), fish ventilation volume (7) and by a fugacity/mass balance approach (8) (9) (10). A detailed appraisal of such methods in this context is given in Crookes & Brooke (11). A publication by Barber (12) focussed on modelling bioaccumulation through dietary uptake is also useful in this context as it includes contributions from gill uptake rate models. A section of the background document to the 2004 dietary protocol (13) was also devoted to this aspect.

⁽¹⁾ In the wild the route leading to greatest exposure in aqueous environments is likely to be through ingestion for very hydrophobic substances and so an estimated BCF is not strictly representative of such a substance's bioaccumulation potential.

Most of these models seem to have been derived using limited databases. For models where details of the database used to build the model are available, it appears that the types of substances used are often of a similar structure or class (in terms of functionality, e.g. organochlorines). This adds to the uncertainty in using a model to predict an uptake rate constant for a different type of substance, in addition to test-specific considerations like species, temperature, etc.

A review of available techniques (11) highlighted that no one method is “more correct” than the others. Therefore, a clear justification should be given for the model used. Where several methods are available for which the use can be justified, it may be prudent to present several estimates of k_1 (and so BCF) or a range of k_1 values (and BCF) according to several uptake estimation methods. However, given the differences in model types and datasets used to develop them, taking a mean value from estimates derived in different ways would not be appropriate.

Some researchers have postulated that BCF estimates of this sort require a bioavailability correction to account for a substance's adsorption to dissolved organic carbon (DOC) under aqueous exposure conditions, to bring the estimate in line with results from aqueous exposure studies (e.g. (13) (14)). However, this correction may not be appropriate given the low levels of DOC required in an aqueous exposure study for a ‘worst case’ estimate (i.e. ratio of bioavailable substance to substance as measured in solution). For highly hydrophobic substances uptake at the gill may become limited by the rate of passive diffusion near the gill surface; in this case it is possible that the correction may be accounting for this effect rather than what it was designed for.

It is advised to focus on methods that require inputs for which data will be readily available for substances tested according to the dietary study described here (i.e. $\log K_{ow}$, fish weight). Other methods that require more complex inputs may be applied, but may need additional measurements in the test or detailed knowledge on the test substance or fish species that may not be widely available. In addition, choice of model may be influenced by the level of validation and applicability domain (see (11) for a review and comparison of different methods).

It should be borne in mind that the resulting k_1 estimate, and estimated BCF, are uncertain and may need to be treated in a weight-of-evidence approach along with the derived BMF and substance parameters (e.g. molecular size) for an overall picture of a substance's bioaccumulation potential. Interpretation and use of these parameters may depend on the regulatory framework.

LITERATURE:

- (1) Sijm D.T.H.M., Pärt P. and Opperhuizen A. (1993), The influence of temperature on the uptake rate constants of hydrophobic compounds determined by the isolated perfused gills of rainbow trout (*Oncorhynchus mykiss*). *Aquat. Toxicol.* 25: 1-14.
- (2) Sijm D.T.H.M., Verberne M.E., Part P. and Opperhuizen A. (1994), Experimentally determined blood and water flow limitations for uptake of hydrophobic compounds using perfused gills of rainbow trout (*Oncorhynchus mykiss*): Allometric applications. *Aquat. Toxicol.* 30: 325-341.
- (3) Sijm D.T.H.M., Verberne M.E., de Jonge W.J., Pärt P. and Opperhuizen A. (1995), Allometry in the uptake of hydrophobic chemicals determined *in vivo* and in isolated perfused gills. *Toxicol. Appl. Pharmacol.* 131: 130-135.
- (4) Barber M.C. (2003), A review and comparison of models for predicting dynamic chemical bioconcentration in fish. *Environ. Toxicol. Chem.* 22: 1963-1992.
- (5) Opperhuizen A. (1986), Bioconcentration of hydrophobic chemicals in fish, in *Aquatic Toxicology and Environmental Fate*, STP 921, Poston, T.M. and Purdy, R., Editors. American Society for Testing and Materials, Philadelphia, PA, USA: 304-315.

- (6) Arnot J.A. and Gobas F.A.P.C. (2004), A food web bioaccumulation model for organic chemicals in aquatic ecosystems. *Environ. Toxicol. Chem.* 23: 2343-2355.
- (7) Thomann R.V. (1989), Bioaccumulation model of organic chemical distribution in aquatic food chains. *Environ. Sci. Technol.* 23: 699-707.
- (8) Hendriks A.J., van der Linde A., Cornelissen G. and Sijm D.T.H.M. (2001). The power of size. 1. Rate constants and equilibrium ratios for accumulation of organic substances related to octanol-water partition ratio and species weight. *Environ. Toxicol. Chem.* 20: 1399-1420.
- (9) Campfens J. and Mackay D. (1997), Fugacity-based model of PCB bioaccumulation in complex aquatic food webs. *Environ. Sci. Technol.* 31: 577-583.
- (10) Arnot J.A. and Gobas F.A.P.C. (2003), A generic QSAR for assessing the bioaccumulation potential of organic chemicals in aquatic food webs. *QSAR Comb. Sci.* 22: 337-345.
- (11) Crookes M. and Brooke D. (2010), Estimation of fish bioconcentration factor (BCF) from depuration data. Draft Report. Environmental Agency, Bristol, UK.
- (12) Barber M.C. (2008), Dietary uptake models used for modelling the bioaccumulation of organic contaminants in fish. *Environ. Toxicol. Chem.* 27: 755-777
- (13) Anonymous (2004), Background document to the fish dietary study protocol, document submitted to the TC-NES WG on PBT.
- (14) Gobas F. and Morrison H. (2000), Bioconcentration and biomagnification in the aquatic environment, in *Handbook of property estimation methods for chemicals*, Boethling, R.S. and Mackay, D., Editors. Lewis Publishers, Boca Raton, FL, USA: 189-231.'

(17) In Part C, Chapter C.20 is replaced by the following:

'C.20 *Daphnia magna* Reproduction Test

INTRODUCTION

This test method is equivalent to OECD test guideline (TG) 211 (2012). OECD test guidelines are periodically reviewed in the light of scientific progress. The reproduction test guideline 211 originates from test guideline 202, Part II, *Daphnia* sp. reproduction test (1984). It had generally been acknowledged that data from tests performed according to that TG 202 could be variable. This led to considerable effort being devoted to the identification of the reasons for this variability with the aim of producing a better test method. Test guideline 211 is based on the outcome of these research activities, ring-tests and validation studies performed in 1992 (1), 1994 (2) and 2008 (3).

The main differences between the initial version (TG 202, 1984), and second version (TG 211, 1998) of the reproduction test guideline are:

- the recommended species to be used is *Daphnia magna*;
- the test duration is 21 days;
- for semi-static tests, the number of animals to be used at each test concentration has been reduced from at least 40, preferably divided into four groups of 10 animals, to at least 10 animals held individually (although different designs can be used for flow-through tests);

- more specific recommendations have been made with regard to test medium and feeding conditions.
- the main differences between the second version of the reproduction test guideline (TG 211, 1998) and this version are:
- appendix 7 has been added to describe procedures for the identification of neonate sex if required. In line with previous versions of this test method sex ratio is an optional endpoint;
- the response variable number of living offspring produced per surviving parental animal has been supplemented with an additional response variable for *Daphnia* reproduction, i.e. the total number of living offspring produced at the end of the test per parent daphnia at the start of the test excluding from the analysis parental accidental and/or inadvertent mortality. The purpose of the added response variable is to align this response variable with other reproduction test methods on invertebrates. Furthermore, in relation to this response variable, it is possible, in this test method, to remove a source of error, namely the effect of inadvertent and/or accidental parental mortality, should that occur during the exposure period.
- additional statistical guidance for test design and for treatment of results has been included both for ECx (e.g. EC10 or EC50) and for NOEC/LOEC approach.
- a limit test is introduced.

Definitions used are given in Appendix 1.

PRINCIPLE OF THE TEST

The primary objective of the test is to assess the effect of chemicals on the reproductive output of *Daphnia magna*. To this end, young female *Daphnia* (the parent animals), aged less than 24 hours at the start of the test, are exposed to the test chemical added to water at a range of concentrations. The test duration is 21 days. At the end of the test, the total number of living offspring produced is assessed. Reproductive output of the parent animals can be expressed in other ways (e.g. number of living offspring produced per animal per day from the first day offspring were observed) but these should be reported in addition to the total number of living offspring produced at the end of the test. Because of the particular design of the semi-static test compared to other invertebrate reproduction test methods, it is also possible to count the number of living offspring produced by each individual parent animal. This enables that, contrary to other invertebrate reproduction test methods, if the parent animal dies accidentally and/or inadvertently during the test period, its offspring production can be excluded from data assessment. Hence, if parental mortality occurs in exposed replicates, it should be considered whether or not the mortality follows a concentration-response pattern, e.g. if there is a significant regression of the response versus concentration of the test chemical with a positive slope (a statistical test like the Cochran-Armitage trend test may be used for this). If the mortality does not follow a concentration-response pattern, then those replicates with parental mortality should be excluded from the analysis of the test result. If the mortality follows a concentration-response pattern, the parental mortality should be assigned as an effect of the test chemical and the replicates should not be excluded from the analysis. If the parent animal dies during the test i.e. accidentally from mishandling or accident, or inadvertently due to unexplained incident not related to the effect of the test chemical or turns out to be male, then the replicate is excluded from the analysis (see more in paragraph 51). The toxic effect of the test chemical on reproductive output is expressed as ECx by fitting the data to an appropriate model by non-linear regression to estimate the concentration that would cause x % reduction in reproductive output, respectively, or alternatively as the NOEC/LOEC value (4). *The test concentrations should preferably bracket the lowest of the used effect concentrations (e.g. EC₁₀) which means that this value is calculated by interpolation and not extrapolation.*

The survival of the parent animals and time to production of first brood should also be reported. Other chemical-related effects on parameters such as growth (e.g. length), and possibly intrinsic rate of population increase, can also be examined (see paragraph 44).

INFORMATION ON THE TEST CHEMICAL

Results of an acute toxicity test (see chapter C.2 of this Annex: *Daphnia* sp. acute immobilisation test) performed with *Daphnia magna* may be useful in selecting an appropriate range of test concentrations in the reproduction tests. The water solubility and the vapour pressure of the test chemical should be known and a reliable analytical method for the quantification of the chemical in the test solutions with reported recovery efficiency and limit of determination should be available.

Information on the test chemical which may be useful in establishing the test conditions includes the structural formula, purity of the chemical, stability in light, stability under the conditions of the test, pKa, P_{ow} and results of a test for ready biodegradability (see chapters C.4 (determination of 'ready' biodegradability), and C.29 (ready biodegradability — CO₂ in sealed vessels) of this Annex).

VALIDITY OF THE TEST

For a test to be valid, the following performance criteria should be met in the control(s):

- the mortality of the parent animals (female *Daphnia*) does not exceed 20 % at the end of the test;
- the mean number of living offspring produced per parent animal surviving at the end of the test is ≥ 60 .

Note: The same validity criterion (20 %) can be used for accidental and inadvertent parental mortality for the controls as well as for each of the test concentrations.

DESCRIPTION OF THE METHOD

Apparatus

Test vessels and other apparatus, which will come into contact with the test solutions, should be made entirely of glass or other chemically inert material. The test vessels will normally be glass beakers.

In addition, some or all of the following equipment will be required:

- oxygen meter (with microelectrode or other suitable equipment for measuring dissolved oxygen in low volume samples);
- adequate apparatus for temperature control;
- pH-meter;
- equipment for the determination of the hardness of water;
- equipment for the determination of the total organic carbon concentration (TOC) of water or equipment for the determination of the chemical oxygen demand (COD);
- adequate apparatus for the control of the lighting regime and measurement of light intensity.

Test Organism

The species to be used in the test is *Daphnia magna* Straus ⁽¹⁾.

Preferably, the clone should have been identified by genotyping. Research (1) has shown that the reproductive performance of Clone A (which originated from IRCHA in France) (5) consistently meets the validity criterion of a mean of ≥ 60 living offspring per parent animal surviving when cultured under the conditions described in this test method. However, other clones are acceptable provided that the *Daphnia* culture is shown to meet the validity criteria for the test.

⁽¹⁾ Other daphnids may be used provided they meet the validity criteria as appropriate (the validity criterion relating to the reproductive output in the controls should be relevant for all species). If other *daphnid* are used they should be clearly identified and their use justified.

At the start of the test, the animals should be less than 24 hours old and should not be first brood progeny. They should be derived from a healthy stock (i.e. showing no signs of stress such as high mortality, presence of males and ephippia, delay in the production of the first brood, discoloured animals, etc.). The stock animals should be maintained in culture conditions (light, temperature, medium, feeding and animals per unit volume) similar to those to be used in the test. If the *Daphnia* culture medium to be used in the test is different from that used for routine *Daphnia* culture, it is good practice to include a pre-test acclimation period of normally about 3 weeks (i.e. one generation) to avoid stressing the parent animals.

Test medium

It is recommended that a fully defined medium be used in this test. This can avoid the use of additives (e.g. seaweed, soil extract), which are difficult to characterise, and therefore improves the opportunities for standardisation between laboratories. Elendt M4 (6) and M7 media (see Appendix 2) have been found to be suitable for this purpose. However, other media (e.g. (7) (8)) are acceptable provided the performance of the *Daphnia* culture is shown to meet the validity criteria for the test.

If media are used which include undefined additives, these additives should be specified clearly and information should be provided in the test report on composition, particularly with regard to carbon content as this may contribute to the diet provided. It is recommended that the total organic carbon (TOC) and/or chemical oxygen demand (COD) of the stock preparation of the organic additive be determined and an estimate of the resulting contribution to the TOC/COD in the test medium made. It is further recommended that TOC levels in the medium (i.e. before addition of the algae) be below 2 mg/l (9).

When testing chemicals containing metals, it is important to recognise that the properties of the test medium (e.g. hardness, chelating capacity) may have a bearing on the toxicity of the test chemical. For this reason, a fully defined medium is desirable. However, at present, the only fully defined media which are known to be suitable for long-term culture of *Daphnia magna* are Elendt M4 and M7. Both media contain the chelating agent EDTA. Work has shown (2) that the 'apparent toxicity' of cadmium is generally lower when the reproduction test is performed in M4 and M7 media than in media containing no EDTA. M4 and M7 are not, therefore, recommended for testing chemicals containing metals, and other media containing known chelating agents should also be avoided. For metal-containing chemicals it may be advisable to use an alternative medium such as, for example, ASTM reconstituted hard fresh water (9), which contains no EDTA. This combination of ASTM reconstituted hard fresh water and seaweed extract (10) is suitable for long-term culturing of *Daphnia magna* (2).

The dissolved oxygen concentration should be above 3 mg/l at the beginning and during the test. The pH should be within the range 6 - 9, and normally it should not vary by more than 1,5 units in any one test. Hardness above 140 mg/l (as CaCO₃) is recommended. Tests at this level and above have demonstrated reproductive performance in compliance with the validity criteria (11) (12).

Test solutions

Test solutions of the chosen concentrations are usually prepared by dilution of a stock solution. Stock solutions should preferably be prepared, without using any solvents or dispersants if possible, by mixing or agitating the test chemical in test medium using mechanical means such as agitating, stirring or ultrasonication, or other appropriate methods. It is preferable to expose test systems to concentrations of the test chemical to be used in the study for as long as is required to demonstrate the maintenance of stable exposure concentrations prior to the introduction of test organisms. If the test chemical is difficult to dissolve in water, procedures described in the OECD Guidance for handling difficult substances should be followed (13). The use of solvents or dispersants should be avoided, but may be necessary in some cases in order to produce a suitably concentrated stock solution for dosing.

A dilution water control with adequate replicates and, if unavoidable, a solvent control with adequate replicates should be run in addition to the test concentrations. Only solvents or dispersants that have been investigated to have no significant or only minimal effects on the response variable should be used in the test. Examples of suitable solvents (e.g. acetone, ethanol, methanol, dimethylformamide and triethylene glycol) and dispersants (e.g. Cremophor RH40, methylcellulose 0,01 % and HCO-40) are given in (13). Where a solvent or dispersant is used, its final concentration should not be greater than 0,1 ml/l (13) and it should be the same concentration in all test vessels, except the dilution water control. However, every effort should be made to keep the solvent concentration to a minimum.

PROCEDURE

Conditions of Exposure

Duration

The test duration is 21 days.

Loading

Parent animals are maintained individually, one per test vessel, usually with 50 - 100 ml (for *Daphnia magna*, smaller volumes may be possible especially for smaller daphnids e.g. *Ceriodaphnia dubia*) of medium in each vessel, unless a flow-through test design is necessary for testing.

Larger volumes may sometimes be necessary to meet requirements of the analytical procedure used for determination of the test chemical concentration, although pooling of replicates for chemical analysis is also allowable. If volumes greater than 100 ml are used, the ration given to the *Daphnia* may need to be increased to ensure adequate food availability and compliance with the validity criteria.

Test animals

For semi-static tests, at least 10 animals individually held at each test concentration and at least 10 animals individually held in the control series.

For flow-through tests, 40 animals divided into four groups of 10 animals at each test concentration has been shown to be suitable (1). A smaller number of test organisms may be used and a minimum of 20 animals per concentration divided into two or more replicates with an equal number of animals (e.g. four replicates each with five daphnids) is recommended. Note that for tests where animals are held in groups, it will not be possible to exclude any offspring from the statistical analysis if inadvertent/ accidental parental mortality occurs when the reproduction has begun, and hence in these cases the reproductive output should be expressed as total number of living offspring produced per parent present at the beginning of the test.

Treatments should be allocated to the test vessels and all subsequent handling of the test vessels should be done in a random fashion. Failure to do this may result in bias that could be construed as being a concentration effect. In particular, if experimental units are handled in treatment or concentration order, then some time-related effect, such as operator fatigue or other error, could lead to greater effects at the higher concentrations. Furthermore, if the test results are likely to be affected by an initial or environmental condition of the test, such as position in the laboratory, then consideration should be given to blocking the test.

Feeding

For semi-static tests, feeding should preferably be done daily, but at least three times per week (i.e. corresponding to media changes). The possible dilution of the exposure concentrations by food addition should be taken into account and avoided as much as possible with well concentrated algae suspensions. Deviations from this (e.g. for flow-through tests) should be reported.

During the test, the diet of the parent animals should preferably be living algal cells of one or more of the following: *Chlorella* sp., *Pseudokirchneriella subcapitata* (formerly *Selenastrum capricornutum*) and *Desmodesmus subspicatus* (formerly *Scenedesmus subspicatus*). The supplied diet should be based on the amount of organic carbon (C) provided to each parent animal. Research (14) has shown that, for *Daphnia magna*, ration levels of between 0,1 and 0,2 mg C/*Daphnia*/day are sufficient for achieving the required number of living offspring to meet the test validity criteria. The ration can be supplied either at a constant rate throughout the period of the test, or, if desired, a lower rate can be used at the beginning and then increased during the test to take account of growth of the parent animals. In this case, the ration should still remain within the recommended range of 0,1 - 0,2 mg C/*Daphnia*/day at all times.

If surrogate measures, such as algal cell number or light absorbance, are to be used to feed the required ration level (i.e. for convenience since measurement of carbon content is time consuming), each laboratory should produce its own nomograph relating the surrogate measure to carbon content of the algal culture (see Appendix 3 for advice on nomograph production). Nomographs should be checked at least annually and more frequently if algal culture conditions have changed. Light absorbance has been found to be a better surrogate for carbon content than cell number (15).

A concentrated algal suspension should be fed to the *Daphnia* to minimise the volume of algal culture medium transferred to the test vessels. Concentration of the algae can be achieved by centrifugation followed by re-suspension in *Daphnia* culture medium.

Light

16 hours light at an intensity not exceeding $15\text{--}20 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ measured at the water surface of the vessel. For light-measuring instruments calibrated in lux, an equivalent range of 1 000-1 500 lux for cool white light corresponds close to the recommended light intensity $15\text{--}20 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

Temperature

The temperature of the test media should be within the range 18-22 °C. However, for any one test, the temperature should not, if possible, vary by more than 2 °C within these limits (e.g. 18-20, 19-21 or 20-22 °C) as daily range. It may be appropriate to use an additional test vessel for the purposes of temperature monitoring.

Aeration

The test vessels should not be aerated during the test.

Test design

Range finding test

When necessary, a range-finding test is conducted with, for example five test chemical concentrations and two replicates for each treatment and control. Additional information, from tests with similar chemicals or from literature, on acute toxicity to *Daphnia* and/or other aquatic organisms may also be useful in deciding on the range of concentrations to be used in the range-finding test.

The duration of the range-finding test is 21 days or of a sufficient duration to reliably predict effect levels. At the end of the test, reproduction of the *Daphnia* is assessed. The number of parents and the occurrence of offspring should be recorded.

Definitive test

Normally there should be at least five test concentrations, bracketing effective concentration (e.g. EC_x), and arranged in a geometric series with a separation factor preferably not exceeding 3,2. An appropriate number of replicates for each test concentration should be used (see paragraphs 24-25). Justification should be provided if fewer than five concentrations are used. Chemicals should not be tested above their solubility limit in test medium. Before conducting the experiment it is advisable to consider the statistical power of the tests design and using appropriate statistical methods (4). In setting the range of concentrations, the following should be borne in mind:

- (i) When EC_x for effects on reproduction is estimated, it is advisable that sufficient concentrations are used to define the EC_x with an appropriate level of confidence. Test concentrations used should preferably bracket the estimated EC_x such that EC_x is found by interpolation rather than extrapolation. It is an advantage for the following statistical analysis to have more test concentrations (e.g. 10) and fewer replicates of each concentration (e.g. 5 thus holding the total number of vessels constant) and with 10 controls.
- (ii) When estimating the LOEC and/or NOEC, the lowest test concentration should be low enough so that the reproductive output at that concentration is not significantly lower than that in the control. If this is not the case, the test should be repeated with a reduced lowest concentration.
- (iii) When estimating the LOEC and/or NOEC, the highest test concentration should be high enough so that the reproductive output at that concentration is significantly lower than that in the control. If this is not the case, the test should be repeated with an increased highest concentration unless the maximum required test concentration for chronic effects testing (i.e., 10 mg/l) was used as the highest test concentration in the initial test.

If no effects are observed at the highest concentration in the range-finding test (e.g. at 10 mg/l), or when the test chemical is highly likely to be of low/ no toxicity based on lack of toxicity to other organisms and/or low/no uptake, the reproduction test may be performed as a limit test, using a test concentration of e.g. 10 mg/l and the control. Ten replicates should be used for both the treatment and the control groups. When a limit test might need to be done in a flow-through system less replicates would be adequate. A limit test will provide the opportunity to demonstrate that there is no statistically significant effect at the limit concentration, but if effects are recorded a full test will normally be required.

Controls

One test-medium control series and also, if relevant, one control series containing the solvent or dispersant should be run in addition to the test series. When used, the solvent or dispersant concentration should be the same as that used in the vessels containing the test chemical. The appropriate number of replicates should be used (see paragraphs 23-24).

Generally in a well-run test, the coefficient of variation around the mean number of living offspring produced per parent animal in the control(s) should be $\leq 25\%$, and this should be reported for test designs using individually held animals.

Test medium renewal

The frequency of medium renewal will depend on the stability of the test chemical, but should be at least three times per week. If, from preliminary stability tests (see paragraph 7), the test chemical concentration is not stable (i.e. outside the range 80 - 120 % of nominal or falling below 80 % of the measured initial concentration) over the maximum renewal period (i.e. 3 days), consideration should be given to more frequent medium renewal, or to the use of a flow-through test.

When the medium is renewed in semi-static tests, a second series of test vessels are prepared and the parent animals transferred to them by, for example, a glass pipette of suitable diameter. The volume of medium transferred with the *Daphnia* should be minimised.

Observations

The results of the observations made during the test should be recorded on data sheets (see examples in Appendixes 4 and 5). If other measurements are required (see paragraph 44), additional observations may be required.

Offspring

The offspring produced by each parent animal should preferably be removed and counted daily from the appearance of the first brood to prevent them consuming food intended for the parent. For the purpose of this test method it is only the number of living offspring that needs to be counted, but the presence of aborted eggs or dead offspring should be recorded.

Mortality

Mortality among the parent animals should be recorded preferably daily, or at least as frequently as offspring are counted.

Other parameters

Although this test method is designed principally to assess effects on reproductive output, it is possible that other effects may also be sufficiently quantified to allow statistical analysis. Reproductive output per surviving parent animal, i.e. number of living offspring produced during the test per surviving parent, may be recorded. This may be compared with the main response variable (reproductive output per parent animal in the start of the test which did not inadvertently or accidentally die during the test). If parental mortality occurs in exposed replicates it should be considered whether or not the mortality follows a concentration-response pattern, e.g. if there is a significant regression of the response versus concentration of the test chemical with a positive slope (a statistical test like the Cochran-Armitage trend test may be used for this). If the mortality does not follow a concentration-response pattern, then those replicates with parental mortality should be excluded from the analysis of the test result. If the mortality follows a concentration-response pattern, the parental mortality should be assigned as an effect of the test chemical and the replicates should not be excluded from the analysis of the test result. Growth measurements are highly desirable since they provide information on possible sublethal effects which may be useful in addition to reproduction measures alone; the measurement of the length of the parent animals (i.e. body length excluding the anal spine) at the end of the test is recommended. Other parameters that can be measured or calculated include time to production of first brood (and subsequent broods), number and size of broods per animal, number of aborted broods, presence of male neonates (OECD, 2008) or ephippia and possibly the intrinsic rate of population increase (see Appendix 1 for definition and Appendix 7 for the identification of the sex of neonates).

Frequency of analytical determinations and measurements

Oxygen concentration, temperature, hardness and pH values should be measured at least once a week, in fresh and old media, in the control(s) and in the highest test chemical concentration.

During the test, the concentrations of test chemical are determined at regular intervals.

In semi-static tests where the concentration of the test chemical is expected to remain within ± 20 per cent of the nominal (i.e. within the range 80 - 120 per cent- see paragraphs 6, 7 and 39), it is recommended that, as a minimum, the highest and lowest test concentrations be analysed when freshly prepared and at the time of renewal on one occasion during the first week of the test (i.e. analyses should be made on a sample from the same solution — when freshly prepared and at renewal). These determinations should be repeated at least at weekly intervals thereafter.

For tests where the concentration of the test chemical is not expected to remain within ± 20 per cent of the nominal, it is necessary to analyse all test concentrations, when freshly prepared and at renewal. However, for those tests where the measured initial concentration of the test chemical is not within ± 20 per cent of nominal but where sufficient evidence can be provided to show that the initial concentrations are repeatable and stable (i.e. within the range 80 - 120 per cent of initial concentrations), chemical determinations could be reduced in weeks 2 and 3 of the test to the highest and lowest test concentrations. In all cases, determination of test chemical concentrations prior to renewal need only be performed on one replicate vessel at each test concentration.

If a flow-through test is used, a similar sampling regime to that described for semi-static tests is appropriate (but measurement of 'old' solutions is not applicable in this case). However, it may be advisable to increase the number of sampling occasions during the first week (e.g. three sets of measurements) to ensure that the test concentrations are remaining stable. In these types of test, the flow-rate of diluent and test chemical should be checked daily.

If there is evidence that the concentration of the chemical being tested has been satisfactorily maintained within ± 20 per cent of the nominal or measured initial concentration throughout the test, then results can be based on nominal or measured initial values. If the deviation from the nominal or measured initial concentration is greater than ± 20 per cent, results should be expressed in terms of the time-weighted mean (see guidance for calculation in Appendix 6).

DATA AND REPORTING

Treatment of results

The purpose of this test is to determine the effect of the test chemical on the reproductive output. The total number of living offspring per parent animal should be calculated for each test vessel (i.e. replicate). In addition, the reproduction can be calculated based on the production of living offspring by the surviving parent organism. However, the ecologically most relevant response variable is the total number of living offspring produced per parent animal which does not die accidentally⁽¹⁾ or inadvertently⁽²⁾ during the test. If the parent animal dies accidentally or inadvertently during the test, or turns out to be male, then the replicate is excluded from the analysis. The analysis will then be based on a reduced number of replicates. If parental mortality occurs in exposed replicates it should be considered whether or not the mortality follows a concentration-response pattern, e.g. if there is a significant regression of the response versus concentration of the test chemical with a positive slope (a statistical test like the Cochran-Armitage trend test may be used for this). If the mortality does not follow a concentration-response pattern, then those replicates with parental mortality should be excluded from the analysis of the test result. If the mortality follows a concentration-response pattern, the parental mortality should be assigned as an effect of the test chemical and the replicates should not be excluded from the analysis of the test result.

In summary, when LOEC and NOEC or EC_x are being used to express the effects, it is recommended to calculate the effect on reproduction by the use of both response variables mentioned above i.e.

- as the total number of living offspring produced per parent animal which does not die accidentally or inadvertently during the test and;
- as the number of living offspring produced per surviving parental animal;

and then to use as the final result the lowest NOEC and LOEC or EC_x value calculated by using either of these two response variables.

Before employing the statistical analysis, e.g. ANOVA procedures, comparison of treatments to the control by Student t-test, Dunnett's test, Williams' test, or stepdown Jonckheere-Terpstra test, it is recommended to consider transformation of data if needed for meeting the requirements of the particular statistical test. As non-parametric alternatives one can consider Dunn's or Mann-Whitney's tests. 95 % confidence intervals are calculated for individual treatment means.

⁽¹⁾ Accidental mortality: non chemical related mortality caused by an accidental incidence (i.e. known cause)

⁽²⁾ Inadvertent mortality: non chemical related mortality with no known cause

The number of surviving parents in the untreated controls is a validity criterion, and should be documented and reported. Also all other detrimental effects, e.g. abnormal behavior and toxicological significant findings, should be reported in the final report as well.

ECx

ECx-values, including their associated lower and upper confidence limits, are calculated using appropriate statistical methods (e.g. logistic or Weibull function, trimmed Spearman-Kärber method, or simple interpolation). To compute the EC10, EC₅₀ or any other ECx, the complete data set should be subjected to regression analysis.

NOEC/LOEC

If a statistical analysis is intended to determine the NOEC/LOEC appropriate statistical methods should be used according to OECD Document 54 on the Current Approaches in the Statistical Analysis of Ecotoxicity Data: a Guidance to Application (4). In general, adverse effects of the test chemical compared to the control are investigated using one-tailed hypothesis testing at $p \leq 0,05$.

Normal distribution and variance homogeneity can be tested using an appropriate statistical test, e.g. the Shapiro-Wilk test and Levene test, respectively ($p \leq 0,05$). One-way ANOVA and subsequent multi-comparison tests can be performed. Multiple comparisons (e.g. Dunnett's test) or step-down trend tests (e.g. Williams' test, or stepdown Jonckheere-Terpstra test) can be used to calculate whether there are significant differences ($p \leq 0,05$) between the controls and the various test chemical concentrations (selection of the recommended test according to OECD Guidance Document 54 (4)). Otherwise, non-parametric methods (e.g. Bonferroni-U-test according to Holm or Jonckheere-Terpstra trend test) could be used to determine the NOEC and the LOEC.

Limit test

If a limit test (comparison of control and one treatment only) has been performed and the prerequisites of parametric test procedures (normality, homogeneity) are fulfilled, metric responses can be evaluated by the Student test (t-test). An unequal-variance t-test (such as Welch test) or a non-parametric test such as the Mann-Whitney-U-test may be used, if these requirements are not fulfilled.

To determine significant differences between the controls (control and solvent or dispersant control), the replicates of each control can be tested as described for the limit test. If these tests do not detect significant differences, all control and solvent control replicates may be pooled. Otherwise all treatments should be compared with the solvent control.

Test report

The test report includes the following:

Test chemical:

- physical nature and relevant physicochemical properties;
- chemical identification data, including purity.

Test species:

- the clone (whether it has been genetically typed), supplier or source (if known) and the culture conditions used. If a different species to *Daphnia magna* is used, this should be reported and justified.

Test conditions:

- test procedure used (e.g. semi-static or flow-through, volume, loading in number of *Daphnia* per litre);

- photoperiod and light intensity;
- test design (e.g. number of replicates, number of parents per replicate);
- details of culture medium used;
- if used, additions of organic material including the composition, source, method of preparation, TOC/COD of stock preparations, estimation of resulting TOC/COD in test medium;
- detailed information on feeding, including amount (in mg C/*daphnia*/day) and schedule (e.g. type of food(s), including, for algae the specific name (species) and, if known, the strain, the culture conditions);
- method of preparation of stock solutions and frequency of renewal (the solvent or dispersant and its concentration should be given, when used).

Results:

- results from any preliminary studies on the stability of the test chemical;
- the nominal test concentrations and the results of all analyses to determine the concentration of the test chemical in the test vessels (see example data sheets in Appendix 5); the recovery efficiency of the method and the limit of determination should also be reported;
- water quality within the test vessels (i.e. pH, temperature and dissolved oxygen concentration, and TOC and/or COD and hardness where applicable) (see example data sheet in Appendix 4);
- the full record of the production of living offspring during the test by each parent animal (see example data sheet in Appendix 4);
- the number of deaths among the parent animals and the day on which they occurred (see example data sheet in Appendix 4);
- the coefficient of variation for control reproductive output (based on total number of living offspring per parent animal alive at the end of the test);
- plot of total number of living offspring produced per parent animal in each replicate excluding any parent animal which may have accidentally or inadvertently died during the test vs. concentration of the test chemical;
- as appropriate plot of total number of living offspring produced per surviving parent animal in each replicate vs. concentration of the test chemical
- where appropriate the Lowest Observed Effect Concentration (LOEC) for reproduction, including a description of the statistical procedures used and an indication of what size of effect could be expected to be detected (a power analysis can be performed before the start of the experiment to provide this) and the No Observed Effect Concentration (NOEC) for reproduction; information on which response variable that has been used for calculating the LOEC and NOEC value (either as total living offspring per maternal organism which did not die accidentally or inadvertently during the test or as total number of living offspring per surviving maternal organism), where appropriate, the LOEC or NOEC for mortality of the parent animals should also be reported;
- where appropriate, the EC_x for reproduction 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 and its standard error;
- other observed biological effects or measurements: report any other biological effects which were observed or measured (e.g. growth of parent animals) including any appropriate justification;
- an explanation for any deviation from the test method.

LITERATURE:

- (1) OECD Test Guidelines Programme. Report of the Workshop on the *Daphnia magna* Pilot Ring Test, Sheffield University, U.K., 20-21 March 1993.
 - (2) OECD (1997). Report of the Final Ring Test of the *Daphnia magna* Reproduction Test. Environmental Health and Safety Publications, Series on Testing and Assessment No.6. OECD, Paris.
 - (3) OECD (2008). Validation report for an enhancement of OECD TG 211 *Daphnia magna* reproduction test. Environmental Health and Safety Publications, Series on Testing and Assessment, No.88. OECD, Paris.
 - (4) OECD (2006). Current approaches in the statistical analysis of ecotoxicity data: a guidance to application. Environmental Health and Safety Publications, Series on Testing and Assessment Number 54. OECD, Paris.
 - (5) Baird, D.J., *et al.* (1991). A comparative study of genotype sensitivity to acute toxic stress using clones of *Daphnia magna* Straus. *Ecotox. and Environ. Safety*, 21, 257-265.
 - (6) Elendt, B.-P. (1990). Selenium deficiency in Crustacea; An ultrastructural approach to antennal damage in *Daphnia magna* Straus. *Protoplasma*, 154, 25-33.
 - (7) EPA (2002). Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms. Fifth Edition. EPA/821/R-02/012. U.S. Environmental Protection Agency, Office of Water, Washington, DC. www.epa.gov/waterscience/methods
 - (8) Vigano, L. (1991). Suitability of commercially available spring waters as standard medium for culturing *Daphnia magna*. *Bull. Environ. Contam. Toxicol.*, 47, 775-782.
 - (9) ASTM. (2008) Standard Guide for Conducting Acute Toxicity Tests with Fishes, Macroinvertebrates, and Amphibians. In: Annual Book of ASTM Standards; Water and Environmental Technology, vol. 11.04; ASTM E729 — 96 (2007) American Society for Testing and Materials, Philadelphia, PA
 - (10) Baird, D.J., *et al.* (1989). The long term maintenance of *Daphnia magna* Straus for use in ecotoxicological tests; problems and prospects. In: Proceedings of the 1st European Conference on Ecotoxicology. Copenhagen 1988. (H. Løkke, H. Tyle and F. Bro-Rasmussen. Eds.) pp 144-148.
 - (11) Parkhurst, B.R., J.L. Forte. And G.P. and Wright (1981) Reproducibility of a life-cycle toxicity test with *Daphnia magna*. *Bull. Environ. Contam. and Toxicol.*, 26: 1-8.
 - (12) Cowgill, U.M. and Milazzo, D.P. (1990). The sensitivity of two cladocerans to water quality variables: salinity and hardness. *Arch. Hydrobiol.*, 120(2): 185-196.
 - (13) OECD (2000), *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures*, Environmental Health and Safety Publications, Series on Testing and Assessment No. 23. OECD, Paris.
 - (14) Sims, I.R., S. Watson. and D. Holmes (1993) Toward a standard *Daphnia* juvenile production test. *Environ. Toxicol. and Chem.*, 12, 2053-2058.
 - (15) Sims, I. (1993). Measuring the growth of phytoplankton: the relationship between total organic carbon with three commonly used parameters of algal growth. *Arch. Hydrobiol.*, 128, 459-466.
-

Appendix 1

DEFINITIONS:

For the purposes of this test method the following definitions are used:

Accidental mortality: non chemical related mortality caused by an accidental incidence (i.e. known cause).

Chemical: a substance or mixture.

EC_x: the concentration of the test chemical dissolved in water that results in a x per cent reduction in reproduction of *Daphnia* within a stated exposure period.

Inadvertent mortality: non chemical related mortality with no known cause.

Intrinsic rate of population increase: a measure of population growth which integrates reproductive output and age-specific mortality (1) (2) (3). In steady state populations it will be zero. For growing populations it will be positive and for shrinking populations it will be negative. Clearly the latter is not sustainable and ultimately will lead to extinction.

Limit of detection: the lowest concentration that can be detected but not quantified.

Limit of determination: the lowest concentration that can be measured quantitatively.

Lowest Observed Effect Concentration (LOEC): the lowest tested concentration at which the chemical is observed to have a statistically significant effect on reproduction and parent mortality (at $p < 0,05$) when compared with the control, within a stated exposure period. 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.

Mortality: an animal is recorded as dead when it is immobile, i.e. when it is not able to swim, or if there is no observed movement of appendages or postabdomen, within 15 seconds after gentle agitation of the test container. (If another definition is used, this should be reported together with its reference).

No Observed Effect Concentration (NOEC): 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.

Offspring: the young *Daphnia* produced by the parent animals in the course of the test.

Parent Animals: those female *Daphnia* present at the start of the test and of which the reproductive output is the object of study.

Reproductive output: the number of living offspring produced by parental animals within the test period

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

LITERATURE

- (1) Wilson, E.O. and Bossert, W.H. (1971). A Primer of Population Biology. Sinauer Associates Inc. Publishers.
- (2) Poole, R.W. (1974). An Introduction to quantitative Ecology. Mc Graw Hill Series in Population Biology, New York, p 532.
- (3) Meyer, J. S., Ingersoll, C. G., McDonald, L.L. and Boyce, M.S. (1986). Estimating uncertainty in population growth rates: Jackknife vs bootstrap techniques. Ecology, 67, 1156-1166.

Appendix 2

PREPARATION OF FULLY DEFINED ELENDET M7 AND M4 MEDIA

Acclimation to Elendt M7 and M4 media

Some laboratories have experienced difficulty in directly transferring *Daphnia* to M4 (1) and M7 media. However, some success has been achieved with gradual acclimation, i.e. moving from own medium to 30 % Elendt, then to 60 % Elendt and then to 100 % Elendt. The acclimation periods may need to be as long as one month.

Preparation

Trace elements

Separate stock solutions (I) of individual trace elements are first prepared in water of suitable purity, e.g. deionised, distilled or reverse osmosis. From these different stock solutions (I) a second single stock solution (II) is prepared, which contains all trace elements (combined solution), i.e:

Stock solution(s) I (single substance)	Amount added to water	Concentration (related to medium M4)	To prepare the combined stock-solution II add the following amount of stock solution I to water	
			ml/l	
	mg/l		M 4	M 7
H ₃ BO ₃	57 190	20 000-fold	1,0	0,25
MnCl ₂ · 4 H ₂ O	7 210	20 000-fold	1,0	0,25
LiCl	6 120	20 000-fold	1,0	0,25
RbCl	1 420	20 000-fold	1,0	0,25
SrCl ₂ · 6 H ₂ O	3 040	20 000-fold	1,0	0,25
NaBr	320	20 000-fold	1,0	0,25
Mo Na ₂ O ₄ · 2 H ₂ O	1 260	20 000-fold	1,0	0,25
CuCl ₂ · 2 H ₂ O	335	20 000-fold	1,0	0,25
ZnCl ₂	260	20 000-fold	1,0	1,0
CoCl ₂ · 6 H ₂ O	200	20 000-fold	1,0	1,0
KI	65	20 000-fold	1,0	1,0
Na ₂ SeO ₃	43,8	20 000-fold	1,0	1,0
NH ₄ VO ₃	11,5	20 000-fold	1,0	1,0
Na ₂ EDTA · 2 H ₂ O	5 000	2 000-fold	—	—
FeSO ₄ · 7 H ₂ O	1 991	2 000-fold	—	—
Both Na ₂ EDTA and FeSO ₄ solutions are prepared singly, poured together and autoclaved immediately. This gives:				
Fe-EDTA solution		1 000-fold	20,0	5,0

M4 and M7 media

M4 and M7 media are prepared using stock solution II, the macro-nutrients and vitamins as follows:

	Amount added to water	Concentration (related to medium M4)	Amount of stock solution added to prepare medium	
	mg/l		ml/l	
			M 4	M 7
Stock solution II (combined trace elements)		20-fold	50	50
Macro nutrient stock solutions (single substance)				
$\text{CaCl}_2 \cdot 2 \text{H}_2\text{O}$	293 800	1 000-fold	1,0	1,0
$\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$	246 600	2 000-fold	0,5	0,5
KCl	58 000	10 000-fold	0,1	0,1
NaHCO_3	64 800	1 000-fold	1,0	1,0
$\text{Na}_2\text{SiO}_3 \cdot 9 \text{H}_2\text{O}$	50 000	5 000-fold	0,2	0,2
NaNO_3	2 740	10 000-fold	0,1	0,1
KH_2PO_4	1 430	10 000-fold	0,1	0,1
K_2HPO_4	1 840	10 000-fold	0,1	0,1
Combined Vitamin stock	—	10 000-fold	0,1	0,1

The combined vitamin stock solution is prepared by adding the 3 vitamins to 1 litre water, as shown below:

	mg/l			
Thiamine hydrochloride	750	10 000-fold		
Cyanocobalamine (B_{12})	10	10 000-fold		
Biotine	7,5	10 000-fold		

The combined vitamin stock is stored frozen in small aliquots. Vitamins are added to the media shortly before use.

N.B: To avoid precipitation of salts when preparing the complete media, add the aliquots of stock solutions to about 500 - 800 ml deionized water and then fill it up to 1 litre.

N.N.B. The first publication of the M4 medium can be found in Elendt, B.P. (1990). Selenium deficiency in crustacea; an ultrastructural approach to antennal damage in *Daphnia magna* Straus. *Protoplasma*, 154, 25-33.

Appendix 3

TOTAL ORGANIC CARBON (TOC) ANALYSIS AND PRODUCTION OF A NOMOGRAPH FOR TOC CONTENT OF ALGAL FEED

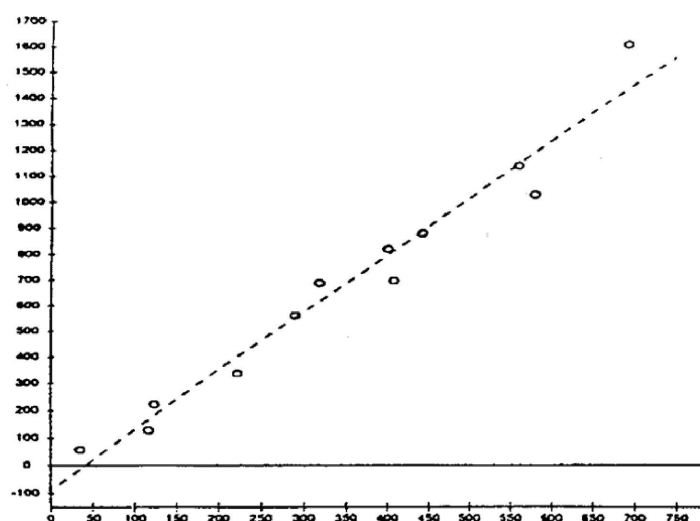
It is recognised that the carbon content of the algal feed will not normally be measured directly but from correlations (i.e. nomographs) with surrogate measures such as algal cell number or light absorbance).

TOC should be measured by high temperature oxidation rather than by UV or persulphate methods. (For advice see: The Instrumental Determination of Total Organic Carbon, Total Oxygen Demand and Related Determinands 1979, HMSO 1980; 49 High Holborn, London WC1V 6HB).

For nomograph production, algae should be separated from the growth medium by centrifugation followed by resuspension in distilled water. Measure the surrogate parameter and TOC concentration in each sample in triplicate. Distilled water blanks should be analysed and the TOC concentration deducted from that of the algal sample TOC concentration.

Nomographs should be linear over the required range of carbon concentrations. Examples are shown below.

N.B. These should not be used for conversions; it is essential that laboratories prepare their own nomographs.



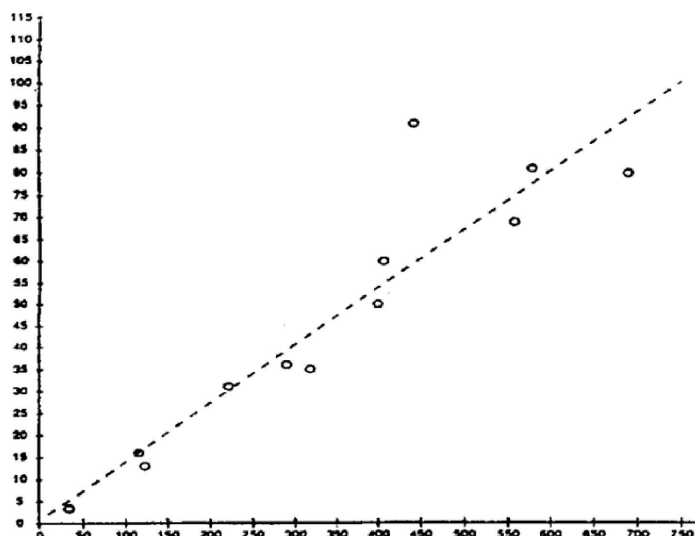
Chlorella vulgaris var. *viridis* (CCAP 211/12).

Regression of mg/l dry weight on mg C/l. Data from concentrated suspensions of semi continuous batch cultured cells, re-suspended in distilled water.

x-axis: mg C/l of concentrated algal feed

y-axis: mg/l dry weight of concentrated algal feed

Correction coefficient – 0,980



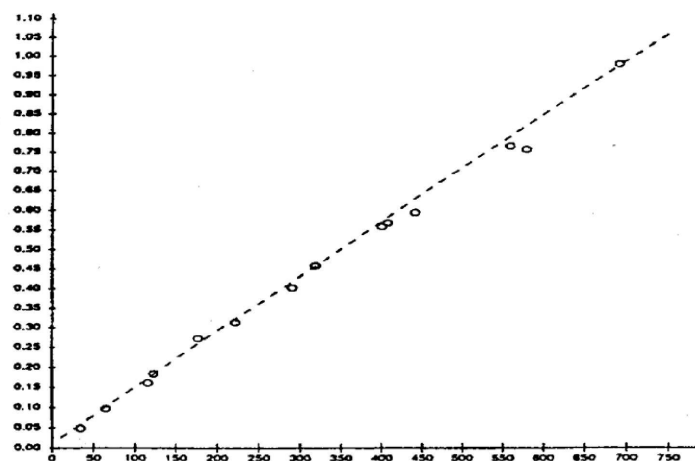
Chlorella vulgaris var. *viridis* (CCAP 211/12).

Regression of cell number on mg C/1. Data from concentrated suspensions of semi continuous batch cultured cells, re-suspended in distilled water.

x-axis: mg C/1 of concentrated algal feed

y-axis: No. cells/1 of concentrated algal feed

Correction coefficient – 0,926



Chlorella vulgaris var. *viridis* (CCAP 211/12).

Regression of absorbance on mg C/1 (1 cm path length). Data from concentrated suspensions of semi continuous batch cultured cells, re-suspended in distilled water.

x-axis: mg C/1 of concentrated algal feed

y-axis: Absorbance at 440 nm of a 1/10 dilution of concentrated algal feed

Correction coefficient – 0,998

Appendix 4

EXAMPLE DATA SHEET FOR RECORDING MEDIUM RENEWAL, PHYSICAL/CHEMICAL MONITORING DATA, FEEDING, DAPHNIA REPRODUCTION AND PARENT MORTALITY

Experiment No:	Date started:					Clone:			Medium:			Type of food:			Test Chemical:			Nominal conc:						
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21		
Medium renewal (tick)																								
pH (*)																								new
																								old
O ₂ (mg/l) (*)																								new
																								old
Temp (°C) (*)																								new
																								old
Food provided (tick)																								
No. live offspring (**)																								Total
Vessel 1																								
2																								
3																								
4																								
5																								
6																								
7																								
8																								
9																								
10																								
																								Total
Cumulative parent mortality (***)																								

(*) Indicate which vessel was used for the experiment
 (**) Record aborted broods as 'AB' in relevant box
 (***) Record mortality of any parental animals as 'M' in relevant box

Appendix 5

EXAMPLE DATA SHEET FOR RECORDING RESULTS OF CHEMICAL ANALYSIS

(a) **Measured concentrations**

[illegible]

(b) Measured concentrations as a percentage of nominal

[illegible]

Appendix 6

CALCULATION OF A TIME-WEIGHTED MEAN

Time-weighted mean

Given that the concentration of the test chemical can decline over the period between medium renewals, it is necessary to consider what concentration should be chosen as representative of the range of concentrations experienced by the parent *Daphnia*. The selection should be based on biological considerations as well as statistical ones. For example, if reproduction is thought to be affected mostly by the peak concentration experienced, then the maximum concentration should be used. However, if the accumulated or longer term effect of the toxic chemical is considered to be more important, then an average concentration is more relevant. In this case, an appropriate average to use is the time-weighted mean concentration, since this takes account of the variation in instantaneous concentration over time.

Figure 1

Example of time-weighted mean

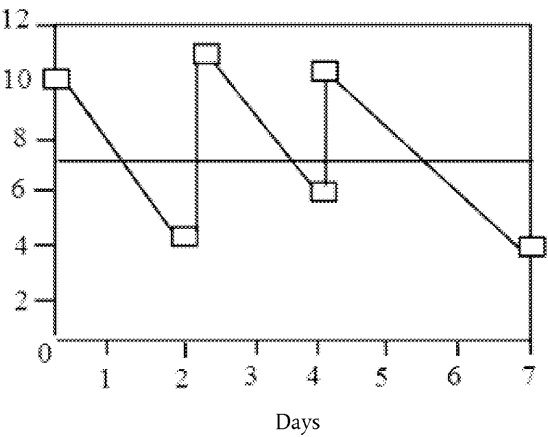


Figure 1 shows an example of a (simplified) test lasting seven days with medium renewal at Days 0, 2 and 4.

- The thin zig-zag line represents the concentration at any point in time. The fall in concentration is assumed to follow an exponential decay process.
- The 6 plotted points represent the observed concentrations measured at the start and end of each renewal period.
- The thick solid line indicates the position of the time-weighted mean.

The time-weighted mean is calculated so that the area under the time-weighted mean is equal to the area under the concentration curve. The calculation for the above example is illustrated in Table 1.

Table 1

Calculation of Time-weighted mean

Renewal No.	Days	Conc 0	Conc 1	Ln(Conc 0)	Ln(Conc 1)	Area
1	2	10,000	4,493	2,303	1,503	13,767
2	2	11,000	6,037	2,398	1,798	16,544

Renewal No.	Days	Conc 0	Conc 1	Ln(Conc 0)	Ln(Conc 1)	Area
3	3	10,000	4,066	2,303	1,403	19,781
Total Days:		7			Total Area:	50,092
					TW Mean:	7,156

Days is the number of days in the renewal period

Conc 0 is the measured concentration at the start of each renewal period

Conc 1 is the measured concentration at the end of each renewal period

Ln(Conc 0) is the natural logarithm of *Conc 0*

Ln(Conc 1) is the natural logarithm of *Conc 1*

Area is the area under the exponential curve for each renewal period. It is calculated by:

$$\text{Area} = \frac{\text{Conc 0} - \text{Conc 1}}{\text{Ln}(\text{Conc 0}) - \text{Ln}(\text{Conc 1})} \times \text{Day}$$

The time-weighted mean (*TW Mean*) is the *Total Area* divided by the *Total Days*.

Of course, for the *Daphnia* reproduction test the table should be extended to cover 21 days.

It is clear that when observations are taken only at the start and end of each renewal period, it is not possible to confirm that the decay process is, in fact, exponential. A different curve would result in a different calculation for *Area*. However, an exponential decay process is not implausible and is probably the best curve to use in the absence of other information.

However, a word of caution is required if the chemical analysis fails to find any chemical at the end of the renewal period. Unless it is possible to estimate how quickly the chemical disappeared from the solution, it is impossible to obtain a realistic area under the curve, and hence it is impossible to obtain a reasonable time-weighted mean.

Appendix 7

GUIDANCE FOR THE IDENTIFICATION OF NEONATE SEX

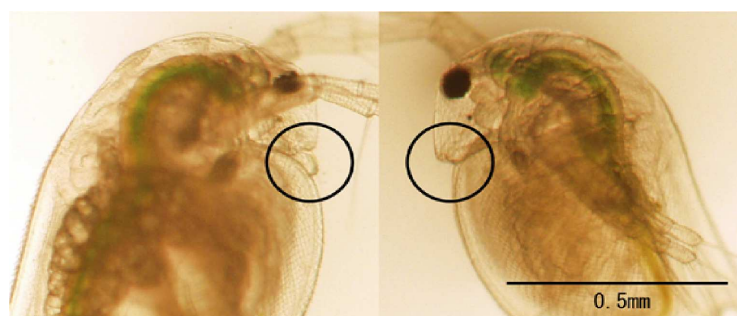
Production of male neonates can occur under changing environmental conditions, such as shortening photoperiod, temperature, decreasing food concentration, and increasing population density (Hobaek and Larson, 1990; Kleiven *et al.*, 1992). Male production is also a known response to certain insect growth regulators (Oda *et al.*, 2005). Under conditions where chemical stressors are inducing a decrease in reproductive offspring from the parthenogenic females, an increased number of males would be expected (OECD, 2008). On the basis of available information, it is not possible to predict which of the sex ratio or of the reproduction endpoint will be more sensitive; however, there are indications (reference “validation report”, part 1) this increase in the number of males might be less sensitive than the decrease in offspring. Since the primary purpose of the test method is to assess the number of offspring produced, the appearance of males is an optional observation. If this optional endpoint is evaluated in a study, then an additional test validity criterion of no more than 5 % males in the controls should be employed.

The most practical and easy way to differentiate sex of *Daphnia* is to use their phenotypic characteristics, as males and females are genetically identical and their sex is environmentally determined. Males and females are different in the length and morphology of the first antennae, which are longer in males than females (Fig. 1). This difference is recognizable right after birth, although other secondary sex characteristics develop as they grow up (e.g. see Fig. 2 in Olmstead and LeBlanc, 2000).

To observe the morphological sex, neonates produced by each test animal should be transferred by pipet and placed into a petri dish with test medium. The medium is kept to a minimum to restrain movement of the animals. Observation of the first antennae can be conducted under a stereomicroscope ($\times 10$ -60).

Figure 1

24-hour-old male (left) and female (right) of *D. magna*. Males can be distinguished from females by the length and morphology of the first antennae as shown in the circles (Tatarazako *et al.*, 2004)



REFERENCES:

- Hobaek A and Larson P. 1990. Sex determination in *Daphnia magna*. Ecology 71: 2255-2268.
- Kleiven O.T., Larsson P., Hobaek A. 1992. Sexual reproduction in *Daphnia magna* requires three stimuli. Oikos 65, 197-206.
- Oda S., Tatarazako N, Watanabe H., Morita M., and Iguchi T. 2005. Production of male neonates in *Daphnia magna* (Cladocera, Crustacea) exposed to juvenile hormones and their analogs. Chemosphere 61:1168-1174.
- OECD, 2008. Validation report for an enhancement of OECD TG 211 *Daphnia magna* reproduction test. OECD Series on Testing and Assessment, Number 88. Organisation for Economic Co-operation and Development, Paris.

Olmstead, A.W., LeBlanc, G.A., 2000. Effects of endocrine-active chemicals on the development characteristics of *Daphnia magna*. Environmental Toxicology and Chemistry 19:2107-2113.

Tatarazako, N., Oda, S., Abe, R., Morita M. and Iguchi T., 2004. Development of a screening method for endocrine disruptors in crustaceans using *Daphnia magna* (Cladocera, Crustacea). Environmental Science 17, 439-449.'

(18) In Part C, Chapter C.29, paragraph 66 is replaced by the following:

'66. A test is considered valid if:

- (a) the mean percentage degradation in vessels F_C containing the reference chemical is $> 60\%$ by the 14th day of incubation; and
- (b) the mean amount of TIC present in the blank controls F_B at the end of the test is $< 3\text{mg C/l}$.

If these limits are not met, the test should be repeated with an inoculum from another source and/or the procedures used should be reviewed. For example, if high blank IC production is a problem the procedure given in paragraphs 27 to 32 should be followed.'

(19) In Part C, the following Chapters are added:

'C.47 Fish, Early-life Stage Toxicity Test

INTRODUCTION

1. This test method is equivalent to OECD test guideline (TG) 210 (2013). Tests with the early-life stages of fish are intended to define the lethal and sub-lethal effects of chemicals on the stages and species tested. They yield information of value for the estimation of the chronic lethal and sub-lethal effects of the chemical on other fish species.
2. Test guideline 210 is based on a proposal from the United Kingdom which was discussed at a meeting of OECD experts convened at Medmenham (United Kingdom) in November 1988 and further updated in 2013 to reflect experience in using the test and recommendations from an OECD workshop on fish toxicity testing, held in September 2010 (1).

PRINCIPLE OF THE TEST

3. The early-life stages of fish are exposed to a range of concentrations of the test chemical dissolved in water. Flow-through conditions are preferred; however, if it is not possible semi-static conditions are acceptable. For details the OECD guidance document on aquatic toxicity testing of difficult substances and mixtures should be consulted (2). The test is initiated by placing fertilised eggs in test chambers and is continued for a species-specific time period that is necessary for the control fish to reach a juvenile life-stage. Lethal and sub-lethal effects are assessed and compared with control values to determine the lowest observed effect concentration (LOEC) in order to determine the (i) no observed effect concentration (NOEC) and/or (ii) EC_x (e.g. EC₁₀, EC₂₀) by using a regression model to estimate the concentration that would cause a $x\%$ change in the effect measured. Reporting of relevant effect concentrations and parameters may depend upon the regulatory framework. The test concentrations should bracket the EC_x so that the EC_x comes from interpolation rather than extrapolation (see Appendix 1 for definitions).

INFORMATION ON THE TEST CHEMICAL

4. Test chemical refers to what is being tested. The water solubility (see chapter A.6 of this Annex) and the vapour pressure (see chapter A.4 of this Annex) of the test chemical should be known and a reliable analytical method for the quantification of the chemical in the test solutions with known and reported accuracy and limit of quantification should be available. Although not necessary to conduct the test, results from an acute toxicity test (see chapters C.1 or C.49 of this Annex), preferably performed with the species chosen for this test, may provide useful information.

5. If the test method is used for the testing of a mixture, its composition should as far as possible be characterised, e.g. by the chemical identity of its constituents, their quantitative occurrence and their substance-specific properties (like those mentioned above). 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.
6. Useful information includes the structural formula, purity of the substance, water solubility, stability in water and light, pK_a , P_{ow} and results of a test for ready biodegradability (e.g. chapters C.4 or C.29 of this Annex).

VALIDITY OF THE TEST

7. For a test to be valid the following conditions apply:
 - the dissolved oxygen concentration should be > 60 % of the air saturation value throughout the test;
 - the water temperature should not differ by more than $\pm 1,5$ °C between test chambers or between successive days at any time during the test, and should be within the temperature ranges specified for the test species (Appendix 2);
 - the analytical measure of the test concentrations is compulsory.
 - overall survival of fertilised eggs and post-hatch success in the controls and, where relevant, in the solvent controls should be greater than or equal to the limits defined in Appendix 2.
8. If a minor deviation from the validity criteria is observed, the consequences should be considered in relation to the reliability of the test data and these considerations should be included in the report. Effects on survival, hatch or growth occurring in the solvent control, when compared to the negative control, should be reported and discussed in the context of the reliability of the test data.

DESCRIPTION OF THE METHOD

Test chambers

9. Any glass, stainless steel or other chemically inert vessels can be used. As silicone is known to have a strong capacity to absorb lipophilic substances, the use of silicone tubing in flow-through studies and use of silicone seals in contact with water should be minimised by the use of e.g. monoblock glass aquaria. The dimensions of the vessels should be large enough to allow proper growth in the control, maintenance of dissolved oxygen concentration (e.g. for small fish species, a 7 L tank volume will achieve this) and compliance with the loading rate criteria given in paragraph 19. It is desirable that test chambers be randomly positioned in the test area. A randomised block design with each treatment being present in each block is preferable to a completely randomised design. The test chambers should be shielded from unwanted disturbance. The test system should preferably be conditioned with concentrations of the test chemical for a sufficient duration to demonstrate stable exposure concentrations prior to the introduction of test organisms.

Selection of species

10. Recommended fish species are given in Table 1. This does not preclude the use of other species, but the test procedure may have to be adapted to provide suitable test conditions. The rationale for the selection of the species and the experimental method should be reported in this case.

Holding of the brood fish

11. Details on holding the brood stock under satisfactory conditions may be found in Appendix 3 and the references cited (3)(4)(5).

Handling of fertilised eggs, embryos and larvae

12. Initially, fertilised eggs, embryos and larvae may be exposed within the main vessel in smaller glass or stainless steel vessels, fitted with mesh sides or ends to permit a flow of test solution through the vessel. Non-turbulent flow-through in these small vessels may be induced by suspending them from an arm arranged to move the vessel up and down but always keeping the organisms submerged. Fertilised eggs of salmonid fishes can be supported on racks or meshes with apertures sufficiently large to allow larvae to drop through after hatching.
13. Where egg containers, grids or meshes have been used to hold eggs within the main test vessel, these restraints should be removed after the larvae hatch, according to the guidance in Appendix 3, except that meshes should be retained to prevent the escape of the larvae. If there is a need to transfer the larvae, they should not be exposed to the air and nets should not be used to release larvae from egg containers. The timing of this transfer varies with the species and should be documented in the report. However, a transfer may not always be necessary.

Water

14. Any water in which the test species shows suitable long-term survival and growth may be used as test water (see Appendix 4). 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-}), ammonia, total residual chlorine pesticides, total organic carbon and suspended solids should be made, for example, on a bi-annual basis where a dilution water is known to be relatively constant in quality. If the water is known to be of variable quality the measurements have to be conducted more often; the frequency is dependent of how variable the quality is. Some chemical characteristics of an acceptable dilution water are listed in Appendix 4.

Test solutions

15. For flow-through tests, a system which continually dispenses and dilutes a stock solution of the test chemical (e.g. metering pump, proportional diluter, saturator system) is required to deliver a series of concentrations to the test chambers. The flow rates of stock solutions and dilution water should be checked at intervals during the test and should not vary by more than 10 % throughout the test. A flow rate equivalent to at least five test chamber volumes per 24 hours has been found suitable (3). However, if the loading rate specified in paragraph 19 is respected, a lower flow rate of e.g. 2-3 test chamber volumes is possible to prevent quick removal of food.
16. 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 using mechanical means (e.g. stirring and/or ultrasonication). Saturation columns (solubility columns) or passive dosing methods (6) can be used for achieving a suitable concentrated stock solution. The use of a solvent carrier is not recommended. However, in case a solvent is necessary, a solvent control should be run in parallel, at the same solvent concentration as the chemical treatments; i.e. the solvent level should preferably be equal across all concentrations as well as the solvent control. For some diluter systems this might be technically difficult; here the solvent concentration in the solvent control should be equal to the highest solvent concentration in the treatment group. For difficult to test substances, the OECD Guidance Document No. 23 on aquatic toxicity testing of difficult substances and mixtures should be consulted (2). If a solvent is used, the choice of solvent will be determined by the chemical properties of the substance. The OECD Guidance Document No. 23 recommends a maximum concentration of 100 µl/l. To avoid potential effect of the solvent on endpoints measured (7), it is recommended to keep solvent concentration as low as possible.
17. For a semi-static test, two different renewal procedures may be followed. Either new test solutions are prepared in clean vessels and surviving eggs and larvae gently transferred into the new vessels, or the test organisms are retained in the test vessels whilst a proportion (at least two thirds) of the test solution / control volume is changed.

PROCEDURE

Conditions of Exposure*Duration*

18. The test should start as soon as possible after the eggs have been fertilised and preferably being immersed in the test solutions before cleavage of the blastodisc commences, or as close as possible after this stage. The test duration will depend upon the species used. Some recommended durations are given in Appendix 2.

Loading

19. The number of fertilised eggs at the start of the test should be sufficient to meet statistical requirements. They should be randomly distributed among treatments, and at least 80 eggs, divided equally between at least four replicate test chambers, should be used per concentration. The loading rate (biomass per volume of test solution) should be low enough in order that a dissolved oxygen concentration of at least 60 % of the air saturation value can be maintained without aeration during the egg and larval stage. For flow-through tests, a loading rate not exceeding 0,5 g/l wet weight per 24 hours and not exceeding 5 g/l of solution at any time has been recommended (3).

Light and temperature

20. The photoperiod and water temperature should be appropriate for the test species (see Appendix 2).

Feeding

21. Food and feeding are critical, and it is essential that the correct food for each life-stage is supplied from an appropriate time and at a level sufficient to support normal growth. Feeding should be approximately equal across replicates unless adjusted to account for mortality. Surplus food and faeces should be removed as necessary, to avoid accumulation of waste. Detailed feeding regimes are given in Appendix 3 but, as experience is gained, food and feeding regimes are continually being refined to improve survival and optimise growth. Live food provides a source of environmental enrichment and therefore should be used in place of or in addition to dry or frozen food whenever appropriate to the species and life stage.

Test concentrations

22. Normally five concentrations of the test chemical, with a minimum of four replicates per concentration, spaced by a constant factor not exceeding 3,2 are required. If available, information on the acute testing, preferable with the same species and/or a range finding test should be considered (1) when selecting the range of test concentrations. However, all sources of information should be considered when selecting the range of test concentrations, including sources like e.g. read across, fish embryo acute toxicity test data. A limit test, or an extended limit test, with fewer than five concentrations as the definitive test may be acceptable where empirical NOECs only are to be established. Justification should be provided if fewer than five concentrations are used. Concentrations of the test chemical higher than the 96 hour LC₅₀ or 10 mg/l, whichever is the lower, need not be tested.

Controls

23. A dilution-water control and, if needed, a solvent control containing the solvent carrier only should be run in addition to the test chemical concentration series (see paragraph 16).

Frequency of Analytical Determinations and Measurements

24. Prior to initiation of the exposure period, proper function of the chemical delivery system across all replicates should be ensured (for example, by measuring test concentrations). Analytical methods required should be established, including an appropriate limit of quantification (LOQ) and sufficient knowledge on the substance stability in the test system. During the test, the concentrations of the test chemical are determined at regular intervals to characterise exposure. A minimum of five determinations is necessary. In flow-through systems, analytical measurements of the test chemical in one replicate per concentration should be made at least once a week changing systematically amongst replicates. Additional analytical determinations will often improve the quality of the test outcome. Samples may need to be filtered to remove any particulate matter (e.g. using a 0,45 µm pore size) or centrifuged to ensure that the determinations are made on the chemical in true solution. In order to reduce adsorption of the test chemical, the filters should be saturated before the use. When the measured concentrations do not remain within 80-120 % of the nominal concentration, the effect concentrations should be determined and expressed relative to the arithmetic mean concentration for flow-through tests (see Appendix 6 of the test method C.20 for the calculation of the arithmetic mean (8)), and expressed relative to the geometric mean of the measured concentrations for semi-static tests (see Chapter 5 in the OECD Guidance Document on aquatic toxicity testing of difficult substances and mixtures (2)).
25. During the test, dissolved oxygen, pH, and temperature should be measured in all test vessels, at least weekly, and salinity and hardness, if warranted, at the beginning and end of the test. Temperature should preferably be monitored continuously in at least one test vessel.

Observations

26. **Stage of embryonic development:** the embryonic stage at the beginning of exposure to the test chemical should be verified as precisely as possible. This can be done using a representative sample of eggs suitably preserved and cleaned.
27. **Hatching and survival:** observations on hatching and survival should be made at least once daily and numbers recorded. If fungus on eggs is observed early in embryonic development (e.g. at day one or two of test), those eggs should be counted and removed. Dead embryos, larvae and juvenile fish should be removed as soon as observed since they can decompose rapidly and may be broken up by the actions of the other fish. Extreme care should be taken when removing dead individuals not to physically damage adjacent eggs/larvae. Signs of death vary according to species and life stage. For example:
- for fertilised eggs: 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;
 - for embryos, larvae and juvenile fish: immobility and/or absence of respiratory movement and/or absence of heartbeat and/or lack of reaction to mechanical stimulus.
28. **Abnormal appearance:** the number of larvae or juvenile fish showing abnormality of body form should be recorded at adequate intervals depending on the duration of the test and the nature of the abnormality described. It should be noted that abnormal larvae and juvenile fish occur naturally and can be of the order of several percent in the control(s) in some species. Where deformities and associated abnormal behaviour are considered so severe that there is considerable suffering to the organism, and it has reached a point beyond which it will not recover, it may be removed from the test. Such animals should be euthanised and treated as mortalities for subsequent data analysis. Normal embryonic development has been documented for most species recommended in this test method (9) (10) (11) (12).
29. **Abnormal behaviour:** abnormalities, e.g. hyperventilation, uncoordinated swimming, atypical quiescence and atypical feeding behaviour should be recorded at adequate intervals depending on the duration of the test (e.g. once daily for warm water species). These effects, although difficult to quantify, can, when observed, aid in the interpretation of mortality data.

30. **Weight:** at the end of the test, all surviving fish are weighed at least on a replicate basis (reporting the number of animals in the replicate and the mean weight per animal): wet weight — (blotted dry) is preferred, however, dry weight data may also be reported (13).
31. **Length:** at the end of the test, individual lengths are measured. Total length is recommended, if however, caudal fin rot or fin erosion occurs, standard length can be used. The same method should be used for all fish in a given test. Individual length can be measured either by e.g. callipers, digital camera, or calibrated ocular micrometer. Typical minimum lengths are defined in Appendix 2.

DATA AND REPORTING

Treatment of results

32. It is recommended that the design of the experiment and selection of statistical test permit adequate power (80 % or higher) to detect changes of biological importance in endpoints where a NOEC is to be reported. Reporting of relevant effect concentrations and parameters may depend upon the regulatory framework. If an EC_x is to be reported, the design of the experiment and selection of regression model should permit estimation of EC_x so that (i) the 95 % confidence interval reported for EC_x does not contain zero and is not overly wide, (ii) the 95 % confidence interval for the predicted mean at EC_x does not contain the control mean (iii) there is no significant lack-of-fit of regression model to the data. Either approach requires the identification of the percent change in each endpoint that is important to detect or estimate. The experimental design should be tailored to allow that. When the above conditions for determining the EC_x are not satisfied, the NOEC approach should be used. 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. Statistical flow diagrams and guidance for each approach are available in Appendixes 5 and 6 to guide in 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.
33. It will be necessary for variations to be analysed within each set of replicates using analysis of variance or contingency table procedures and appropriate statistical analysis methods be 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 Jonckheere-Terpstra or Williams' test is recommended for continuous responses and a step-down Cochran-Armitage test for quantal responses that are consistent with a monotone concentration-response and with no evidence of extra-binomial variance (14). When there is evidence of extra-binomial variance, the Rao-Scott modification of the Cochran-Armitage test is recommended (15) (16) or Williams or Dunnett's (after an arcsin-square-root transform) or Jonckheere-Terpstra test applied to replicate proportions. Where the data are not consistent with a monotone concentration-response, Dunnett's or Dunn's or the Mann-Whitney method may be found useful for continuous responses and Fisher's Exact test for quantal responses (14) (17) (18). Care should be taken where applying any statistical method or model to ensure that the requirements of the method or model are satisfied (e.g. chamber to chamber variability is estimated and accounted for in the experimental design and test or model used). Data are to be evaluated for normality and Appendix 5 indicates what should be done on the residuals from an ANOVA. Appendix 6 discusses additional considerations for the regression approach. Transformations to meet the requirements of a statistical test should be considered. However, transformations to enable the fitting of a regression model require great care, as, for example, a 25 % change in the untransformed response does not correspond to a 25 % change in a transformed response. In all analyses, the test chamber, not the individual fish, is the unit of analysis and the experimental unit and both hypothesis tests and regression should reflect that (3) (14) (19) (20).

Test report

34. The test report should include the following information:

Test chemical:

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, e.g., by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents

Test species:

- scientific name, strain, source and method of collection of the fertilised eggs and subsequent handling.

Test conditions:

- test procedure used (e.g. semi-static or flow-through, loading);
- photoperiod(s);
- test design (e.g. number of test chambers and replicates, number of eggs per replicate, material and size of the test chamber (height, width, volume), water volume per test chamber);
- 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;
- water quality within test vessels, pH, hardness, temperature and dissolved oxygen concentration;
- detailed information on feeding (e.g. type of food(s), source, amount given and frequency).

Results reported individually (or on a replicate basis) and as mean and coefficient of variation, as appropriate, for the following endpoints:

- evidence that controls met the overall survival acceptability standard of the test species (Appendix 2);
- data on mortality at each stage (embryo, larval and juvenile) and cumulative mortality;
- days to hatch, numbers of larvae hatched each day, and end of hatching;
- number of healthy fish at end of test;
- data for length (specify either standard or total) and weight of surviving animals;
- incidence, description and number of morphological abnormalities, if any;
- incidence, description and number of behavioural effects, if any;

- approach for the statistical analysis (regression analysis or analysis of the variance) and treatment of data (statistical test or model used);
- no observed effect concentration for each response assessed (NOEC);
- lowest observed effect concentration (at $p = 0,05$) for each response assessed (LOEC);
- 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 the test method.

Discussion of the results, including any influence of deviations from the test method on the outcome of the test.

Table 1

Fish species recommended for testing

FRESHWATER	ESTUARINE and MARINE
<i>Oncorhynchus mykiss</i> Rainbow trout	<i>Cyprinodon variegatus</i> Sheepshead minnow
<i>Pimephales promelas</i> Fathead minnow	<i>Menidia</i> sp. Silerside
<i>Danio rerio</i> Zebrafish	
<i>Oryzias latipes</i> Japanese ricefish or Medaka	

LITERATURE:

- (1) OECD (2012), Fish Toxicity Testing Framework, Environmental Health and Safety Publications Series on Testing and Assessment No.171, OECD, Paris.
- (2) OECD (2000), Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures, Environmental Health and Safety Publications, Series on Testing and Assessment. No. 23, OECD Paris.
- (3) ASTM (1988), Standard Guide for Conducting Early Life-Stage Toxicity Tests with Fishes. American Society for Testing and Materials, E 1241-88. 26 pp.
- (4) Brauhn, J.L. and R.A. Schoettger (1975), Acquisition and Culture of Research Fish: Rainbow trout, Fathead minnows, Channel catfish and Bluegills, Ecological Research Series, EPA-660/3-75-011, Duluth, Minnesota.
- (5) Brungs, W.A. and B.R. Jones (1977), Temperature Criteria for Freshwater Fish: Protocol and Procedures, Ecological Research Series EPA-600/3-77-061, Duluth, Minnesota.
- (6) Adolfsson-Erici, *et al.* (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.
- (7) Hutchinson, T.H. *et al.* (2006), Acute and chronic effects of carrier solvents in aquatic organisms: A critical review, Aquatic Toxicology, 76, 69-92.
- (8) Chapter C.20 of this Annex, *Daphnia magna* Reproduction Test.

-
- (9) Hansen, D.J. and P.R. Parrish (1977), Suitability of sheepshead minnows (*Cyprindon variegatus*) for life-cycle toxicity tests, In *Aquatic Toxicology and Hazard Evaluation* (edited by F.L. Mayer and J.L. Hamelink), ASTM STP 634.
- (10) Kimmel, H. B. *et al.* (1995), Stages of embryonic development of the zebrafish. *Developmental Dynamics*, 203:253–310.
- (11) Gonzalez-Doncel, M. *et al.* (2005), A quick reference guide to the normal development of *Oryzias latipes* (Teleostei, Adrinichthyidae) *Journal of Applied Ichthyology*, 20:1–14.
- (12) Devlin, E.W. *et al.* (1996), Prehatching Development of the Fathead Minnow, *Pimephales promelas* Rafinesque. EPA/600/R-96/079. USEPA, Office of Research and Development, Washington, D.C..
- (13) Oris, J.T., S.C. Belanger, and A.J. Bailer, (2012), Baseline characteristics and statistical implications for the OECD 210 Fish Early Life Stage Chronic Toxicity Test, *Environmental Toxicology and Chemistry* 31; 2, 370 - 376.
- (14) OECD (2006). *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application*, Environmental Health and Safety Publications Series on Testing and Assessment No.54, OECD, Paris.
- (15) Rao, J.N.K. and A.J. Scott (1992), A simple method for the analysis of clustered binary data, *Biometrics* 48, 577-585.
- (16) Rao, J.N.K. and A.J. Scott (1999), A simple method for analyzing overdispersion in clustered Poisson data, *Statistics in Medicine* 18, 1373-1385.
- (17) Dunnett C.W. (1955), A multiple comparisons procedure for comparing several treatments with a control, *Journal of American Statistical Association*, 50, 1096-1121.
- (18) Dunnett C.W. (1964), New tables for multiple comparisons with a control. *Biometrics*, 20, 482-491.
- (19) Rand, G.M. and S.R. Petrocelli (1985), *Fundamentals of Aquatic Toxicology*. Hemisphere Publication Corporation, New York.
- (20) McClave, J.T., J.H. Sullivan and J.G. Pearson (1980). *Statistical Analysis of Fish Chronic Toxicity Test Data*, Proceedings of 4th Aquatic Toxicology Symposium, ASTM, Philadelphia.
-

Appendix 1

DEFINITIONS:

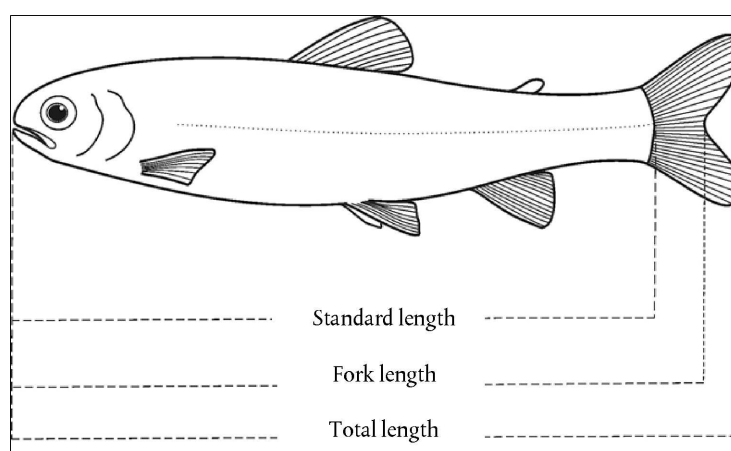
Fork length (FL): refers to the length from the tip of the snout to the end of the middle caudal fin rays and is used in fishes in which it is difficult to tell where the vertebral column ends (www.fishbase.org)

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



Chemical: a substance or a mixture

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.

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. Appendixes 5 and 6 provide guidance.

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.

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

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

IUPAC: International Union of Pure and Applied Chemistry.

SMILES: Simplified Molecular Input Line Entry Specification.

TEST CONDITIONS, DURATION AND SURVIVAL CRITERIA FOR RECOMMENDED SPECIES

SPECIES	TEST CONDITIONS			RECOMMENDED DURATION OF TEST	Typical minimum mean total length of control fish at the end of the study (mm) ⁽¹⁾	SURVIVAL OF CONTROLS (minimum)	
	Temperature (°C)	Salinity (‰)	Photoperiod (hrs)			Hatching success	Post-hatch success

Freshwater:

<i>Oncorhynchus mykiss</i> Rainbow trout	10 ± 1,5 ⁽²⁾		12 - 16 ⁽³⁾	2 weeks after controls are free-feeding (or 60 days post-hatch)	40	75 %	75 %
<i>Pimephales promelas</i> Fathead minnow	25 ± 1,5		16	32 days from start of test (or 28 days post-hatch)	18	70 %	75 %
<i>Danio rerio</i> Zebrafish	26 ± 1,5		12 - 16 ⁽⁴⁾	30 days post-hatch	11	70 %	75 %
<i>Oryzias latipes</i> Japanese Ricefish or Medaka	25 ± 2		12 - 16 ⁽⁴⁾	30 days post-hatch	17	80 %	80 %

Estuarine and Marine:

<i>Cyprinodon variegatus</i> Sheepshead minnow	25 ± 1,5	15-35 ⁽⁵⁾	12 - 16 ⁽⁴⁾	32 days from start of test (or 28 days post-hatch)	17	75 %	80 %
<i>Menidia sp.</i> Silverside	22 - 25	15-35 ⁽⁵⁾	13	28 days	20	80 %	60 %

Key:

- ⁽¹⁾ Typical minimum mean total length is not a validity criterion but deviations below the figure indicated should be carefully examined in relation to the sensitivity of the test. The minimum mean total length is derived from a selection of data available at the current time.
- ⁽²⁾ The particular strain of rainbow trout tested may necessitate the use of other temperatures. Brood stock must be held at the same temperature as that to be used for the eggs. After receipt of eggs from a commercial breeder, a short adaptation (e.g. 1-2 h) to test temperature after arrival is necessary.
- ⁽³⁾ Darkness for larvae until one week after hatching except when they are being inspected, then subdued lighting throughout test (12-16 hour photoperiod) ⁽⁴⁾.
- ⁽⁴⁾ For any given test conditions, light regime should be constant.
- ⁽⁵⁾ For any given test this shall be performed to ± 2 ‰.

FEEDING AND HANDLING GUIDANCE FOR BROOD AND TEST ANIMALS OF RECOMMENDED SPECIES

SPECIES	FOOD (*)				POST-HATCH TRANSFER TIME	TIME TO FIRST FEEDING
	Brood fish	Newly-hatched larvae	Juveniles			
			Type	Frequency		
Freshwater:						
<i>Oncorhynchus mykiss</i> Rainbow trout	trout food	None ^(a)	trout starter BSN	2-4 feeds per day	14-16 days post-hatch or at swim-up (not essential)	19 days post hatch or at swim-up
<i>Pimephales promelas</i> Fathead minnow	BSN, flake food, FBS	BSN	BSN48, flake food	2-3 times a day	once hatching is 90 %	2 day post hatch
<i>Danio rerio</i> Zebrafish	BSN, flake food	Commercial larvae food, protozoa ^(b) , protein ^(c)	BSN48, flake food,	BSN once daily; flake food twice daily	once hatching is 90 %	2 days post hatch
<i>Oryzias latipes</i> Japanese Ricefish or Medaka	flake food	BSN, flake food (or protozoa or rotifers)	BSN48, flake food (or rotifers)	BSN once daily; flake food twice daily or flake food and rotifers once daily	not applicable	6-7 days post spawn
Estuarine and Marine:						
<i>Cyprinodon variegatus</i> Sheepshead minnow	BSN, flake food, FBS	BSN	BSN48	2-3 feeds per day	not applicable	1 day post hatch/swim-up
<i>Menidia sp.</i> Silerside	BSN48, flake food	BSN	BSN48	2-3 feeds per day	not applicable	1 day post hatch/swim-up

Key:

(*) Food should be given to satiation. Surplus food and faeces should be removed, as necessary to avoid accumulation of waste

FBS frozen brine shrimps; adults *Artemia* sp

BSN brine shrimp nauplii; newly hatched

BSN48 brine shrimp nauplii; 48 hours old

^(a) yolk-sac larvae require no food^(b) filtered from mixed culture^(c) granules from fermentation process.

Appendix 4

SOME CHEMICAL CHARACTERISTICS OF AN ACCEPTABLE DILUTION WATER

Component	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 5

STATISTICAL GUIDANCE FOR NOEC DETERMINATION

General

The replicate tank is the unit of analysis. Thus, for continuous measurements, such as size, the replicate mean or median should be calculated and these replicate values are the data for analysis. The power of the tests used should be demonstrated, preferably based on an adequate historical database for each lab. The size effect that can be detected with 75-80 % power should be provided for each endpoint with the statistical test to be used.

The databases available at the time of development of this test method establish the power possible under the recommended statistical procedures. An individual lab should demonstrate its ability to meet this power requirement either by conducting its own power analysis or by demonstrating that the Coefficient of Variation (CV) for each response does not exceed the 90th percentile of CVs used in developing the TG. Table 1 provides these CVs. If only replicate means or medians are available, then the within-replicate CV can be ignored.

Table 1

90th Percentile CVs for selected Freshwater Species

Species	Response	CV_Between Replicates	CV_Within Replicates
Rainbow Trout	Length	17,4	9,8
	Weight	10,1	28
Fathead Minnow	Length	16,9	13,5
	Weight	11,7	38,7
Zebrafish	Length	43,7	11,7
	Weight	11,9	32,8

For almost all statistical tests used to evaluate laboratory toxicology studies, the comparisons of interest are of treatment groups to control. For that reason, it is not appropriate to require a significant ANOVA F-test before using Dunnett's or Williams' test or a significant Kruskal-Wallis test before using the Jonckheere-Terpstra, Mann-Whitney, or Dunn test (Hochberg and Tamhane 1987, Hsu 1996, Dunnett 1955, 1964, Williams 1971, 1972, 1975, 1977, Robertson *et al.* 1988, Jonckheere 1954, Dunn 1964).

Dunnett's test has a built-in multiplicity adjustment and its false positive and false negative rates are adversely affected by using the F-test as a gatekeeper. Similarly, the step-down Williams and Jonckheere-Terpstra tests using a 0,05 significance level at every step preserve an overall 5 % false positive rate and that rate and the power of the tests are adversely affected by using the F- or Kruskal-Wallis test as a gatekeeper. Mann-Whitney and Dunn's test have to be adjusted for multiplicity and the Bonferroni-Holm adjustment is advised.

A thorough discussion of most of the recommendations on hypothesis testing and verification of assumptions underlying these tests is given in OECD (2006), which also contains an extensive bibliography.

Treatment of Controls when a Solvent is Used

If a solvent is used, then both a dilution water control and a solvent control should be included. The two controls should be compared for each response and combined for statistical analysis if no significant difference is found between the controls. Otherwise, the solvent control should be used for NOEC determination or EC_x estimation and the water control is not used. See restriction in the validity criteria (Paragraph 7)

For length, weight, proportion of egg hatch or larval mortality or abnormal larvae, and first or last day of hatch or swim-up, a T-test or Mann-Whitney test should be used to compare the dilution water- control and the solvent control at the 0,05 significance level, ignoring all treatment groups. The results of these tests should be reported.

Size Measurements (length and weight)

Individual fish length and weight values can be normally or log-normally distributed. In either case, the replicate mean values tend to be normally distributed by virtue of the Central Limit Theorem and confirmed from data from well over 100 ELS studies of three freshwater species. Alternatively, where the data or historical databases suggest a log-normal distribution for individual fish size values, the replicate mean logarithm of the individual fish values can be calculated and the data for analysis can then be the anti-logs of these replicate mean logarithms.

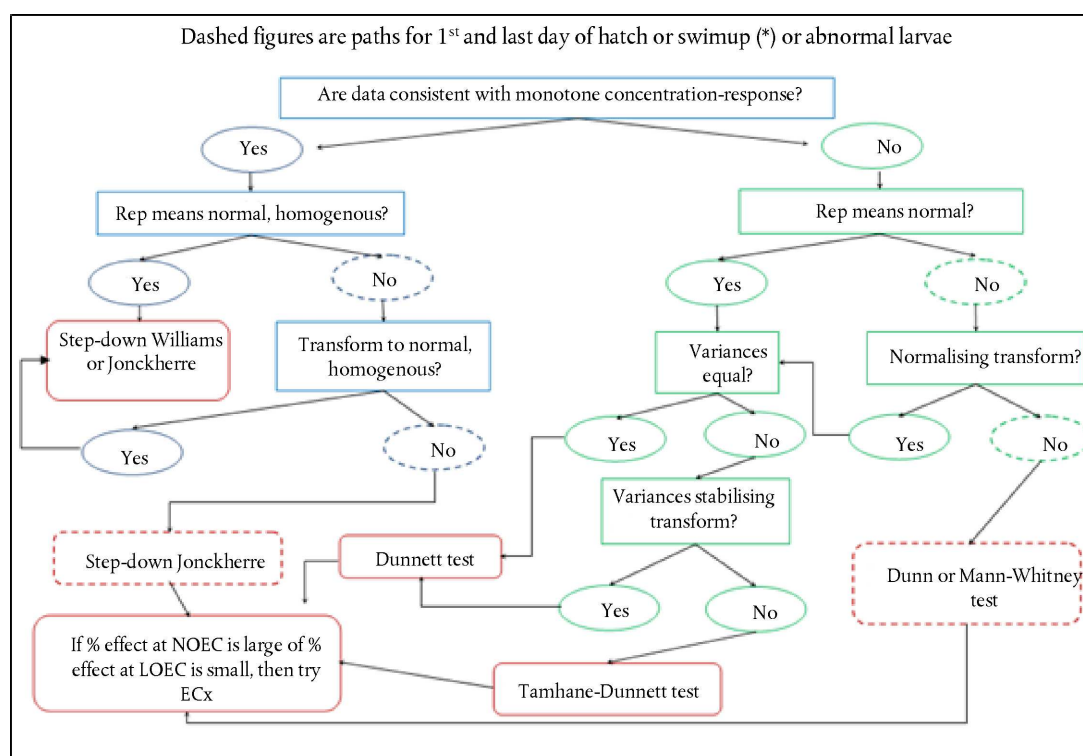
Data should be evaluated for consistency with a normal distribution and variance homogeneity. For this purpose, the residuals from an ANOVA model with concentration as the single explanatory class variable should be used. Visual determination from scatterplots and histograms or stem-and-leaf plots can be used. Alternatively, a formal test such as the Shapiro-Wilk or Anderson-Darling can be used. Consistency with variance homogeneity can be assessed from a visual examination of the same scatter plot or formally from Levene's test. Only parametric tests (e. g. Williams, Dunnett) need be evaluated for normality or variance homogeneity.

Attention should be paid to possible outliers and their effect on analysis. Tukey's outlier test and visual inspection of the same plots of residuals described above can be used. It should be recalled that observations are entire replicates, so omitting an outlier from analysis should be done only after careful consideration.

The statistical tests that make use of the characteristics of the experimental design and biological expectation are step-down trend tests, such as Williams and Jonckheere-Terpstra. These tests assume a monotone concentration-response and the data should be assessed for consistency with that assumption. This can be done visually from a scatter plot of the replicate means against test concentration. It will be helpful to overlay that scatter plot with a piecewise linear plot connecting the concentration means weighted by replicate sample size. Great deviation of this piecewise linear plot from monotonicity would indicate a possible need to use non-trend tests. Alternatively, formal tests can be used. A simple formal test is to compute linear and quadratic contrasts of the concentration means. If the quadratic contrast is significant and the linear contrast is not significant that is an indication of a possible problem with monotonicity which should be further evaluated from plots. Where normality or variance homogeneity may be an issue, these contrasts can be constructed from rank-order transformed data. Alternative procedures, such as Bartholomew's test for monotonicity can be used, but add complexity.

Figure 2

NOEC Flow-Chart Size Measurements (length and weight)



(*) These responses never satisfy assumptions for parametric analysis or models.

Unless the data are not consistent with the requirements for these tests, the NOEC is determined by a step-down application of Williams' or the Jonckheere-Terpstra test. OECD (2006) provides details on these procedures. For data not consistent with the requirements for a step-down trend test, Dunnett's test or the Tamhane-Dunnett (T3) test can be used, both of which have built-in adjustments for multiplicity. These tests assume normality and, in the case of Dunnett, variance homogeneity. Where those conditions are not satisfied, Dunn's non-parametric test can be used. OECD (2006) contains details for all of these tests. Figure 2 is giving an overview, how to find the test of choice.

Egg Hatch and Larval Survival

The data are proportions of eggs that hatch or larvae that survive in individual replicates. These proportions should be assessed for extra-binomial variance, which is common but not universal for such measurements. The flowchart in figure 3 is guidance for the test of choice; see text for detailed descriptions.

Two tests are commonly used. These are Tarone's C(α) test (Tarone, 1979) and chi-squared tests, each applied separately to every test concentration. If extra-binomial variance is found in even one test concentration, then methods that accommodate that should be used.

Formula 1

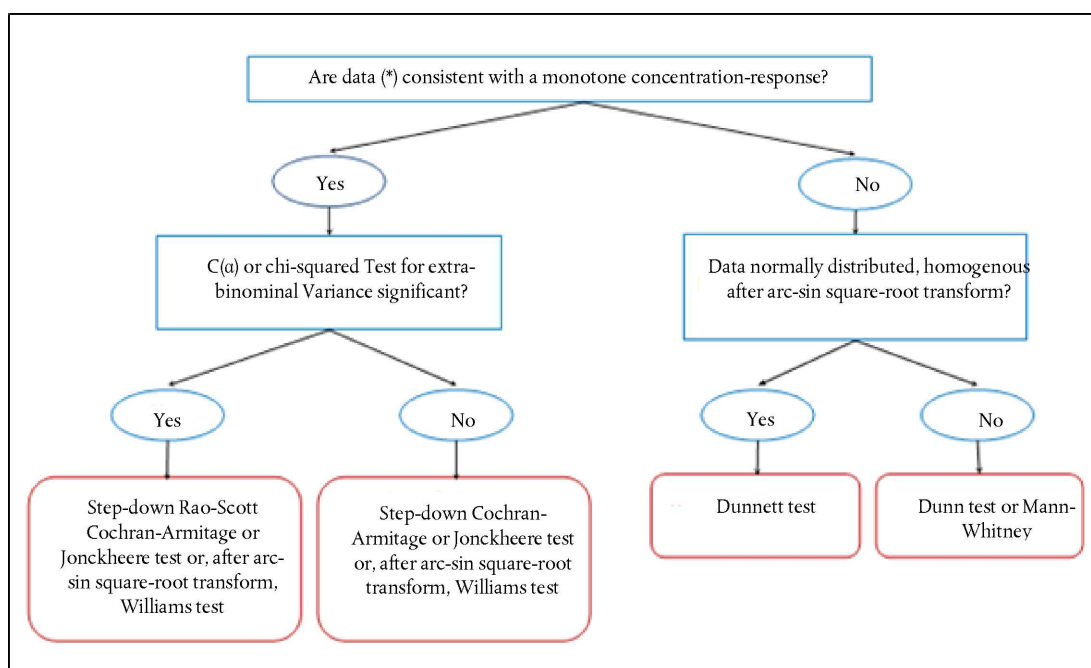
Tarone's C (α) test (Tarone 1979)

$$Z = \frac{\sum_{j=1}^m \frac{(x_j - n_j \hat{p})^2}{\hat{p}(1 - \hat{p})} - \sum_{j=1}^m n_j}{\left\{ 2 \sum_{j=1}^m n_j (n_j - 1) \right\}^{1/2}}$$

Where \hat{p} is the mean proportion for a given concentration, \mathbf{m} is the number of replicate tanks, \mathbf{n}_j is the number of subjects in replicate j , and \mathbf{x}_j is the number of subjects in that replicate responding, e.g. not hatched or dead. This test is applied to each concentration separately. This test can be seen as an adjusted chi-squared test, but limited power simulations done by Tarone have shown it to be more powerful than a chi-squared test.

Figure 3

NOEC Flow Chart for Egg Hatch and Larval Mortality



(*) Data are replicate proportion.

Where there is no significant evidence of extra-binomial variance, the step-down Cochran-Armitage test can be used. This test ignores replicates, so where there is such evidence, the Rao-Scott adjustment to the Cochran-Armitage test (RSCA) takes replicates, replicate sizes, and extra-binomial variance into account and is recommended. Alternative tests include the step-down Williams and Jonckheere-Terpstra tests and Dunn's test as described for size measurements. These tests apply whether or not there is extra-binomial variance, but have somewhat lower power (Agresti 2002, Morgan 1992, Rao and Scott 1992, 1999, Fung *et al.* 1994, 1996).

First or Last Day of Hatch or Swim-up

The response is an integer, giving the test day on which the indicated observation is observed for a given replicate tank. The range of values is generally very limited and there are often high proportions of tied values, e.g. the same first day of hatch is observed in all control replicates and, perhaps in one or two low test concentrations. Parametric tests such as Williams and Dunn's test are not appropriate for such data. Unless there is evidence on serious non-monotonicity, the step-down Jonckheere-Terpstra test is very powerful for detecting effects of the test chemical. Otherwise, Dunn's test can be used.

Larval Abnormalities

The response is the count of larvae found to be abnormal in some way. This response is frequently of low incidence and has some of the same problems as first day of hatch, as well as sometimes exhibiting erratic in concentration-response. If the data at least roughly follow a monotone concentration shape, the step-down Jonckheere-Terpstra test is powerful for detecting effects. Otherwise, Dunn's test can be used.

REFERENCES:

Agresti, A. (2002); Categorical Data Analysis, second edition, Wiley, Hoboken.

Dunnett C. W. (1955); A multiple comparison procedure for comparing several treatments with a control, J. American Statistical Association 50, 1096-1121.

Dunn O. J. (1964); Multiple Comparisons Using Rank Sums, Technometrics 6, 241-252.

Dunnett C. W. (1964); New tables for multiple comparisons with a control, Biometrics 20, 482-491.

Fung, K.Y., D. Krewski, J.N.K. Rao, A.J. Scott (1994); Tests for Trend in Developmental Toxicity Experiments with Correlated Binary Data, Risk Analysis 14, 639-648.

Fung, K.Y., D. Krewski, R.T. Smythe (1996); A comparison of tests for trend with historical controls in carcinogen bioassay, Canadian Journal of Statistics 24, 431-454.

Hochberg, Y. and A. C. Tamhane (1987); Multiple Comparison Procedures, Wiley, New York.

Hsu, J.C. (1996); Multiple Comparisons: Theory and Methods; Chapman and Hall/CRC Press, Boca Raton.

Jonckheere A. R. (1954); A distribution-free k-sample test against ordered alternatives, Biometrika 41, 133.

Morgan, B.J.T. (1992); Analysis of Quantal Response Data, Chapman and Hall, London.

OECD (2006). Current approaches in the statistical analysis of ecotoxicity data: A guidance to application. Series on Testing and Assessment, No. 54. Organisation for Economic Co-operation and Development, OECD, Paris..

Rao J.N.K. and Scott A.J. (1992) — A simple method for the analysis of clustered binary data, Biometrics 48, 577-585.

Rao J.N.K. and Scott A.J. (1999) — A simple method for analyzing overdispersion in clustered Poisson data, Statistics in Medicine 18, 1373-1385.

Robertson, T., Wright F.T. and Dykstra R.L. (1988); Order restricted statistical inference, Wiley.

Tarone, R.E. (1979); Testing the goodness of fit of the Binomial distribution, Biometrika 66, 585-590.

Williams D.A. (1971); A test for differences between treatment means when several dose levels are compared with a zero dose control, Biometrics 27, 103-117.

Williams D.A. (1972); The comparison of several dose levels with a zero dose control, Biometrics 28, 519-531.

Williams D. A. (1975); The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratology, Biometrics 31, 949-952.

Williams D.A. (1977); Some inference procedures for monotonically ordered normal means, Biometrika 64, 9-14.

Appendix 6

STATISTICAL GUIDANCE FOR REGRESSION ESTIMATES

General

The observations used to fit a model are replicate means (length and weight) or replicate proportions (egg hatch and larval mortality) (OECD 2006).

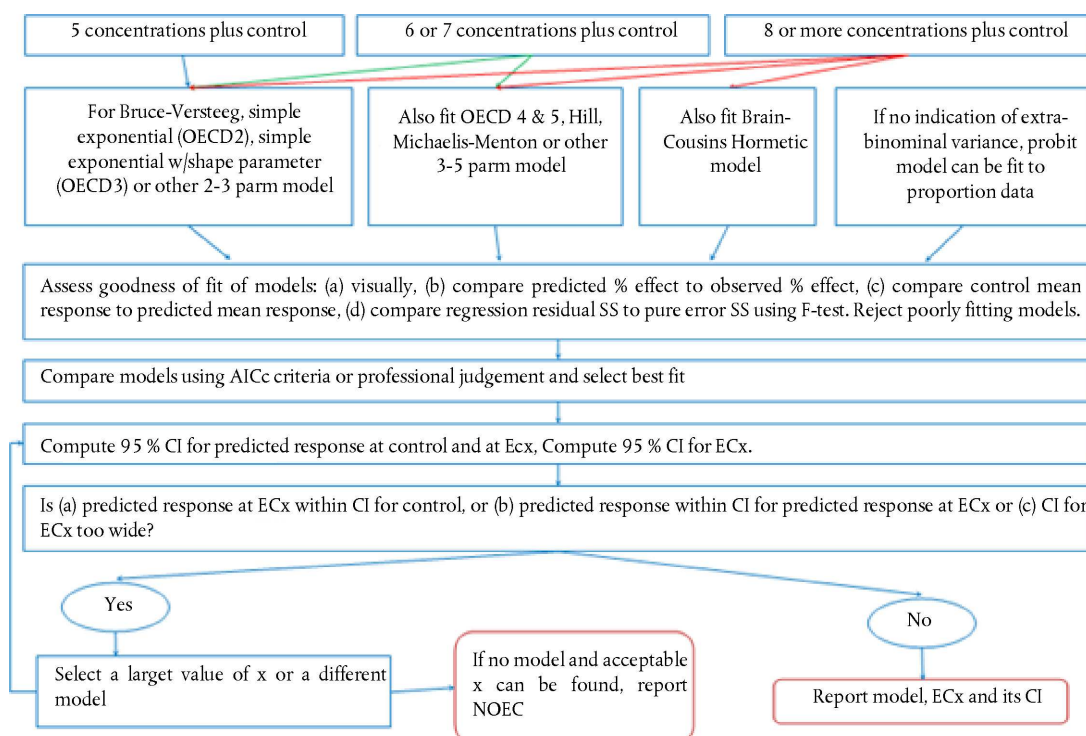
Weighted regression using replicate sample size as weight is generally advised. Other weighting schemes are possible, such as weighting by predicted mean response or a combination of this and replicate sample size. Weighting by reciprocal of within-concentration sample variance is not recommended (Bunke *et al.* 1999, Seber and Wild, 2003, Motulsky and Christopoulos 2004, Huet *et al.* 2003).

Any transformation of responses prior to analysis should preserve the independence of the observations and ECx and its confidence bounds should be expressed in the original units of measurement, rather than in transformed units. For example, a 20 % change in the logarithm of length is not equivalent to a 20 % change in length (Lyles *et al.* 2008, Draper and Smith 1999).

The flowchart in figure 4 gives an overview for ECx estimations. The details are described in the text below.

Figure 4

Flow chart for ECx Estimation of Replicate Mean Length, Weight, or Proportion of Egg Hatch or Larval Mortality, see text for more details



Considerations for Egg Hatch and Larval Mortality

For egg hatch and larval mortality, it is generally best to fit a decreasing model unless one is fitting a probit model as described below. That is, one should model the proportion of eggs that do not hatch or larvae that die. The reason for this is that EC_x refers to a concentration at which there is a change equal to x % of the control mean response. If there are 5 % control eggs that fail to hatch and one models failure to hatch, then EC₂₀ refers to a concentration at which there is a change equal to 20 % of the 5 % control failure to hatch, and that is a change of $0,2 \times 0,05 = 0,01$ or 1 percentage point to 6 % failure to hatch. Such a small change cannot be estimated in any meaningful way from the data available and is not biologically important. Whereas if one models the proportion of eggs that hatch, the control proportion would be 95 % in this example and a 20 % reduction from the control mean would be a change of $0,95 \times 0,2 = 0,18$, so from 95 % hatch success to 77 % (= 95 – 18) hatching success and that effects concentration can be estimated and is presumably of greater interest. This is not an issue with size measurements, though adverse effects on size generally mean a decrease in size.

Models for Size (length or weight) and Egg Hatch Success or Larval Survival.

Except for the Brain-Cousens hormetic model, all of these models are described and recommended in OECD (2006). What are called OECD 2-5, are also discussed for ecotoxicity experiments in Slob (2002). There are, of course, many other models that might be useful. Bunke, *et al.* (1999) lists numerous models not included here and references to other models are plentiful. Those listed below are suggested as particularly appropriate in ecotoxicity experiments and widely used.

With 5 test concentrations plus control

- Bruce-Versteeg
- Simple Exponential (OECD 2)
- Exponential with shape parameter (OECD 3)
- Simple Exponential with Lower Bound (OECD 4)

With 6 or more test concentrations plus control

- Exponential with shape parameter and lower bound (OECD 5)
- Michaelis-Menten
- Hill

Where there is visual evidence of hormesis (unlikely with egg hatch success or larval survival, but sometimes observed in size observations)

- Brain-Cousens Hormetic; Brain and Cousens (1989)

Alternative models for egg hatch failure and larval mortality

- Increasing models for these responses can be fit by probit (or logistic) models if there is no evidence of extra-binomial variance and control incidence is estimated in the model fit. This is not the preferred method, as it treats the individual, not the replicate, as the unit of analysis (Morgan 1992, O'Hara Hines and Lawless 1993, Collett 2002, 2003).

Goodness of fit of a single model

- Visually compare observed and predicted percent decrease at each test concentration (Motulsky and Christopoulos 2004, Draper and Smith 1999).

- Compare regression error mean square against the pure error mean square using an F-test (Draper and Smith 1999).
- Check that every term in the model is significantly different from zero (i.e., determine whether all model terms are important), (Motulsky and Christopoulos 2004).
- Plots of residuals from regression vs. test concentration, possibly on a log(conc) scale. There should be no pattern to this plot; the points should be randomly scattered about a horizontal line at zero height.
- The data should be evaluated for normality and variance homogeneity in the same way as indicated in Appendix 5.
- In addition, normality of the residuals about the regression model should be assessed using the same methods indicated in Appendix 5 for the residuals from ANOVA

Compare models

- Use Akiake's AICc criteria. Smaller AICc values denote better fits and if $AICc(B) - AICc(A) \geq 10$, the model A is almost certainly better than model B (Motulsky and Christopoulos (2004).
- Compare the two models visually by how well they meet the single model criteria above.
- The parsimony principal is advised, whereby the simplest model that fits the data reasonably well is used (Ratkowsky 1993, Lyles et.al 2008).

Quality of EC_x estimate

The confidence interval (CI) for EC_x should not be too wide. Statistical judgment is needed in deciding how wide the confidence interval can be and EC_x still be useful. Simulations for regression models fit to egg hatching and size data show that about 75 % of confidence intervals for EC_x ($x = 10, 20$ or 30) span no more than two test concentrations. This provides a general guide for what is acceptable and a practical guide for what is achievable. Numerous authors assert the need to report confidence intervals for all model parameters and that wide confidence intervals for model parameters indicate unacceptable models (Ott and Longnecker 2008, Alvord and Rossio 1993, Motulsky and Christopoulos 2004, Lyles *et al.* 2008, Seber and Wild 2003, Bunke *et al.* 1999, Environment Canada 2005).

The CI for EC_x (or any other model parameter) should not contain zero (Motulsky and Christopoulos 2004). This is the regression equivalent the minimum significant difference that is often cited in hypothesis testing approaches (e. g. Wang *et al* 2000). It also corresponds to the confidence interval for the mean responses at the LOEC not contain the control mean. One should wonder whether the parameter estimates are scientifically plausible. E.g. if the confidence interval for y_0 is ± 20 %, no EC_{10} estimate is plausible. If the model predicts a 20 % effect at a concentration C and the maximum observed effect at C and lower concentrations is 10 %, then the EC_{20} is not plausible (Motulsky and Christopoulos 2004, Wang *et al.* 2000, Environment Canada 2005).

EC_x should not require extrapolation outside the range of positive concentrations (Draper and Smith 1999, OECD 2006). For example, a general guide might be for EC_x to be no more than about 25 % below the lowest tested concentration or above the highest tested concentration.

REFERENCES:

Alvord, W.G., Rossio, J.L. (1993); Determining confidence limits for drug potency in immunoassay, *Journal of Immunological Methods* 157, 155-163.

Brain P. and Cousens R. (1989); An equation to describe dose responses where there is stimulation of growth at low doses. *Weed res.* 29: 93-96.

Bunke, O., Droge, B. and Polzehl, J. (1999). Model selection, transformations and variance estimation in nonlinear regression. *Statistics* 33, 197-240.

Collett, D. (2002); *Modelling Binary Data*, second edition, Chapman and Hall, London.

Collett, D. (2003); *Modelling Survival Data in Medical Research*, second edition, Chapman and Hall, London.

Draper, N.R. and Smith, H. (1999); *Applied Regression Analysis*, third edition. New York: John Wiley & Sons.

Environment Canada (2005); *Guidance Document on Statistical Methods for Environmental Toxicity Tests*, Report EPS 1/RM/46

Huet, S., A. Bouvier, M.-A. Poursat, E. Jolivet (2003); *Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples*, Springer Series in Statistics, New York.

Lyles, R. H., C. Poindexter, A. Evans, M. Brown, and C.R. Cooper (2008); Nonlinear Model-Based Estimates of IC50 for Studies Involving Continuous Therapeutic Dose-Response Data, *Contemp Clin Trials*. 2008 November; 29(6): 878–886.

Morgan, B.J.T. (1992); *Analysis of Quantal Response Data*, Chapman and Hall, London.

Motulsky, H., A. Christopoulos (2004); *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, Oxford University Press, USA.

O'Hara Hines, R. J. and J. F. Lawless (1993); *Modelling Overdispersion in Toxicological Mortality Data Grouped over Time*, *Biometrics* Vol. 49, pp. 107-121

OECD (2006); *Current approaches in the statistical analysis of ecotoxicity data: A guidance to application*. Series Testing and Assessment, No. 54, Organisation for Economic Co-operation and Development, OECD, Paris.

Ott, R.L., M.T. Longnecker, *An Introduction to Statistical Methods and Data Analysis*, sixth edition, 2008, Brooks-Cole, Belmont, CA

Ratkowsky, D.A. (1993); Principles of nonlinear regression, *Journal of Industrial Microbiology* 12, 195-199.

Seber, G.A.F., C.J. Wild, *Nonlinear Regression*, Wiley, 2003

Slob W. (2002); Dose-response modelling of continuous endpoints. *Toxicol. Sci.*, 66, 298-312

Wang, Q., D.L. Denton, and R. Shukla (2000); Applications and Statistical Properties Of Minimum Significant Difference-Based Criterion Testing In a Toxicity Testing Program, *Environmental Toxicology and Chemistry*, Vol. 19, pp. 113–117, 2000.

C.48 Fish Short Term Reproduction Assay

INTRODUCTION

1. This test method is equivalent to OECD test guideline (TG) 229 (2012). The need to develop and validate a fish assay capable of detecting endocrine active chemicals originates from the concerns that environmental levels of chemicals may cause adverse effects in both humans and wildlife due to the interaction of these chemicals with the endocrine system. In 1998, the OECD initiated a high-priority activity to revise existing guidelines and to develop new guidelines for the screening and testing of potential endocrine disruptors. One element of the activity was to develop a test guideline for the screening of chemicals active on the endocrine system of fish species. The Fish Short Term Reproduction Assay underwent an extensive validation programme consisting of inter-laboratory studies with selected chemicals to demonstrate the relevance and reliability of the assay for the detection of chemicals that impact reproduction in fish by various mechanisms including endocrine modalities (1, 2, 3, 4, 5). All endpoints of the OECD test guideline have been validated on the fathead minnow, and a subset of endpoints has been validated in the Japanese medaka (i.e. vitellogenin and secondary sex characteristics) and the zebrafish (i.e. vitellogenin). The validation work has been peer-reviewed by a panel of experts nominated by the National Coordinators of the OECD Test Guideline Programme (6) in part, and by an independent panel of experts commissioned by the United States Environmental Protection Agency (29). The assay is not designed to identify specific mechanisms of hormonal disruption because the test animals possess an intact hypothalamic-pituitary-gonadal (HPG) axis, which may respond to chemicals that impact on the HPG axis at different levels.
2. This test method describes an *in vivo* screening assay where sexually mature male and spawning female fish are held together and exposed to a chemical during a limited part of their life-cycle (21 days). At termination of the 21-day exposure period, two biomarker endpoints are measured in males and females as indicators of endocrine activity of the test chemical; these endpoints are vitellogenin and secondary sexual characteristics. Vitellogenin is measured in fathead minnow, Japanese medaka and zebrafish, whereas secondary sex characteristics are measured in fathead minnow and Japanese medaka. Additionally, quantitative fecundity is monitored daily throughout the test. Gonads are also preserved and histopathology may be evaluated to assess the reproductive fitness of the test animals and to add to the weight of evidence of other endpoints.
3. This bioassay serves as an *in vivo* reproductive screening assay and its application should be seen in the context of the "OECD Conceptual Framework for the Testing and Assessment of Endocrine Disrupting Chemicals" (30). In this Conceptual Framework the Fish Short Term Reproduction Assay is proposed at Level 3 as an *in vivo* assay providing data about selected endocrine mechanism(s)/pathway(s).

INITIAL CONSIDERATIONS AND LIMITATIONS

4. Vitellogenin (VTG) is normally produced by the liver of female oviparous vertebrates in response to circulating endogenous oestrogen. It is a precursor of egg yolk proteins and, once produced in the liver, travels in the bloodstream to the ovary, where it is taken up and modified by developing eggs. Vitellogenin is almost undetectable in the plasma of immature female and male fish because they lack sufficient circulating oestrogen; however, the liver is capable of synthesising and secreting vitellogenin in response to exogenous oestrogen stimulation.
5. The measurement of vitellogenin serves for the detection of chemicals with various oestrogenic modes of action. The detection of oestrogenic chemicals is possible via the measurement of vitellogenin induction in male fish, and it has been abundantly documented in the scientific peer-reviewed literature (e.g. (7)). Vitellogenin induction has also been demonstrated following exposure to aromatisable androgens (8, 9). A reduction in the circulating level of oestrogen in females, for instance through the inhibition of the aromatase converting the endogenous androgen to the natural oestrogen 17 β -estradiol, causes a decrease in the VTG level which is used to detect chemicals having aromatase inhibiting properties (10, 11). The biological relevance of the vitellogenin response following oestrogenic/aromatase inhibition is established and has been broadly documented. However, it is possible that production of VTG in females can also be affected by general toxicity and non-endocrine toxic modes of action, e.g. hepatotoxicity.

6. Several measurement methods have been successfully developed and standardised for routine use. This is the case of species-specific Enzyme-Linked Immunosorbent Assay (ELISA) methods using immunochemistry for the quantification of VTG produced in small blood or liver samples collected from individual fish (12, 13, 14, 15, 16, 17, 18). Fathead minnow blood, zebrafish blood or head/tail homogenate, and medaka liver are sampled for VTG measurement. In medaka, there is a good correlation between VTG measured from blood and from liver (19). Appendix 6 provides the recommended procedures for sample collection for VTG analysis. Kits for the measurement of VTG are widely available; such kits should be based on a validated species-specific ELISA method.
7. Secondary sex characteristics in male fish of certain species are externally visible, quantifiable and responsive to circulating levels of endogenous androgens; this is the case for the fathead minnow and the medaka — but not for zebrafish which does not possess quantifiable secondary sex characteristics. Females maintain the capacity to develop male secondary sex characteristics, when they are exposed to androgenic chemicals in water. Several studies are available in the scientific literature to document this type of response in fathead minnow (20) and medaka (21). A decrease in secondary sex characteristics in males should be interpreted with caution because of low statistical power, and should be based on expert judgement and weight of evidence. There are limitations to the use of zebrafish in this assay, due to the absence of quantifiable secondary sex characteristics responsive to androgenic acting chemicals.
8. In the fathead minnow, the main indicator of exogenous androgenic exposure is the number of nuptial tubercles located on the snout of the female fish. In the medaka, the number of papillary processes constitutes the main marker of exogenous exposure to androgenic chemicals in female fish. Appendix 5A and Appendix 5B indicate the recommended procedures to follow for the evaluation of sex characteristics in fathead minnow and in medaka, respectively.
9. The 21-day fish assay includes the evaluation of quantitative egg production and preservation of gonads for optional histopathology examination. Some regulatory authorities may require this additional endpoint for a more complete evaluation of the reproductive fitness of the test animals, or in cases where vitellogenin and secondary sex characteristics did not respond to the chemical exposure. Although some endpoints may be highly diagnostic (e.g. VTG induction in males and tubercle formation in females), not all endpoints (e.g. fecundity and gonad histopathology) in the assay are intended to unequivocally identify specific cellular mechanisms of action. Rather, the suite of endpoints, collectively, allows inferences to be made with regard to possible endocrine disturbances and thus provide guidance for further testing. Although not endocrine specific, fecundity, due to its demonstrated sensitivity across known endocrine active chemicals (5), is an important endpoint to include because when it and other endpoints are unaffected one is more confident that a compound is not likely endocrine active. However, when fecundity is affected it will contribute heavily in weight of evidence inferences. Guidance on data interpretation and acceptance of test results is provided further in this test method.
10. Definitions used in this test method are given in Appendix 1.

PRINCIPLE OF THE TEST

11. In the assay, male and female fish in a reproductive status are exposed together in test vessels. Their adult and reproductive status enables a clear differentiation of each sex, and thus a sex-related analysis of each endpoint, and ensures their sensitivity towards exogenous chemicals. At test termination, sex is confirmed by macroscopic examination of the gonads following ventral opening of the abdomen with scissors. An overview of the relevant bioassay conditions are provided in Appendix 2. The assay is normally initiated with fish sampled from a population that is in spawning condition; senescent animals should not be used. Guidance on

the age of fish and on the reproductive status is provided in the section on Selection of fish. The assay is conducted using three chemical exposure concentrations as well as a water control, and a solvent control if necessary. Two vessels or replicates per treatment are used for zebrafish (each vessel containing 5 males and 5 females). Four vessels or replicates per treatment are used for fathead minnow (each vessel containing 2 males and 4 females). This is to accommodate the territorial behaviour of male fathead minnow while maintaining sufficient power of the assay. Four vessels or replicates per treatment are used for medaka (each vessel containing 3 males and 3 females). The exposure is conducted for 21-days and sampling of fish is performed at day 21 of exposure. Quantitative fecundity is monitored daily.

12. On sampling at day 21, all animals are killed humanely. Secondary sex characteristics are measured in fathead minnow and medaka (see Appendix 5A and Appendix 5B); blood samples are collected for determination of VTG in zebrafish and fathead minnow, alternatively head/tail can be collected for the determination of VTG in zebrafish (Appendix 6); liver is collected for VTG analysis in medaka (Appendix 6); gonads are fixed either in whole or dissected for potential histopathological evaluation (22).

TEST ACCEPTANCE CRITERIA

13. For the test results to be acceptable the following conditions apply:
 - the mortality in the water (or solvent) controls should not exceed 10 per cent at the end of the exposure period;
 - the dissolved oxygen concentration should be at least 60 per cent of the air saturation value (ASV) throughout the exposure period;
 - the water temperature should not differ by more than $\pm 1,5$ °C between test vessels at any one time during the exposure period and be maintained within a range of 2 °C within the temperature ranges specified for the test species (Appendix 2);
 - 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;
 - evidence that fish are actively spawning in all replicates prior to initiating chemical exposure and in control replicates during the test.

DESCRIPTION OF THE METHOD

Apparatus

14. 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 2);
 - (e) spawning substrate for fathead minnow and zebrafish, Appendix 4 gives the necessary details.
 - (f) suitably accurate balance (i.e. accurate to $\pm 0,5$ mg).

Water

15. 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. 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 $\pm 0,5$ pH units. In order to ensure that the dilution water will not unduly influence the test result (for example by complexation of test chemical); samples should be taken at intervals for analysis. Measurements of heavy metals (e.g. Cu, Pb, Zn, Hg, Cd, and Ni), major anions and cations (e.g. Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , and SO_4^{2-}), pesticides (e.g. total organophosphorus and total organochlorine pesticides), total organic carbon and suspended solids should be made, for example, every three months where dilution water is known to be relatively constant in quality. If water quality has been demonstrated to be constant over at least one year, determinations can be less frequent and intervals extended (e.g. every six months). Some chemical characteristics of acceptable dilution water are listed in Appendix 3.

Test solutions

16. 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 using mechanical means (e.g. stirring or ultrasonication). Saturation columns (solubility columns) can be used for achieving a suitable concentrated stock solution. The use of a solvent carrier is not recommended. However, in case a solvent is necessary, a solvent control should be run in parallel, at the same solvent concentration as the chemical treatments. For difficult to test chemicals, a solvent may be technically the best solution; the OECD guidance document on aquatic toxicity testing of difficult substances and mixtures should be consulted (23). The choice of solvent will be determined by the chemical properties of the substance or mixture. The OECD guidance document recommends a maximum of 100 $\mu\text{l/l}$, which should be observed. However a recent review (24) highlighted additional concerns when using solvents for endocrine activity testing. Therefore it is recommended that the solvent concentration, if necessary, is minimised wherever technically feasible (dependent on the physical-chemical properties of the test chemical).
17. A flow-through test system will be used. Such a system continually dispenses and dilutes a stock solution of the test chemical (e.g. metering pump, proportional diluter, saturator system) in order to deliver a series of concentrations to the test chambers. The flow rates of stock solutions and dilution water should be checked at intervals, preferably daily, during the test and should not vary by more than 10 % throughout the test. Care should be taken to avoid the use of low-grade plastic tubing or other materials that may contain biologically active chemicals. When selecting the material for the flow-through system, possible adsorption of the test chemical to this material should be considered.

Holding of fish

18. Test fish should be selected from a laboratory population, preferably from a single 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. It is important that the loading rate and stocking density (for definitions, see Appendix 1) be appropriate for the test species used (see Appendix 2).
19. Following a 48-hour settling-in period, mortalities are recorded and the following criteria applied:
 - mortalities of greater than 10 % of population in seven days: reject the entire batch;
 - mortalities of between 5 % and 10 % of population: acclimation for seven additional days; if more than 5 % mortality during second seven days, reject the entire batch;
 - mortalities of less than 5 % of population in seven days: accept the batch.
20. Fish should not receive treatment for disease during the acclimation period, in the pre-exposure period, or during the exposure period.

Pre-exposure and selection of fish

21. The one to two-week pre-exposure period is recommended with animals placed in vessels similar to the actual test. Fish should be fed *ad libitum* throughout the holding period and during the exposure phase. The exposure phase is started with sexually dimorphic adult fish from a laboratory supply of reproductively mature animals (e.g. with clear secondary sexual characteristics visible as far as fathead minnow and medaka are concerned), and actively spawning. For general guidance only (and not to be considered in isolation from observing the actual reproductive status of a given batch of fish), fathead minnows should be approximately 20 (\pm 2) weeks of age, assuming they have been cultured at 25 ± 2 °C throughout their lifespan. Japanese medaka should be approximately 16 (\pm 2) weeks of age, assuming they have been cultured at 25 ± 2 °C throughout their lifespan. Zebrafish should be approximately 16 (\pm 2) weeks of age, assuming they have been cultured at 26 ± 2 °C throughout their lifespan. Egg production should be assessed daily during the pre-exposure phase. It is recommended that spawning be observed in all replicate tanks prior to inclusion in the exposure phase of the assay. Quantitative guidance on desirable daily egg production cannot be provided at this stage, but it is relatively common to observe average spawns of > 10 eggs/female/day for each species. A randomised block design according to egg production output should be used to allocate replicates to the various experimental levels to ensure balanced distribution of replicates.

TEST DESIGN

22. Three concentrations of the test chemical, one control (water) and, if needed, one solvent control are used. The data may be analysed in order to determine statistically significant differences between treatment and control responses. These analyses will inform whether further longer term testing for adverse effects (namely, survival, development, growth and reproduction) is required for the chemical, rather than for use in risk assessment (25).
23. For zebrafish, on day 21 of the experiment, males and females from each treatment level (5 males and 5 females in each of the two replicates) and from the control(s) are sampled for the measurement of vitellogenin. For medaka, on day 21 of the experiment, males and females from each treatment level (3 males and 3 females in each of the four replicates) and from the control(s) are sampled for the measurement of vitellogenin and secondary sex characteristics. For fathead minnow, on day 21 of exposure, males and females (2 males and 4 females in each of the four replicates) and from the control(s) are sampled for the measurement of vitellogenin and secondary sex characteristics. Quantitative assessment of fecundity is required, and gonadal tissues should be fixed in whole or dissected for potential histopathological evaluation, if required.

Selection of test concentrations

24. For the purposes of this test, the highest test concentration should be set by the maximum tolerated concentration (MTC) determined from a range finder or from other toxicity data, or 10 mg/l, or the maximum solubility in water, whichever is lowest. The MTC is defined as the highest test concentration of the chemical which results in less than 10 % mortality. Using this approach assumes that there are existing empirical acute toxicity data or other toxicity data from which the MTC can be estimated. Estimating the MTC can be inexact and typically requires some professional judgment.
25. Three test concentrations, spaced by a constant factor not exceeding 10, and a dilution-water control (and solvent control if necessary) are required. A range of spacing factors between 3,2 and 10 is recommended.

PROCEDURE

Selection and weighing of test fish

26. It is important to minimise variation in weight of the fish at the beginning of the assay. Suitable size ranges for the different species recommended for use in this test are given in Appendix 2. For the whole batch of fish used in the test, the range in individual weights for male and female fish at the start of the test should be kept, if possible, within ± 20 % of the arithmetic mean weight of the same sex. It is recommended to weigh a subsample of the fish stock before the test in order to estimate the mean weight.

Conditions of exposure*Duration*

27. The test duration is 21 days, following a pre-exposure period. The recommended pre-exposure period is one to two weeks.

Feeding

28. Fish should be fed *ad libitum* with an appropriate food (Appendix 2) at a sufficient rate to maintain body condition. Care should be taken to avoid microbial growth and water turbidity. As a general guidance, the daily ration may be divided into two or three equal portions for multiple feeds per day, separated by at least three hours between each feed. A single larger ration is acceptable particularly for weekends. Food should be withheld from the fish for 12 hours prior to sampling/necropsy.
29. Fish food should be evaluated for the presence of contaminants such as organochlorine pesticides, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs). Food with an elevated level of phytoestrogens that would compromise the response of the assay to known oestrogen agonist (e.g. 17 β - estradiol) should be avoided.
30. Uneaten food and faecal material should be removed from the test vessels at least twice weekly, e.g. by carefully cleaning the bottom of each tank using a siphon.

Light and temperature

31. The photoperiod and water temperature should be appropriate for the test species (see Appendix 2).

Frequency of analytical determinations and measurements

32. 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 on the chemical stability in the test system. During the test, the concentrations of the test chemical are determined at regular intervals, as follows: the flow rates of diluent and toxicant stock solution should be checked preferably daily but as a minimum twice per week, and should not vary by more than 10 % throughout the test. It is recommended that the actual test chemical concentrations be measured in all vessels at the start of the test and at weekly intervals thereafter.
33. 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.

34. Samples may need to be filtered (e.g. using a 0,45 µm pore size) or centrifuged. If needed, then centrifugation is the recommended procedure. However, if the test material does not adsorb to filters, filtration may also be acceptable.
35. During the test, dissolved oxygen, temperature, and pH should be measured in all test vessels at least once per week. Total hardness and alkalinity should be measured in the controls and one vessel at the highest concentration at least once per week. Temperature should preferably be monitored continuously in at least one test vessel.

Observations

36. A number of general (e.g. survival) and biological responses (e.g. VTG levels) are assessed over the course of the assay or at termination of the assay. The daily quantitative monitoring of fecundity is required. Measurement and evaluation of these endpoints and their utility are described below.

Survival

37. Fish should be examined daily during the test period and any mortality should be recorded and the dead fish removed as soon as possible. Dead fish should not be replaced in either the control or treatment vessels. Sex of fish that die during the test should be determined by macroscopic evaluation of the gonads.

Behaviour and appearance

38. Any abnormal behaviour (relative to controls) should be noted; this might include signs of general toxicity including hyperventilation, uncoordinated swimming, loss of equilibrium, and atypical quiescence or feeding. Additionally external abnormalities (such as haemorrhage, discoloration) should be noted. Such signs of toxicity should be considered carefully during data interpretation since they may indicate concentrations at which biomarkers of endocrine activity are not reliable. Such behavioural observations may also provide useful qualitative information to inform potential future fish testing requirements. For example, territorial aggressiveness in normal males or masculinised females has been observed in fathead minnows under androgenic exposure; in zebrafish, the characteristic mating and spawning behaviour after the dawn onset of light is reduced or hindered by oestrogenic or anti-androgenic exposure.
39. Because some aspects of appearance (primarily colour) can change quickly with handling, it is important that qualitative observations be made prior to removal of animals from the test system. Experience to date with fathead minnows suggests that some endocrine active chemicals may initially induce changes in the following external characteristics: body colour (light or dark), coloration patterns (presence of vertical bands), and body shape (head and pectoral region). Therefore observations of physical appearance of the fish should be made over the course of the test, and at conclusion of the study

Fecundity

40. Daily quantitative observations of spawning should be recorded on a replicate basis. Egg production should be recorded as the number of eggs/surviving female/day on a replicate basis. Eggs will be removed daily from the test chambers. Spawning substrates should be placed in the test chamber for the fathead minnow and zebrafish to enable fish to spawn in normal conditions. Appendix 4 gives further details of recommended spawning substrates for zebrafish (Appendix 4A) and fathead minnow (Appendix 4B). It is not considered necessary to provide spawning substrate for medaka.

Humane killing of fish

41. At day 21, i.e. at termination of the exposure, the fish should be euthanised with appropriate amounts of Tricaine (Tricaine methane sulfonate, Metacain, 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; blood or tissue is then sampled for VTG determination, as explained in the vitellogenin section.

Observation of secondary sex characteristics

42. Some endocrine active chemicals may induce changes in specialised secondary sex characteristics (number of nuptial tubercles in male fathead minnow, papillary processes in male medaka). Notably, chemicals with certain modes of action may cause abnormal occurrence of secondary sex characteristic in animals of the opposite sex; for example, androgen receptor agonists, such as trenbolone, methyltestosterone and dihydrotestosterone, can cause female fathead minnows to develop pronounced nuptial tubercles or female medaka to develop papillary processes (11, 20, 21). It also has been reported that oestrogen receptor agonists can decrease nuptial tubercle numbers and size of the dorsal nape pad in adult males of fathead minnow (26, 27). Such gross morphological observations may provide useful qualitative and quantitative information to inform potential future fish testing requirements. The number and size of nuptial tubercles in fathead minnow and papillary processes in medaka can be quantified directly or more practically in preserved specimens. Recommended procedures for the evaluation of secondary sex characteristics in fathead minnow and medaka are available from Appendix 5A and Appendix 5B, respectively.

Vitellogenin (VTG)

43. Blood is collected from the caudal artery/vein with a heparinised microhematocrit capillary tubule, or alternatively by cardiac puncture with a syringe. Depending upon the size of the fish, collectable blood volumes generally range from 5 to 60 µl per individual for fathead minnows and 5-15 µl per individual for zebrafish. Plasma is separated from the blood via centrifugation, and stored with protease inhibitors at – 80 °C, until analysed for VTG. Alternatively, in medaka the liver will be used, and in zebrafish the head/tail homogenate can be used as tissue-source for VTG determination (Appendix 6). The measurement of VTG should be based upon a validated homologous ELISA method, using homologous VTG standard and homologous antibodies. It is recommended to use a method capable to detect VTG levels as low as few ng/ml plasma (or ng/mg tissue), which is the background level in unexposed male fish.
44. Quality control of VTG analysis will be accomplished through the use of standards, blanks and at least duplicate analyses. For each ELISA method, a test for matrix effect (effect of sample dilution) should be run to determine the minimum sample dilution factor. Each ELISA plate used for VTG assays should include the following quality control samples: at least 6 calibration standards covering the range of expected VTG concentrations, and at least one non-specific binding assay blank (analysed in duplicate). Absorbance of these blanks should be less than 5 % of the maximum calibration standard absorbance. At least two aliquots (well-duplicates) of each sample dilution will be analysed. Well-duplicates that differ by more than 20 % should be re-analysed.
45. The correlation coefficient (R^2) for calibration curves should be greater than 0,99. However, a high correlation is not sufficient to guarantee adequate prediction of concentration in all ranges. In addition to having a sufficiently high correlation for the calibration curve, the concentration of each standard, as calculated from the calibration curve, should all fall between 70 and 120 % of its nominal concentration. If the nominal concentrations trend away from the calibration regression line (e.g. at lower concentrations), it may be necessary to split the calibration curve into low and high ranges or to use a nonlinear model to adequately fit the absorbance data. If the curve is split, both line segments should have $R^2 > 0,99$.
46. The limit of detection (LOD) is defined as the concentration of the lowest analytical standard, and limit of quantitation (LOQ) is defined as the concentration of the lowest analytical standard multiplied by the lowest dilution factor.
47. On each day that VTG assays are performed, a fortification sample made using an inter-assay reference standard will be analysed (Appendix 7). The ratio of the expected concentration to the measured concentration will be reported along with the results from each set of assays performed on that day.

Evaluation of gonadal histopathology

48. Performance of gonadal histopathology may be required by regulatory authorities to study the target organ on the HPG axis following chemical exposure. In this respect, gonads are fixed either whole body or dissected. When histopathology is required, specific endocrine-related responses on the gonads will be looked for in the assessment of the endocrine activity of the test chemical. These diagnostic responses essentially include the presence of testicular oocytes, Leydig cell hyperplasia, decreased yolk formation, increased spermatogonia and perfollicular hyperplasia. Other gonadal lesions like oocyte atresia, testicular degeneration, and stage changes, may have various causes. The Guidance document on fish gonadal histopathology specifies procedures that will be used in the dissection, fixation, sectioning and histopathological evaluation of the gonads (22).

DATA AND REPORTING

Evaluation of Biomarker Responses by Analysis of Variance (ANOVA)

49. To identify potential activity of a chemical, responses are compared between treatments and control groups using analysis of variance (ANOVA). Where a solvent control is used, an appropriate statistical test should be performed between the dilution water and solvent controls for each endpoint. Guidance on how to handle dilution water and solvent control data in the subsequent statistical analysis can be found in OECD, 2006c (28). All biological response data should be analysed and reported separately by sex. If the required assumptions for parametric methods are not met — non-normal distribution (e.g. Shapiro-Wilk's test) or heterogeneous variance (Bartlett's test or Levene's test), consideration should be given to transforming the data to homogenise variances prior to performing the ANOVA, or to carrying out a weighted ANOVA. Dunnett's test (parametric) on multiple pair-wise comparisons or a Mann-Whitney with Bonferroni adjustment (non-parametric) may be used for non-monotonous dose-response. Other statistical tests may be used (e.g. Jonckheere-Terpstra test or Williams test) if the dose-response is approximately monotone. A statistical flowchart is provided in Appendix 8 to help in the decision on the most appropriate statistical test to be used. Additional information can also be obtained from the OECD Document on Current Approaches to Statistical Analysis of Ecotoxicity Data (28).

Reporting of test results

50. Study data should include:

Testing facility:

- Responsible personnel and their study responsibilities
- Each laboratory should have demonstrated proficiency using a range of representative chemicals

Test Chemical:

- Characterisation of test chemical
- Physical nature and relevant physicochemical properties
- Method and frequency of preparation of test concentrations
- Information on stability and biodegradability

Solvent:

- Characterisation of solvent (nature, concentration used)
- Justification of choice of solvent (if other than water)

Test animals:

- Species and strain
- Supplier and specific supplier facility
- Age of the fish at the start of the test and reproductive/spawning status
- Details of animal acclimation procedure
- Body weight of the fish at the start of the exposure (from a sub-sample of the fish stock)

Test Conditions:

- Test procedure used (test-type, loading rate, stocking density, etc.);
- Method of preparation of stock solutions and flow-rate;
- The nominal test concentrations, weekly measured concentrations of the test solutions and analytical method used, means of the measured values and standard deviations in the test vessels and evidence that the measurements refer to the concentrations of the test chemical in true solution;
- Dilution water characteristics (including pH, hardness, alkalinity, temperature, dissolved oxygen concentration, residual chlorine levels, total organic carbon, suspended solids and any other measurements made)
- Water quality within test vessels: pH, hardness, temperature and dissolved oxygen concentration;
- Detailed information on feeding (e.g. type of food(s), source, amount given and frequency and analyses for relevant contaminants if available (e.g. PCBs, PAHs and organochlorine pesticides).

Results

- Evidence that the controls met the acceptance criteria of the test;
- Data on mortalities occurring in any of the test concentrations and control;
- Statistical analytical techniques used, treatment of data and justification of techniques used;
- Data on biological observations of gross morphology, including secondary sex characteristics, egg production and VTG;
- Results of the data analyses preferably in tabular and graphical form;
- Incidence of any unusual reactions by the fish and any visible effects produced by the test chemical

GUIDANCE FOR THE INTERPRETATION AND ACCEPTANCE OF THE TEST RESULTS

51. This section contains a few considerations to be taken into account in the interpretation of test results for the various endpoints measured. The results should be interpreted with caution where the test chemical appears to cause overt toxicity or to impact on the general condition of the test animal.
52. In setting the range of test concentrations, care should be taken not to exceed the maximum tolerated concentration to allow a meaningful interpretation of the data. It is important to have at least one treatment where there are no signs of toxic effects. Signs of disease and signs of toxic effects should be thoroughly assessed and reported. For example, it is possible that production of VTG in females can also be affected by general toxicity and non-endocrine toxic modes of action, e.g. hepatotoxicity. However, interpretation of effects may be strengthened by other treatment levels that are not confounded by systemic toxicity.

53. There are a few aspects to consider for the acceptance of test results. As a guide, the VTG levels in control groups of males and females should be distinct and separated by about three orders of magnitude in fathead minnow and zebrafish, and about one order of magnitude for medaka. Examples of the range of values encountered in control and treatment groups are available in the validation reports (1, 2, 3, 4). High VTG values in control males could compromise the responsiveness of the assay and its ability to detect weak oestrogen agonists. Low VTG values in control females could compromise the responsiveness of the assay and its ability to detect aromatase inhibitors and oestrogen antagonists. The validation studies were used to build that guidance.
54. Concerning the quantification of egg production, this is subject to important variations [the coefficient of variation (CV) may range from 20 to 60 %] that impinge the ability of the assay to detect a significant decrease in egg production smaller than 70 % as the CV approaches 50 % or more. When the CV is confined to lower values (around 20-30 %), then the assay will have acceptable power (80 %) to detect 40-50 % decrease in egg production. The test design used for the fathead minnow, including four replicates per treatment level, should allow more power to the fecundity endpoint, compared to a test design with 2 replicates only.
55. If a laboratory has not performed the assay before or substantial changes (e.g. change of fish strain or supplier) have been made it is advisable that a technical proficiency study is conducted. It is recommended that chemicals covering a range of modes of action or impacts on a number of the test endpoints are used. In practice, each laboratory is encouraged to build its own historical control data for males and females and to perform a positive control chemical for estrogenic activity (e.g. 17 β -estradiol at 100 ng/l, or a known weak agonist) resulting in increased VTG in male fish, a positive control chemical for aromatase inhibition (e.g. fadrozole or prochloraz at 300 μ g/l) resulting in decreased VTG in female fish, and a positive control chemical for androgenic activity (e.g. 17 β -trenbolone at 5 μ g/l) resulting in induction of secondary sex characteristics in female fathead minnow and medaka. All these data can be compared to available data from the validation studies (1, 2, 3) to ensure laboratory proficiency.
56. In general, VTG measurements should be considered positive if there is a statistically significant increase in VTG in males ($p < 0,05$), or a statistically significant decrease in females ($p < 0,05$) at least at the highest dose tested compared to the control group, and in the absence of signs of general toxicity. A positive result is further supported by the demonstration of a biologically plausible relationship between the dose and the response curve. As mentioned earlier, the VTG decrease may not entirely be of endocrine origin; however a positive result should generally be interpreted as evidence of endocrine activity *in vivo*, and should normally initiate actions for further clarification.
57. Gonadal histopathology evaluation may be required by regulatory authorities to determine the reproductive fitness of the test animals and to allow a weight of evidence assessment of the test results. Performance of gonadal histopathology may not be necessary in cases where either, VTG or secondary sex characteristics is positive (i.e. VTG increase or decrease, or induction of secondary sex characteristics).

LITERATURE:

- (1) OECD (2006a). Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine active Substances (Phase 1A). OECD Environmental Health and Safety Publications Series on Testing and Assessment No.60, Paris.
- (2) OECD (2006b). Report of the Initial Work Towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine active Substances (Phase 1B). OECD Environmental Health and Safety Publications Series on Testing and Assessment No.61, Paris.
- (3) OECD (2007). Final report of the Validation of the 21-day Fish Screening Assay for the Detection of Endocrine Active Substances. Phase 2: Testing Negative Substances. OECD Environmental Health and Safety Publications Series on Testing and Assessment No.78, Paris.

- (4) Owens JW (2007). Phase 3 report of the validation of the OECD Fish Screening Assay. CEFIC LRI Project, Endocrine. <http://www.cefic-lri.org/index.php?page=projects> (accessed 18/09/08).
- (5) US EPA (2007). Validation of the Fish Short-Term Reproduction Assay: Integrated Summary Report. 15 December 2007. US Environmental Protection Agency, Washington, DC. 104 pp.
- (6) OECD (2008). Report of the Validation Peer Review for the 21-Day Fish Endocrine Screening Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report. OECD Environmental Health and Safety Publications Series on Testing and Assessment No.94, Paris.
- (7) Sumpter J.P. and S. Jobling (1995). Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. *Environmental Health Perspectives*; 103 Suppl 7:173-8 Review.
- (8) Pawlowski S., *et al.* (2004). Androgenic and estrogenic effects of the synthetic androgen 17alpha-methyltestosterone on sexual development and reproductive performance in the fathead minnow (*Pimephales promelas*) determined using the gonadal recrudescence assay. *Aquatic Toxicology*; 68(3):277-91.
- (9) Andersen L., *et al.* (2006). Short-term exposure to low concentrations of the synthetic androgen methyltestosterone affects vitellogenin and steroid levels in adult male zebrafish (*Danio rerio*). *Aquatic Toxicology*; 76(3-4):343-52.
- (10) Ankley G.T., *et al.* (2002). Evaluation of the aromatase inhibitor fadrozole in a short-term reproduction assay with the fathead minnow (*Pimephales promelas*). *Toxicological Sciences*; 67(1):121-30.
- (11) Panter G.H., *et al.* (2004). Successful detection of (anti-)androgenic and aromatase inhibitors in pre-spawning adult fathead minnows (*Pimephales promelas*) using easily measured endpoints of sexual development. *Aquatic Toxicology*; 70(1):11-21.
- (12) Parks L.G., *et al.* (1999). Fathead minnow (*Pimephales promelas*) vitellogenin: purification, characterization and quantitative immunoassay for the detection of estrogenic compounds. *Comparative Biochemistry and Physiology. Part C Pharmacology, toxicology and endocrinology*; 123(2):113-25.
- (13) Panter G.H., *et al.* (1999). Application of an ELISA to quantify vitellogenin concentrations in fathead minnows (*Pimephales promelas*) exposed to endocrine disrupting chemicals. CEFIC-EMSG research report reference AQ001. CEFIC, Brussels, Belgium.
- (14) Fenske M., *et al.* (2001). Development and validation of a homologous zebrafish (*Danio rerio* Hamilton-Buchanan) vitellogenin enzyme-linked immunosorbent assay (ELISA) and its application for studies on estrogenic chemicals. *Comp. Biochem. Phys. C* 129 (3): 217-232.
- (15) Holbech H., *et al.* (2001). Development of an ELISA for vitellogenin in whole body homogenate of zebrafish (*Danio rerio*). *Comparative Biochemistry and Physiology. Part C Pharmacology, toxicology and endocrinology*; 130: 119-131
- (16) Rose J., *et al.* (2002). Vitellogenin induction by 17β-estradiol and 17α-ethinylestradiol in male zebrafish (*Danio rerio*). *Comp. Biochem. Physiol. C* 131: 531-539.
- (17) Brion F., *et al.* (2002). Development and validation of an enzyme-linked immunosorbent assay to measure vitellogenin in the zebrafish (*Danio rerio*). *Environmental Toxicology and Chemistry*; vol 21: 1699-1708.
- (18) Yokota H., *et al.* (2001). Development of an ELISA for determination of the hepatic vitellogenin in Medaka (*Oryzias latipes*). *Jpn J Environ Toxicol* 4:87-98.
- (19) Tatarazako N., *et al.* (2004). Validation of an enzyme-linked immunosorbent assay method for vitellogenin in the Medaka. *Journal of Health Science* 50:301-308.
- (20) Ankley G.T., *et al.* (2003). Effects of the androgenic growth promoter 17-beta-trenbolone on fecundity and reproductive endocrinology of the fathead minnow. *Environmental Toxicology and Chemistry*; 22(6): 1350-60.

-
- (21) Seki M, *et al* (2004). Fish full life-cycle testing for androgen methyltestosterone on medaka (*Oryzias latipes*). *Environmental Toxicology and Chemistry*; 23(3):774-81.
- (22) OECD (2010). Guidance Document on Fish Gonadal Histopathology. OECD Environmental Health and Safety Publications Series on Testing and Assessment No. 123, Paris.
- (23) OECD (2000) Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. Environmental Health and Safety Publications. Series on Testing and Assessment. No. 23. Paris
- (24) Hutchinson T.H., *et al.* (2006a). Acute and chronic effects of carrier solvents in aquatic organisms: A critical review. *Review. Aquatic Toxicology*, 76; pp.69–92.
- (25) Hutchinson T.H., *et al.* (2006b). Screening and testing for endocrine disruption in fish-biomarkers as “signposts”, not “traffic lights,” in risk assessment. *Environmental Health Perspectives*; 114 Suppl 1:106-14.
- (26) Miles-Richardson S.R., *et al.* (1999). Effects of waterborne exposure to 17 β -estradiol on secondary sex characteristics and gonads of the fathead minnow (*Pimephales promelas*). *Aquat. Toxicol.* 47, 129-145.
- (27) Martinovic D., *et al.* (2008). Characterization of reproductive toxicity of vinclozolin in the fathead minnow and co-treatment with an androgen to confirm an anti-androgenic mode of action. *Environ. Toxicol. Chem.* 27, 478-488.
- (28) OECD (2006c), Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application, OECD Environmental Health and Safety Publications, Series on Testing and Assessment No.54, OECD, Paris.
- (29) US EPA (2008), Peer-Review Results for the Fish Short-Term Reproduction Assay, dated 30 January 2008, US Environmental Protection Agency, Washington DC. 110 pp.
- (30) OECD (2012), *OECD Conceptual Framework for Testing and Assessment of Endocrine Disruptors*, OECD Environmental Health and Safety Publications, Series on Testing and Assessment No. 150, OECD, Paris.
-

Appendix 1

ABBREVIATIONS & DEFINITIONS:

Chemical: a substance or a mixture

CV: coefficient of variation

ELISA: Enzyme-Linked Immunosorbent Assay

HPG axis: hypothalamic-pituitary-gonadal axis

Loading rate: the wet weight of fish per volume of water.

MTC: Maximum Tolerated Concentration, representing about 10 % of the LC₅₀

Stocking density: is the number of fish per volume of water.

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

VTG: vitellogenin is a phospholipoglycoprotein precursor to egg yolk protein that normally occurs in sexually active females of all oviparous species.

Appendix 2

EXPERIMENTAL CONDITIONS FOR THE FISH ENDOCRINE SCREENING ASSAY

1. Recommended species	Fathead minnow (<i>Pimephales promelas</i>)	Medaka (<i>Oryzias latipes</i>)	Zebrafish (<i>Danio rerio</i>)
2. Test type	Flow-through	Flow-through	Flow-through
3. Water temperature	25 ± 2 °C	25 ± 2 °C	26 ± 2 °C
4. Illumination quality	Fluorescent bulbs (wide spectrum)	Fluorescent bulbs (wide spectrum)	Fluorescent bulbs (wide spectrum)
5. Light intensity	10-20 µE/m ² /s, 540-1 000 lux, or 50-100 ft-c (ambient laboratory levels)	10-20 µE/m ² /s, 540-1 000 lux, or 50-100 ft-c (ambient laboratory levels)	10-20 µE/m ² /s, 540-1 000 lux, or 50-100 ft-c (ambient laboratory levels)
6. Photoperiod (dawn / dusk transitions are optional, however not considered necessary)	16 h light, 8 h dark	12-16 h light, 12-8 h dark	12-16 h light, 12-8 h dark
7. Loading rate	< 5 g per l	< 5 g per l	< 5 g per l
8. Test chamber size	10 l (minimum)	2 l (minimum)	5 l (minimum)
9. Test solution volume	8 l (minimum)	1,5 l (minimum)	4 l (minimum)
10. Volume exchanges of test solutions	Minimum of 6 daily	Minimum of 5 daily	Minimum of 5 daily
11. Age of test organisms	See paragraph 21	See paragraph 21	See paragraph 21
12. Approximate wet weight of adult fish (g)	Females: 1,5 ± 20 % Males: 2,5 ± 20 %	Females: 0,35 ± 20 % Males: 0,35 ± 20 %	Females: 0,65 ± 20 % Males: 0,4 ± 20 %
13. No. of fish per test vessel	6 (2 males and 4 females)	6 (3 males and 3 females)	10 (5 males and 5 females)
14. No. of treatments	= 3 (plus appropriate controls)	= 3 (plus appropriate controls)	= 3 (plus appropriate controls)
15. No. vessels per treatment	4 minimum	4 minimum	2 minimum
16. No. of fish per test concentration	16 adult females and 8 males (4 females and 2 males in each replicate vessel)	12 adult females and 12 males (3 females and 3 males in each replicate vessel)	10 adult females and 10 males (5 females and 5 males in each replicate vessel)

17. Feeding regime	Live or frozen adult or nauplii brine shrimp two or three times daily (<i>ad libitum</i>), commercially available food or a combination of the above	Brine shrimp nauplii two or three times daily (<i>ad libitum</i>), commercially available food or a combination of the above	Brine shrimp nauplii two or three times daily (<i>ad libitum</i>), commercially available food or a combination of the above
18. Aeration	None unless DO concentration falls below 60 % air saturation	None unless DO concentration falls below 60 % air saturation	None unless DO concentration falls below 60 % air saturation
19. Dilution water	Clean surface, well or reconstituted water or dechlorinated tap water	Clean surface, well or reconstituted water or dechlorinated tap water	Clean surface, well or reconstituted water or dechlorinated tap water
20. Pre- exposure period	7-14 days recommended	7-14 days recommended	7-14 days recommended
21. Chemical exposure duration	21-d	21-d	21-d
22. Biological endpoints	<ul style="list-style-type: none"> — survival — behaviour — fecundity — 2y sex characteristics — VTG — optionally gonadal histopathology 	<ul style="list-style-type: none"> — survival — behaviour — fecundity — 2y sex characteristics — VTG — optionally gonadal histopathology 	<ul style="list-style-type: none"> — survival — behaviour — fecundity — VTG — optionally gonadal histopathology
23. Test acceptability	Dissolved oxygen ≥ 60 % of saturation; mean temperature of 25 ± 2 °C; 90 % survival of fish in the controls; measured test concentrations within 20 % of mean measured values per treatment level.	Dissolved oxygen ≥ 60 % of saturation; mean temperature of 25 ± 2 °C; 90 % survival of fish in the controls; measured test concentrations within 20 % of mean measured values per treatment level.	Dissolved oxygen ≥ 60 % of saturation; mean temperature of 26 ± 2 °C; 90 % survival of fish in the controls; measured test concentrations within 20 % of mean measured values per treatment level.

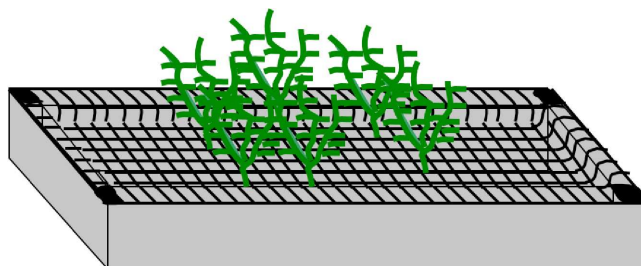
*Appendix 3***SOME CHEMICAL CHARACTERISTICS OF ACCEPTABLE DILUTION WATER**

COMPONENT	CONCENTRATIONS
Particulate matter	< 20 mg/l
Total organic carbon	< 2 mg/l
Unionised ammonia	< 1 µg/l
Residual chlorine	< 10 µg/l
Total organophosphorus pesticides	< 50 ng/l
Total organochlorine pesticides plus polychlorinated biphenyls	< 50 ng/l
Total organic chlorine	< 25 ng/l

Appendix 4A

SPAWNING SUBSTRATE FOR ZEBRAFISH

Spawning tray: all glass instrument dish, for example $22 \times 15 \times 5,5$ cm (l × w × d), covered with a removable stainless steel wire lattice (mesh width 2mm). The lattice should cover the opening of the instrument dish at a level below the brim.



On the lattice, spawning substrate should be fixed. It should provide structure for the fish to move into. For example, artificial aquaria plants made of green plastic material are suitable (NB: possible adsorption of the test chemical to the plastic material should be considered). The plastic material should be leached out in sufficient volume of warm water for sufficient time to ensure that no chemicals may be disposed to the test water. When using glass materials it should be ensured that the fish are neither injured nor cramped during their vigorous actions.

The distance between the tray and the glass panes should be at least 3 cm to ensure that the spawning is not performed outside the tray. The eggs spawned onto the tray fall through the lattice and can be sampled 45-60 min after the start of illumination. The transparent eggs are non-adhesive and can easily be counted by using transversal light. When using five females per vessel, egg numbers up to 20 at a day can be regarded as low, up to 100 as medium and more than 100 as high numbers. The spawning tray should be removed, the eggs collected and the spawning tray re-introduced in the test vessel, either as late as possible in the evening or very early in the morning. The time until re-introduction should not exceed one hour since otherwise the cue of the spawning substrate may induce individual mating and spawning at an unusual time. If a situation needs a later introduction of the spawning tray, this should be done at least 9 hours after start of the illumination. At this late time of the day, spawning is not induced any longer.

Appendix 4B

SPAWNING SUBSTRATE FOR FATHEAD MINNOW

Two or three combined plastic/ceramic/glass or stainless steel spawning tiles and trays are placed in each of the test chamber (e.g. 80 mm length of grey semi-circular guttering sitting on a lipped tray of 130mm length) (see picture). Properly seasoned PVC or ceramic tiles have demonstrated to be appropriate for a spawning substrate (Thorpe *et al*, 2007).

It is recommended that the tiles are abraded to improve adhesion. The tray should also be screened to prevent fish from access to the fallen eggs unless the egg adhesion efficiency has been demonstrated for the spawning substrate used.



The base is designed to contain any eggs that do not adhere to the tile surface and would therefore fall to the bottom of the tank (or those eggs laid directly onto the flat plastic base). All spawning substrates should be leached for a minimum of 12 hours, in dilution water, before use.

Thorpe KL, Benstead R, Hutchinson TH, Tyler CR, 2007. An optimised experimental test procedure for measuring chemical effects on reproduction in the fathead minnow, *Pimephales promelas*. *Aquatic Toxicology*, 81, 90–98.

Appendix 5A

ASSESSMENT OF SECONDARY SEX CHARACTERISTICS IN FATHEAD MINNOW FOR THE DETECTION OF CERTAIN ENDOCRINE ACTIVE CHEMICALS**Overview**

Potentially important characteristics of physical appearance in adult fathead minnows in endocrine disrupter testing include body colour (i.e., light/dark), coloration patterns (i.e., presence or absence of vertical bands), body shape (i.e., shape of head and pectoral region, distension of abdomen), and specialised secondary sex characteristics (i.e., number and size of nuptial tubercles, size of dorsal pad and ovipositor).

Nuptial tubercles are located on the head (dorsal pad) of reproductively-active male fathead minnows, and are usually arranged in a bilaterally-symmetric pattern (Jensen *et al.* 2001). Control females and juvenile males and females exhibit no tubercle development (Jensen *et al.* 2001). There can be up to eight individual tubercles around the eyes and between the nares of the males. The greatest numbers and largest tubercles are located in two parallel lines immediately below the nares and above the mouth. In many fish there are groups of tubercles below the lower jaw; those closest to the mouth generally occur as a single pair, while the more ventral set can be comprised of up to four tubercles. The actual numbers of tubercles is seldom more than 30 (range, 18-28; Jensen *et al.* 2001). The predominant tubercles (in terms of numbers) are present as a single, relatively round structure, with the height approximately equivalent to the radius. Most reproductively-active males also have, at least some, tubercles which are enlarged and pronounced such that they are indistinguishable as individual structures.

Some types of endocrine-disrupting chemicals can cause the abnormal occurrence of certain secondary sex characteristics in the opposite sex; for example, androgen receptor agonists, such as 17 α -methyltestosterone or 17 β -trenbolone, can cause female fathead minnows to develop nuptial tubercles (Smith 1974; Ankley *et al.* 2001; 2003), while oestrogen receptor agonists may decrease number or size of nuptial tubercles in males (Miles-Richardson *et al.* 1999; Harries *et al.* 2000).

Below is a description of the characterisation of nuptial tubercles in fathead minnows based on procedures used at the U.S. Environmental Protection Agency lab in Duluth, MN. Specific products and/or equipment can be substituted with comparable materials available.

Viewing is best accomplished using an illuminated magnifying glass or 3X illuminated dissection scope. View fish dorsally and anterior forward (head toward viewer).

- Place fish in small Petri dish (e.g. 100 mm in diameter), anterior forward, and ventral down. Focus viewfinder to allow identification of tubercles. Gently and slowly roll fish from side to side to identify tubercle areas. Count and score tubercles.
- Repeat the observation on the ventral head surface by placing the fish dorsal anterior forward in the Petri dish.
- Observations should be completed within 2 min for each fish.

Tubercle Counting and Rating

Six specific areas have been identified for assessment of tubercle presence and development in adult fathead minnows. A template was developed to map the location and quantity of tubercles present (see end of this appendix). The number of tubercles is recorded and their size can be quantitatively ranked as: 0- absence, 1-present, 2-enlarged and 3-pronounced for each organism (Fig. 1).

Rate 0 — absence of any tubercle. Rating 1 — present, is identified as any tubercle having a single point whose height is nearly equivalent to its radius. Rating 2 — enlarged, is identified by tissue resembling an asterisk in appearance, usually having a large radial base with grooves or furrows emerging from the centre. Tubercle height is often more jagged but can be somewhat rounded at times. Rating 3 — pronounced, is usually quite large and rounded with less definition in structure. At times these tubercles will run together forming a single mass along an individual or combination of areas (B, C and D, described below). Coloration and design are similar to rating 2 but at times are fairly indiscriminate. Using this rating system generally will result in overall tubercle scores of < 50 in a normal control male possessing a tubercle count of 18 to 20 (Jensen *et al.* 2001).

Figure 1



The actual number of tubercles in some fish may be greater than the template boxes for a particular rating area. If this happens, additional rating numbers may be marked within, to the right or to the left of the box. The template therefore does not need to display symmetry. An additional technique for mapping tubercles which are paired or joined vertically along the horizontal plane of the mouth could be done by double-marking two tubercle rating points in a single box.

Mapping regions:

A — Tubercles located around eye. Mapped dorsal to ventral around anterior rim of eye. Commonly multiple in mature control males, not present in control females, generally paired (one near each eye) or single in females exposed to androgens.

B — Tubercles located between nares, (sensory canal pores). Normally in pairs for control males at more elevated levels (2- enlarged or 3- pronounced) of development. Not present in control females with some occurrence and development in females exposed to androgens.

C — Tubercles located immediately anterior to nares, parallel to mouth. Generally enlarged or pronounced in mature control males. Present or enlarged in less developed males or androgen-treated females.

D — Tubercles located parallel along mouth line. Generally rated developed in control males. Absent in control females but present in androgen-exposed females.

E — Tubercles located on lower jaw, close to mouth, usually small and commonly in pairs. Varying in control or treated males, and treated females.

F — Tubercles located ventral to E. Commonly small and paired. Present in control males and androgen-exposed females.

REFERENCES:

- (1) Ankley GT, Jensen KM, Kahl MD, Korte JJ, Makynen ME. 2001. Description and evaluation of a short-term reproduction test with the fathead minnow (*Pimephales promelas*). *Environ Toxicol Chem* 20:1276-1290.
- (2) Ankley GT, Jensen KM, Makynen EA, Kahl MD, Korte JJ, Hornung MW, Henry TR, Denny JS, Leino RL, Wilson VS, Cardon MC, Hartig PC, Gray EL. 2003. Effects of the androgenic growth promoter 17- β trenbolone on fecundity and reproductive endocrinology of the fathead minnow. *Environ Toxicol Chem* 22:1350-1360.
- (3) Harries JE, Runnalls T, Hill E, Harris CA, Maddix S, Sumpter JP, Tyler CR. 2000. Development of a reproductive performance test for endocrine disrupting chemicals using pair-breeding fathead minnows (*Pimephales promelas*). *Environ Sci Technol* 34:3003-3011.

- (4) Jensen KM, Korte JJ, Kahl MD, Pasha MS, Ankley GT. 2001. Aspects of basic reproductive biology and endocrinology in the fathead minnow (*Pimephales promelas*). *Comp Biochem Physiol C* 128:127-141.
- (5) Kahl MD, Jensen KM, Korte JJ, Ankley GT. 2001. Effects of handling on endocrinology and reproductive performance of the fathead minnow. *J Fish Biol* 59:515-523.
- (6) Miles-Richardson SR, Kramer VJ, Fitzgerald SD, Render JA, Yamini B, Barbee SJ, Giesy JP. 1999. Effects of waterborne exposure of 17-estradiol on secondary sex characteristics and gonads of fathead minnows (*Pimephales promelas*). *Aquat Toxicol* 47:129-145.
- (7) Smith RJF. 1974. Effects of 17 α -methyltestosterone on the dorsal pad and tubercles of fathead minnows (*Pimephales promelas*). *Can J Zool* 52:1031-1038.

Tubercle Template:

ID _____

Date _____

Total Score _____

Numerical Rating

1-present

2-enlarged

3-pronounced

	A	X1	X1	X1	X1
--	---	----	----	----	----

	B	X1	X1	X1	X1
--	---	----	----	----	----

	C	X1	X1	X1	X1	X1	X1	X1	X1	X1	X1
	D	X1	X1	X1	X1	X1	X1	X1	X1	X1	X1

		E	X1	X1	
	F	X1	X1	X1	X1

Appendix 5B

ASSESSMENT OF SECONDARY SEX CHARACTERISTICS IN MEDAKA FOR THE DETECTION OF CERTAIN ENDOCRINE ACTIVE CHEMICALS

Below is a description of the measurement of papillary processes (*), which are the secondary sex characteristics in medaka (*Oryzias latipes*).

- (1) After the excision of the liver (Appendix 6), the carcass is placed into a conical tube containing about 10 ml of 10 % neutral buffered formalin (upside: head, downside: tail). If the gonad is fixed in a solution other than 10 % neutral buffered formalin, make a transverse cut across the carcass between anterior region of anal fin and anus using razor, taking care not to harm the gonopore and gonad itself (Fig.3). Place the cranial side of the fish body into the fixative solution to preserve the gonad, and the tail side of the fish body into the 10 % neutral buffered formalin as described above.
- (2) After placing the fish body into 10 % neutral buffered formalin, grasp the anterior region of the anal fin with tweezers and fold it for about 30 seconds to keep the anal fin open. When grasping the anal fin with tweezers, grasp a few fin rays in the anterior region with care not to scratch the papillary processes.
- (3) After keeping the anal fin open for about 30 seconds, store the fish body in 10 % neutral buffered formalin at room temperature until the measurement of the papillary processes (measurement should be conducted after fixing for at least 24 hours).

Measurement

- (1) After fixing the fish body in the 10 % neutral buffered formalin for at least 24 hours, pick up the fish carcass from the conical tube and wipe the formalin on the filter paper (or paper towel).
- (2) Place the fish abdomen side up. Then cut the anal fin using small dissection scissors carefully (it is preferable to cut the anal fin with small amount of pterygiophore).
- (3) Grasp the anterior region of the severed anal fin with tweezers and put it on a glass slide with a several drops of water. Then cover the anal fin with a cover glass. Be careful not to scratch the papillary processes when grasping the anal fin with tweezers.
- (4) Count the number of the joint plate with papillary processes using the counter under a biological microscope (upright microscope or inverted microscope). The papillary processes are recognised when a small formation of processes is visible on the posterior margin of joint plate. Write the number of joint plate with papillary processes in each fin ray to the worksheet (e.g. first fin ray: 0, second fin ray: 10, third fin ray: 12, etc.) and enter the sum of this number on the Excel spreadsheet by individual fish. If necessary, take a photograph of the anal fin and count the number of joint plate with papillary processes on the photograph.
- (5) After the measurement, put the anal fin into the conical tube described in (1) and store it.

(*) Papillary processes normally appear only in adult males and are found on fin rays from the second to the seventh or eighth counting from the posterior end of the anal fin (Fig.1 and 2). However, processes rarely appear on the first fin ray from the posterior end of the anal fin. This SOP covers the measurement of processes on the first fin ray (the fin ray number refers to the order from the posterior end of the anal fin in this SOP).

Fig.1.

Diagram showing sexual difference in shape and size of the anal fin. A, male; B, female. Oka, T. B., 1931. On the processes on the fin rays of the male of *Oryzias latipes* and other sex characters of this fish. J. Fac. Sci., Tokyo Univ., IV, 2: 209-218.

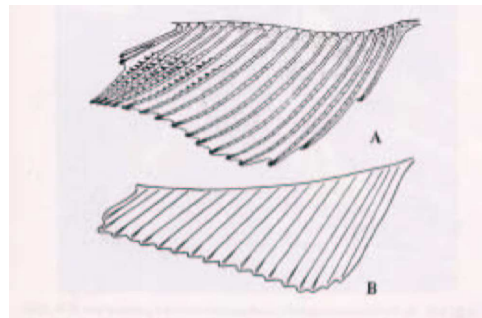


Fig.2.

A, Processes on joint plates of anal fin-ray. J.P., joint plate; A.S., axial space; P., process. B, Distal extremity of fin-ray. Actinotrichia (Act.) are on the tip. Oka, T. B., 1931. On the processes on the fin rays of the male of *Oryzias latipes* and other sex characters of this fish. J. Fac. Sci., Tokyo Univ., IV, 2: 209-218.

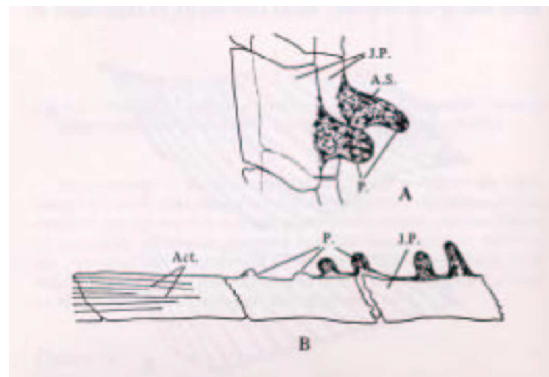
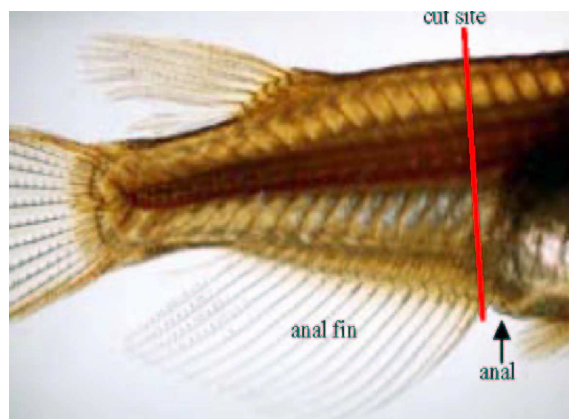


Fig.3.

Photograph of fish body showing the cut site when the gonad is fixed in the fixing solution other than 10 % neutral buffered formalin. In that case, the remaining body will be cut off between anterior region of anal fin and anal using razor (red bar), and the head side of fish body will be put into the fixing solution for gonad and the tail side of the fish body will be put into the 10 % neutral buffered formalin.



Appendix 6

RECOMMENDED PROCEDURES FOR SAMPLE COLLECTION FOR VITELLOGENIN ANALYSIS

Care should be taken to avoid cross-contamination between VTG samples of males and females.

Procedure 1A: Fathead Minnow, Blood Collection from the Caudal Vein/Artery

After anaesthetisation, the caudal peduncle is partially severed with a scalpel blade and blood is collected from the caudal vein/artery with a heparinised microhematocrit capillary tube. After the blood has been collected, the plasma is quickly isolated by centrifugation for 3 min at 15 000 g (or alternatively for 10 min. at 15 000 g at 4 ° C). If desired, percent haematocrit can be determined following centrifugation. The plasma portion is then removed from the microhematocrit tube and stored in a centrifuge tube with 0,13 units of aprotinin (a protease inhibitor) at – 80 °C until determination of VTG can be made. Depending on the size of the fathead minnow (which is sex-dependent), collectable plasma volumes generally range from 5 to 60 microliters per fish (Jensen *et al.* 2001).

Procedure 1B: Fathead Minnow, Blood Collection from Heart

Alternatively, blood may also be collected by cardiac puncture using a heparinised syringe (1 000 units of heparin per ml). The blood is transferred into Eppendorf tubes (held on ice) and then centrifuged (5 min, 7 000 g, room temperature). The plasma should be transferred into clean Eppendorf tubes (in aliquots if the volume of plasma makes this feasible) and promptly frozen at – 80 °C, until analysed (Panter *et al.*, 1998).

Procedure 2A: Japanese Medaka, Excision of the Liver in Medaka*Removal of the test fish from the test chamber*

- (1) Test fish should be removed from the test chamber using the small spoon-net. Be careful not to drop the test fish into other test chambers.
- (2) In principle, the test fish should be removed in the following order: control, solvent control (where appropriate), lowest concentration, middle concentration, highest concentration and positive control. In addition, all males should be removed from one test chamber before the remaining females are removed.
- (3) The sex of each test fish is identified on the basis of external secondary sex characteristics (e.g. the shape of the anal fin).
- (4) Place the test fish in a container for transport and carry it to the workstation for excision of the liver. Check the labels of the test chamber and the transport container for accuracy and to confirm that the number of fish that have been removed from the test chamber and that the number of fish remaining in the test chamber are consistent with expectation.
- (5) If the sex cannot be identified by the fish's external appearance, remove all fish from the test chamber. In this case, the sex should be identified by observing the gonad or secondary sex characteristics under a stereoscopic microscope.

Excision of the liver

- (1) Transfer the test fish from the container for transport to the anaesthetic solution using the small spoon-net.
- (2) After the test fish is anaesthetised, transfer the test fish on the filter paper (or a paper towel) using tweezers (commodity type). When grasping the test fish, apply the tweezers to the sides of the head to prevent breaking the tail.
- (3) Wipe the water on the surface of the test fish on the filter paper (or the paper towel).

- (4) Place the fish abdomen side up. Then make a small transverse incision partway between the ventral neck region and the mid-abdominal region using dissection scissors.
- (5) Insert the dissection scissors into the small incision, and incise the abdomen from a point caudal to the branchial mantle to the cranial side of the anus along the midline of the abdomen. Be careful not to insert the dissection scissors too deeply so as to avoid damaging the liver and gonad.
- (6) Conduct the following operations under the stereoscopic microscope.
- (7) Place the test fish abdomen side up on the paper towel (glass Petri dish or slide glass are also available).
- (8) Extend the walls of the abdominal cavity with precision tweezers and exteriorise the internal organs. It is also acceptable to exteriorise the internal organs by removing one side of the wall of the abdominal cavity if necessary.
- (9) Expose the connected portion of the liver and gallbladder using another pair of precision tweezers. Then grasp the bile duct and cut off the gallbladder. Be careful not to break the gallbladder.
- (10) Grasp the oesophagus and excise the gastrointestinal tract from the liver in the same way. Be careful not to leak the contents of the gastrointestinal tract. Excise the caudal gastrointestinal tract from the anus and remove the tract from the abdominal cavity.
- (11) Trim the mass of fat and other tissues from the periphery of the liver. Be careful not to scratch the liver.
- (12) Grasp the hepatic portal area using the precision tweezers and remove the liver from the abdominal cavity.
- (13) Place the liver on the slide glass. Using the precision tweezers, remove any additional fat and extraneous tissue (e.g. abdominal lining), if needed, from the surface of the liver.
- (14) Measure the liver weight with 1,5 ml microtube as a tare using an electronic analytical balance. Record the value on the worksheet (read: 0,1 mg). Confirm the identification information on the microtube label.
- (15) Close the cap of the microtube containing the liver. Store it in a cooling rack (or ice rack).
- (16) Following the excision of one liver, clean the dissection instruments or replace them with clean ones.
- (17) Remove livers from all of the fish in the transport container as described above.
- (18) After the livers have been excised from all of the fish in the transport container (i.e., all males or females in a test chamber), place all liver specimens in a tube rack with a label for identification and store it in a freezer. When the livers are donated for pre-treatment shortly after the excision, the specimens are carried to the next workstation in a cooling rack (or ice rack).

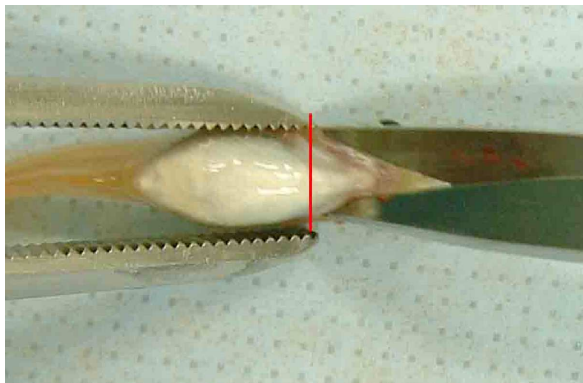
Following liver excision, the fish carcass is available for gonad histology and measurement of secondary sex characteristics.

Specimen

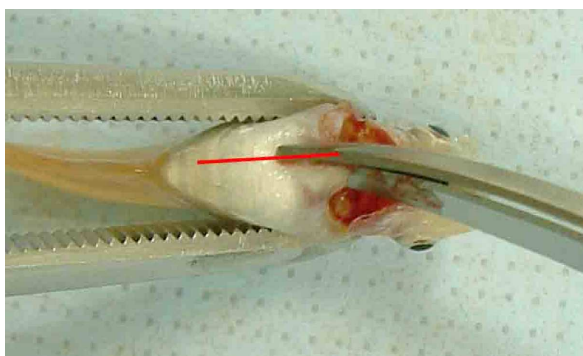
Store the liver specimens taken from the test fish at $\leq -70\text{ }^{\circ}\text{C}$ if they are not used for the pre-treatment shortly after the excision.

Figure 1

A cut is made just anterior to pectoral fins with scissors

*Figure 2*

The midline of abdomen is incised with scissors to a point approximately 2 mm cranial to the anus

*Figure 3*

The abdominal walls are spread with forceps for exposure of the liver and other internal organs

(Alternatively, the abdominal walls may be pinned laterally).

Arrow shows liver

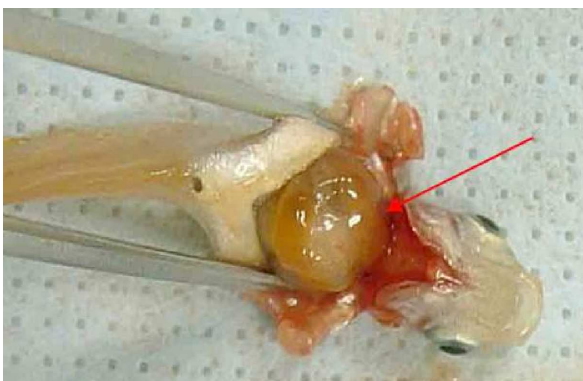
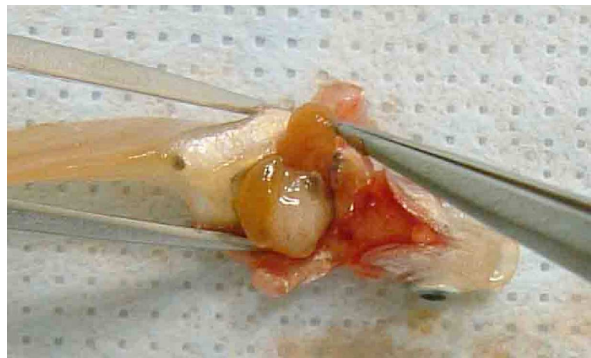
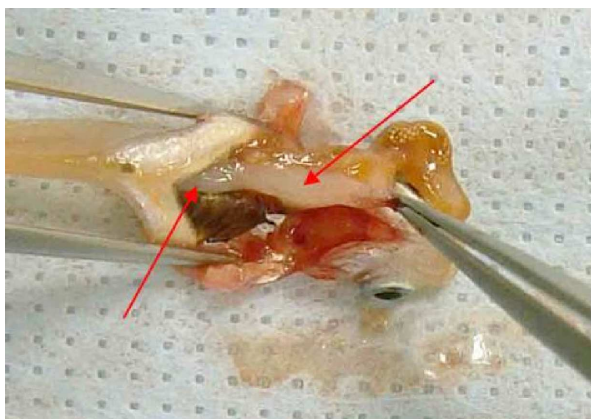


Figure 4

The liver is bluntly dissected and excised using forceps

*Figure 5*

The intestines are gently retracted using forceps

*Figure 6*

Both ends of the intestines and any mesenteric attachments are severed using scissors

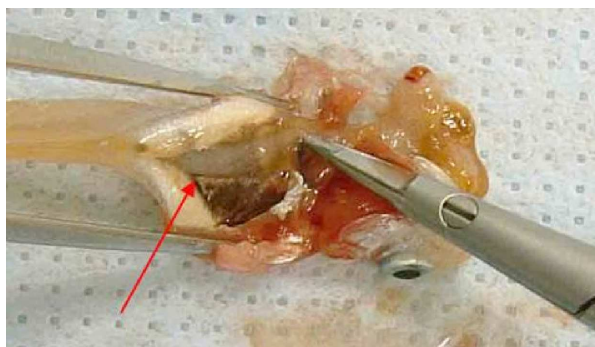
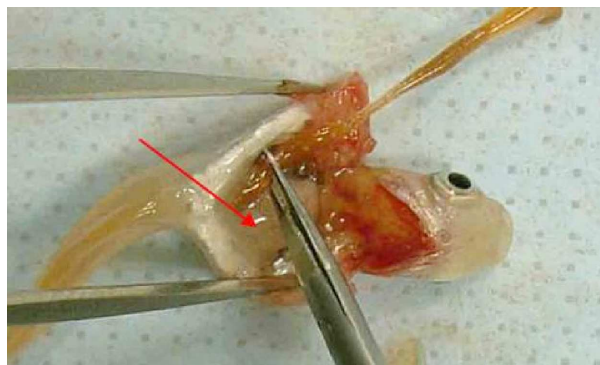
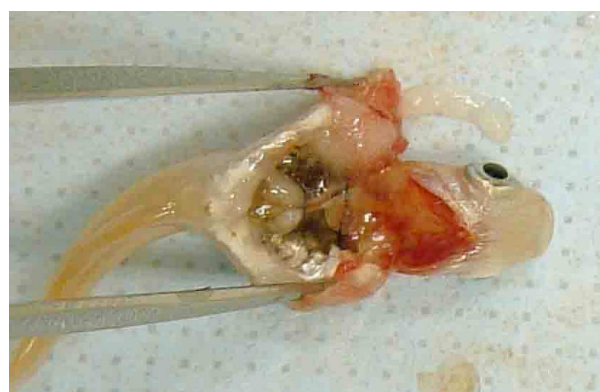


Figure 7 (female)

The procedure is identical for the female

*Figure 8*

The completed procedure



Procedure 2 B: Japanese Medaka (*Oryzias latipes*), Liver Pre-treatment for Vitellogenin Analysis

Take the bottle of homogenate buffer from the ELISA kit and cool it with crushed ice (temperature of the solution: $\leq 4^{\circ}\text{C}$). If homogenate buffer from EnBio ELISA system is used, thaw the solution at room temperature, and then cool the bottle with crushed ice.

Calculate the volume of homogenate buffer for the liver on the basis of its weight (add 50 μl of homogenate buffer per mg liver weight). For example, if the weight of the liver is 4,5 mg, the volume of homogenate buffer for the liver is 225 μl . Prepare a list of the volume of homogenate buffer for all livers.

Preparation of the liver for pre-treatment

- (1) Take the 1,5 ml microtube containing the liver from the freezer just before the pre-treatment.
- (2) Pre-treatment of the liver from males should be performed before females to prevent vitellogenin contamination. In addition, the pre-treatment for test groups should be conducted in the following order: control, solvent control (where appropriate), lowest concentration, middle concentration, highest concentration and positive control.

- (3) The number of 1,5 ml microtubes containing liver samples taken from the freezer at a given time should not exceed the number that can be centrifuged at that time.
- (4) Arrange the 1,5 ml microtubes containing liver samples in the order of specimen number on the ice rack (no need to thaw the liver).

Operation of the pre-treatment

(1) Addition of the homogenate buffer

Check the list for the volume of the homogenate buffer to be used for a particular sample of liver and adjust the micropipette (volume range: 100-1 000 µl) to the appropriate volume. Attach a clean tip to the micropipette.

Take the homogenate buffer from the reagent bottle and add the buffer to the 1,5 ml microtube containing the liver.

Add the homogenate buffer to all of 1,5 ml microtubes containing the liver according to the procedure described above. There is no need to change the micropipette tip to a new one. However, if the tip is contaminated or suspected to be contaminated, the tip should be changed.

(2) Homogenisation of the liver

- Attach a new pestle for homogenisation to the microtube homogeniser.
- Insert the pestle into the 1,5 ml microtube. Hold the microtube homogeniser to press the liver between the surface of the pestle and the inner wall of the 1,5 ml microtube.
- Operate the microtube homogeniser for 10 to 20 seconds. Cool the 1,5 ml microtube with crushed ice during the operation.
- Lift up the pestle from the 1,5 ml microtube and leave it at rest for about 10 seconds. Then conduct a visual check of the state of the suspension.
- If pieces of liver are observed in the suspension, repeat the operations (3) and (4) to prepare satisfactory liver homogenate.
- Cool the suspended liver homogenate on the ice rack until centrifugation.
- Change the pestle to the new one for each homogenate.
- Homogenise all livers with homogenate buffer according to the procedure described above.

(3) Centrifugation of the suspended liver homogenate

- Confirm the temperature of the refrigerated centrifuge chamber at $\leq 5^{\circ}\text{C}$.
- Insert the 1,5 ml microtubes containing the suspended liver homogenate in refrigerated centrifuge (adjust the balance if necessary).
- Centrifuge the suspended liver homogenate at 13 000 g for 10 min at $\leq 5^{\circ}\text{C}$. However, if the supernatants are adequately separated, centrifugal force and time may be adjusted as needed.
- Following centrifugation, check that the supernatants are adequately separated (surface: lipid, intermediate: supernatant, bottom layer: liver tissue). If the separation is not adequate, centrifuge the suspension again under the same conditions.
- Remove all specimens from the refrigerated centrifuge and arrange them in the order of specimen number on the ice rack. Be careful not to resuspend each separated layer after the centrifugation.

(4) Collection of the supernatant

- Place four 0,5 ml microtubes for storage of the supernatant into the tube rack.
- Collect 30 µl of each supernatant (separated as the intermediate layer) with the micropipette and dispense it to one 0,5 ml microtube. Be careful not to collect the lipid on the surface or the liver tissue in the bottom layer.
- Collect the supernatant and dispense it to other two 0,5 ml microtubes in the same manner as described above.
- Collect the rest of the supernatant with the micropipette (if feasible: ≥ 100 µl). Then dispense the supernatant to the remaining 0,5 ml microtube. Be careful not to collect the lipid on the surface or the liver tissue in the bottom layer.
- Close the cap of the 0,5 ml microtube and write the volume of the supernatant on the label. Then immediately cool the microtubes on the ice rack.
- Change the tip of the micropipette to the new one for each supernatant. If a large amount of lipid becomes attached to the tip, change it to the new one immediately to avoid contamination of the liver extract with fat.
- Dispense all of the centrifuged supernatant to four 0,5 ml microtubes according to the procedure described above.
- After dispensing the supernatant to the 0,5 ml microtubes, place all of them in the tube rack with the identification label, and then freeze them in the freezer immediately. If the VTG concentrations are measured immediately after the pre-treatment, keep one 0,5 ml microtube (containing 30 µl of supernatant) cool in the tube rack and transfer it to the workstation where the ELISA assay is conducted. In such case, place the remaining microtubes in the tube racks and freeze them in the freezer.
- After the collection of the supernatant, discard the residue adequately.

Storage of the specimen

Store the 0,5 ml microtubes containing the supernatant of the liver homogenate at ≤ -70 °C until they are used for the ELISA.

Procedure 3A: Zebrafish, Blood Collection from the Caudal Vein / Artery

Immediately following anaesthesia, the caudal peduncle is severed transversely, and the blood is removed from the caudal artery/vein with a heparinised microhematocrit capillary tube. Blood volumes range from 5 to 15 microliters depending on fish size. An equal volume of aprotinin buffer (6 micrograms/ml in PBS) is added to the microcapillary tube, and plasma is separated from the blood via centrifugation (5 minutes at 600 g). Plasma is collected in the test tubes and stored at -20 °C until analysed for VTG or other proteins of interest.

Procedure 3B: Zebrafish, Blood Collection by Cardiac Puncture

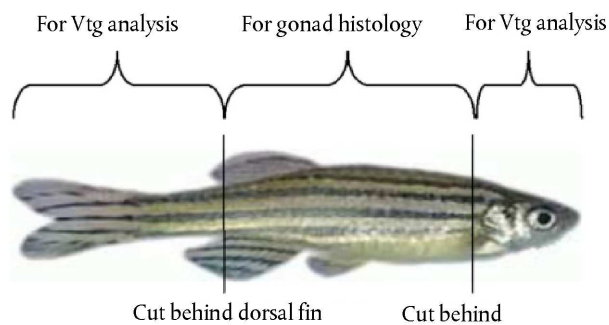
To avoid coagulation of blood and degradation of protein the samples are collected within Phosphate-buffered saline (PBS) buffer containing heparin (1 000 units/ml) and the protease inhibitor aprotinin (2 TIU/ml). As ingredients for the buffer, heparin, ammonium-salt and lyophilised aprotinin are recommended. For blood sampling, a syringe (1ml) with a fixed thin needle (e.g. Braun Omnikan-F) is recommended. The syringe should be prefilled with buffer (approximately 100 microliter) to completely elute the small blood volumes from each fish. The blood samples are taken by cardiac puncture. At first the fish should be anesthetized with MS-222 (100 mg/l). The proper plane of anaesthesia allows the user to distinguish the heartbeat of the zebrafish. While puncturing the heart, keep the syringe piston under weak tension. Collectable blood volumes range between 20 - 40 microliters. After cardiac puncture, the blood/buffer-mixture should be filled into the test tube. Plasma is separated from the blood via centrifugation (20 min; 5 000 g) and should be stored at -80 °C until required for analysis.

Procedure 3C: SOP: Zebrafish, homogenisation of head & tail

1. The fish are anaesthetised and euthanised in accordance with the test description.
2. The head and tail are cut of the fish in accordance with Figure 1.

Important: All dissection instruments, and the cutting board should be rinsed and cleaned properly (e.g. with 96 % ethanol) between handling of each single fish to prevent "vitellogenin pollution" from females or induced males to uninduced males.

Figure 1



3. The weight of the pooled head and tail from each fish is measured to the nearest mg.
4. After being weighed, the parts are placed in appropriate tubes (e.g. 1,5 ml eppendorf) and frozen at -80°C until homogenisation or directly homogenised on ice with two plastic pistils. (Other methods can be used if they are performed on ice and the result is a homogenous mass). Important: The tubes should be numbered properly so that the head and tail from the fish can be related to their respective body-section used for gonad histology.
5. When a homogenous mass is achieved, 4 x the tissue weight of ice-cold **homogenisation buffer** (*) is added. Keep working with the pistils until the mixture is homogeneous. Important note: New pistils are used for each fish.
6. The samples are placed on ice until centrifugation at 4°C at 50 000 g for 30 min.
7. Use a pipette to dispense portions of 20 μl supernatant into **at least two** tubes by dipping the tip of the pipette below the fat layer on the surface and carefully sucking up the supernatant without fat- or pellet fractions.
8. The tubes are stored at -80°C until use.

(*) Homogenisation buffer:

- (50 mM Tris-HCl pH 7,4; 1 % Protease inhibitor cocktail (Sigma)): 12 ml Tris-HCl pH 7,4 + 120 μl Protease inhibitor cocktail.
- TRIS: TRIS-ULTRA PURE (ICN) e.g. from Bie & Berntsen, Denmark.
- Protease inhibitor cocktail: From Sigma (for mammalian tissue) Product number P 8340.

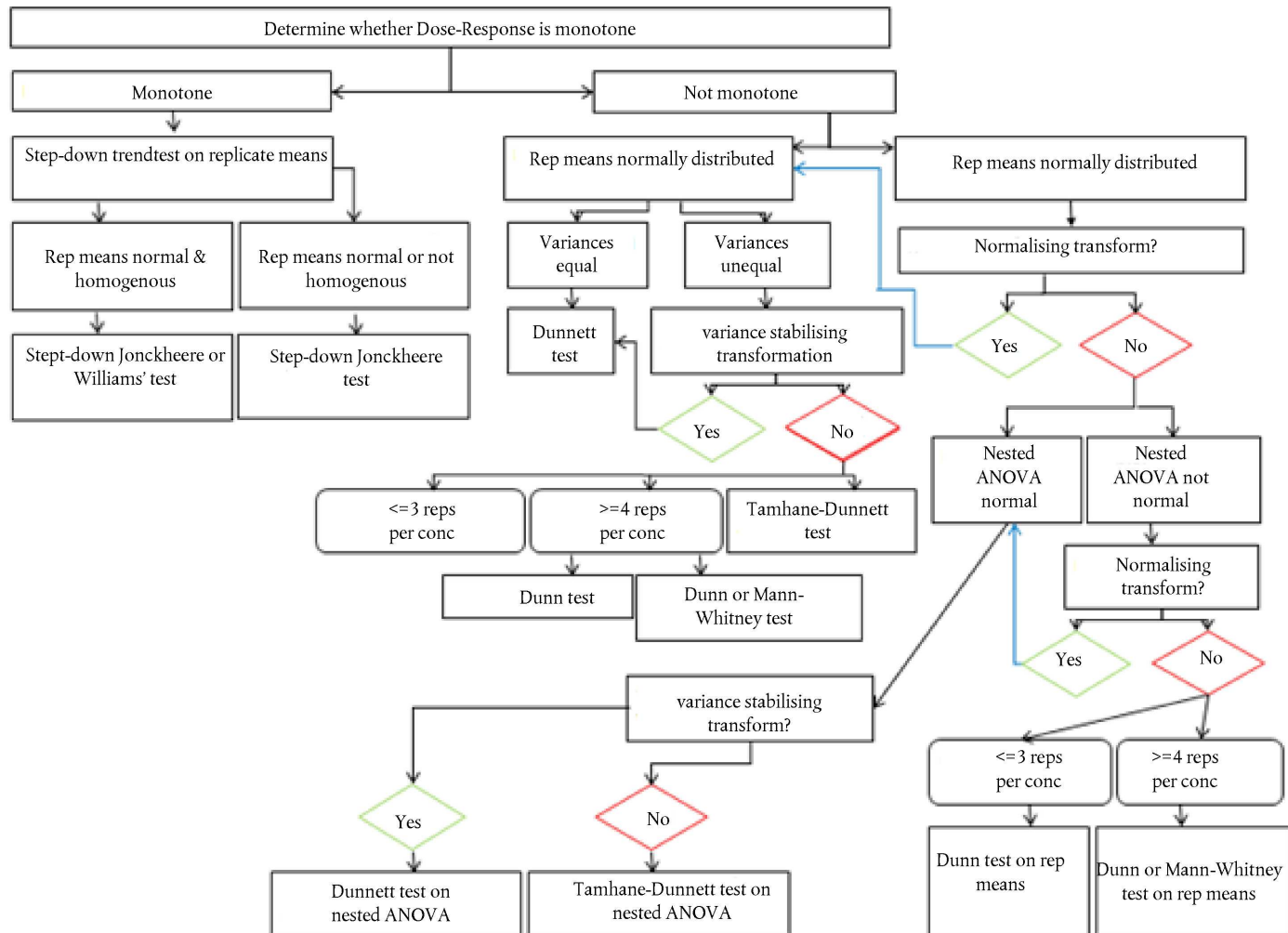
NOTE: The homogenisation buffer should be used the same day as manufactured. Place on ice during use.

*Appendix 7***VITELLOGENIN FORTIFICATION SAMPLES AND INTER-ASSAY REFERENCE STANDARD**

On each day that VTG assays are performed, a fortification sample made using an inter-assay reference standard will be analysed. The VTG used to make the inter-assay reference standard will be from a batch different from the one used to prepare calibration standards for the assay being performed.

The fortification sample will be made by adding a known quantity of the inter-assay standard to a sample of control male plasma. The sample will be fortified to achieve a VTG concentration between 10 and 100 times the expected vitellogenin concentration of control male fish. The sample of control male plasma that is fortified may be from an individual fish or may be a composite from several fish.

A subsample of the unfortified control male plasma will be analysed in at least two duplicate wells. The fortified sample also will be analysed in at least two duplicate wells. The mean quantity of vitellogenin in the two unfortified control male plasma samples will be added to the calculated quantity of VTG added to fortification the samples to determine an expected concentration. The ratio of this expected concentration to the measured concentration will be reported along with the results from each set of assays performed on that day.



C.49 Fish Embryo Acute Toxicity (FET) Test

INTRODUCTION

1. This test method (TM) is equivalent to OECD test guideline (TG) 236 (2013). It describes a Fish Embryo Acute Toxicity (FET) test with the zebrafish (*Danio rerio*). This test is designed to determine acute toxicity of chemicals on embryonic stages of fish. The FET-test is based on studies and validation activities performed on zebrafish (1)(2)(3)(4)(5)(6)(7)(8)(9)(10)(11)(12)(13)(14). The FET-test has been successfully applied to a wide range of chemicals exhibiting diverse modes of action, solubilities, volatilities, and hydrophobicities (reviewed in 15 and 16).
2. Definitions used in this test method are given in Appendix 1.

PRINCIPLE OF THE TEST

3. Newly fertilised zebrafish eggs are exposed to the test chemical for a period of 96 hrs. Every 24 hrs, up to four apical observations are recorded as indicators of lethality (6): (i) coagulation of fertilised eggs, (ii) lack of somite formation, (iii) lack of detachment of the tail-bud from the yolk sac, and (iv) lack of heartbeat. At the end of the exposure period, acute toxicity is determined based on a positive outcome in any of the four apical observations recorded, and the LC_{50} is calculated.

INITIAL CONSIDERATIONS

4. Useful information about substance-specific properties include the structural formula, molecular weight, purity, stability in water and light, pK_a and K_{ow} , water solubility and vapour pressure as well as results of a test for ready biodegradability (TM C.4 (17) or TM 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. A reliable analytical method for the quantification of the substance in the test solutions with known and reported accuracy and limit of detection should be available.
5. If the test method is used for the testing of a mixture, its composition should, as far as possible, be characterised, e.g. by the chemical identity of its constituents, their quantitative occurrence and their substance-specific properties (see paragraph 4). 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.
6. Concerning substances that may be activated via metabolism, there is evidence that zebrafish embryos do have biotransformation capacities (19)(20)(21)(22). However, the metabolic capacity of embryonic fish is not always similar to that of juvenile or adult fish. For instance, the protoxicant allyl alcohol (9) has been missed in the FET. Therefore, if there are any indications that metabolites or other transformation products of relevance may be more toxic than the parent compound, it is also recommended to perform the test with these metabolites/transformation products and to also use these results when concluding on the toxicity of the test chemical, or alternatively perform another test which takes metabolism into further account.
7. For substances with a molecular weight ≥ 3 kDa, a very bulky molecular structure, and substances causing delayed hatch which might preclude or reduce the post-hatch exposure, embryos are not expected to be sensitive because of limited bioavailability of the substance, and other toxicity tests might be more appropriate.

VALIDITY OF THE TEST

8. For the test results to be valid, the following criteria apply:
 - (a) The overall fertilisation rate of all eggs collected should be ≥ 70 % in the batch tested.

- (b) The water temperature should be maintained at 26 ± 1 °C in test chambers at any time during the test.
- (c) Overall survival of embryos in the negative (dilution-water) control, and, where relevant, in the solvent control should be ≥ 90 % until the end of the 96 hrs exposure.
- (d) Exposure to the positive control (e.g. 4,0 mg/l 3,4-dichloroaniline for zebrafish) should result in a minimum mortality of 30 % at the end of the 96 hrs exposure.
- (e) Hatching rate in the negative control (and solvent control if appropriate) should be ≥ 80 % at the end of 96 hrs exposure.
- (f) At the end of the 96 hrs exposure, the dissolved oxygen concentration in the negative control and highest test concentration should be ≥ 80 % of saturation.

DESCRIPTION OF THE METHOD

9. An overview of recommended maintenance and test conditions is available in Appendix 2.

Apparatus

10. The following equipment is needed:

- (a) Fish tanks made of chemically inert material (e.g. glass) and of a suitable capacity in relation to the recommended loading (see "Maintenance of brood fish", paragraph 14);
- (b) Inverted microscope and/or binocular with a capacity of at least 80-fold magnification. If the room used for recording observations cannot be adjusted to 26 ± 1 °C, a temperature-controlled cross movement stage or other methods to maintain temperature are necessary;
- (c) Test chambers; e.g., standard 24-well plates with a depth of approx. 20 mm. (see "Test chambers", paragraph 11);
- (d) e.g., self-adhesive foil to cover the 24-well plates;
- (e) Incubator or air-conditioned room with controlled temperature, allowing to maintain 26 ± 1 °C in wells (or test chambers);
- (f) pH-meter;
- (g) Oxygen meter;
- (h) Equipment for determination of hardness of water and conductivity;
- (i) Spawn trap: instrument trays of glass, stainless steel or other inert materials; wire mesh (grid size $2 \pm 0,5$ mm) of stainless steel or other inert material to protect the eggs once laid; spawning substrate (e.g. plant imitates of inert material) (TM C.48, Appendix 4a (23));
- (j) Pipettes with widened openings to collect eggs;
- (k) Glass vessels to prepare different test concentrations and dilution water (beakers, graduated flasks, graduated cylinders and graduated pipettes) or to collect zebrafish eggs (e.g. beakers, crystallisation dishes);
- (l) If alternative exposure systems, such as flow-through (24) or passive dosing (25) are used for the conduct of the test, appropriate facilities and equipment are needed.

Test chambers

11. Glass or polystyrene test chambers should be used (e.g. 24-well plates with a 2,5-5 ml filling capacity per well). In case adsorption to polystyrene is suspected (e.g., for non-polar, planar substances with high K_{ow}), inert materials (glass) should be used to reduce losses due to adsorption (26). Test chambers should be randomly positioned in the incubator.

Water and test conditions

12. Dilution of the maintenance water is recommended to achieve hardness levels typical of a wide variety of surface waters. Dilution water should be prepared from reconstituted water (27). The resulting degree of hardness should be equivalent to 100-300 mg/l $CaCO_3$ in order to prevent excessive precipitation of calcium carbonate. Other well-characterised surface or well water may be used. The reconstituted water may be adapted to maintenance water of low hardness by dilution with deionised water up to a ratio of 1:5 to a minimum hardness of 30-35 mg/l $CaCO_3$. The water is aerated to oxygen saturation prior to addition of the test chemical. Temperature should be kept at 26 ± 1 °C, in the wells, throughout the test. The pH should be in a range between pH 6,5 and 8,5, and not vary within this range by more than 1,5 units during the course of the test. If the pH is not expected to remain in this range, then pH adjustment should be done prior to initiating the test. The pH adjustment should be made in such a way that the stock solution concentration is not changed to any significant extent and that no chemical reaction or precipitation of the test chemical is caused. Use of hydrogen chloride (HCl) and sodium hydroxide (NaOH) to correct pH in the solutions containing the test chemical is recommended.

Test solutions

13. Test solutions of the selected concentrations can be prepared, e.g. by dilution of a stock solution. The stock solutions should preferably be prepared by simply mixing or agitating the test chemical in the dilution water by mechanical means (e.g. stirring and/or ultra-sonification). If the test chemical is difficult to dissolve in water, procedures described in the OECD Guidance Document No. 23 for handling difficult substances and mixtures should be followed (28). The use of solvents should be avoided, but may be required in some cases in order to produce a suitably concentrated stock solution. Where a solvent is used to assist in stock solution preparation, its final concentration should not exceed 100 µl/l and should be the same in all test vessels. When a solvent is used, an additional solvent control is required.

Maintenance of brood fish

14. A breeding stock of unexposed, wild-type zebrafish with well-documented fertilisation rate of eggs is used for egg production. Fish should be free of macroscopically discernible symptoms of infection and disease and should not have undergone any pharmaceutical (acute or prophylactic) treatment for 2 months before spawning. Breeding fish are maintained in aquaria with a recommended loading capacity of 1 l water per fish and a fixed 12 - 16 hour photoperiod (29)(30)(31)(32)(33). Optimal filtering rates should be adjusted; excess filtering rates causing heavy perturbation of the water should be avoided. For feeding conditions, see Appendix 2. Surplus feeding should be avoided, and water quality and cleanness of the aquaria should be monitored regularly and be reset to the initial state, if necessary.

Proficiency Testing

15. As a reference chemical, 3,4-dichloroaniline (used in the validation studies (1)(2)), should be tested in a full concentration-response range to check the sensitivity of the fish strain used, preferably twice a year. For any laboratory initially establishing this assay, the reference chemical should be used. A laboratory can use this chemical to demonstrate their technical competence in performing the assay prior to submitting data for regulatory purposes.

Egg production

16. Zebrafish eggs may be produced via spawning groups (in individual spawning tanks) or via mass spawning (in the maintenance tanks). In the case of spawning groups, males and females (e.g., at a ratio of 2:1) in a breeding group are placed in spawning tanks a few hours before the onset of darkness on the day prior to the test. Since spawning groups of zebrafish may occasionally fail to spawn, the parallel use of at least three spawning tanks is recommended. To avoid genetic bias, eggs are collected from a minimum of three breeding groups, mixed and randomly selected.
17. For the collection of eggs, spawn traps are placed into the spawning tanks or maintenance tanks before the onset of darkness on the day prior to the test or before the onset of light on the day of the test. To prevent predation of eggs by adult zebrafish, the spawn traps are covered with inert wire mesh of appropriate mesh size (approx. $2 \pm 0,5$ mm). If considered necessary, artificial plants made of inert material (e.g., plastic or glass) can be fixed to the mesh as spawning stimulus (3)(4)(5)(23)(35). Weathered plastic materials which do not leach (e.g., phthalates) should be used. Mating, spawning and fertilisation take place within 30 min after the onset of light and the spawn traps with the collected eggs can be carefully removed. Rinsing eggs with reconstituted water after collection from spawning traps is recommended.

Egg differentiation

18. At 26 °C, fertilised eggs undergo the first cleavage after about 15 min and the consecutive synchronous cleavages form 4, 8, 16 and 32 cell blastomers (see Appendix 3)(35). At these stages, fertilised eggs can be clearly identified by the development of a blastula.

PROCEDURE

Conditions of exposure

19. Twenty embryos per concentration (one embryo per well) are exposed to the test chemical. Exposure should be such that ± 20 % of the nominal chemical concentration are maintained throughout the test. If this is not possible in a static system, a manageable semi-static renewal interval should be applied (e.g. renewal every 24 hrs). In these cases exposure concentrations need to be verified as a minimum in the highest and lowest test concentrations at the beginning and the end of each exposure interval (see paragraph 36). If an exposure concentration of ± 20 % of the nominal concentrations cannot be maintained, all concentrations need to be measured at the beginning and the end of each exposure interval (see paragraph 36). Upon renewal, care should be taken that embryos remain covered by a small amount of old test solutions to avoid drying. The test design can be adapted to meet the testing requirements of specific substances (e.g., flow-through (24) or passive dosing systems (25) for easily degradable or highly adsorptive substances (29), or others for volatile substances (36)(37)). In any case, care should be taken to minimise any stress to the embryos. Test chambers should be conditioned at least for 24 hrs with the test solutions prior to test initiation. Test conditions are summarised in Appendix 2.

Test concentrations

20. Normally, five concentrations of the test chemical spaced by a constant factor not exceeding 2,2 are required to meet statistical requirements. Justification should be provided, if fewer than five concentrations are used. The highest concentration tested should preferably result in 100 % lethality, and the lowest concentration tested should preferably give no observable effect, as defined in paragraph 28. A range-finding test before the definitive test allows selection of the appropriate concentration range. The range-finding is typically performed using ten embryos per concentration. The following instructions refer to performing the test in 24-well plates. If different test chambers (e.g. small Petri dishes) are used or more concentrations are tested, instructions have to be adjusted accordingly.

21. Details and visual instructions for allocation of concentrations across 24-well plates are available in paragraph 27 and Appendix 4, Figure 1.

Controls

22. Dilution water controls are required both as negative control and as internal plate controls. If more than 1 dead embryo is observed in the internal plate control, the plate is rejected, thus reducing the number of concentrations used to derive the LC_{50} . If an entire plate is rejected the ability to evaluate and discern observed effects may become more difficult, especially if the rejected plate is the solvent control plate or a plate in which treated embryos are also affected. In the first case the test must be repeated. In the second one the loss of an entire treatment group(s) due to internal control mortality may limit the ability to evaluate effects and determine LC_{50} values.
23. A positive control at a fixed concentration of 4 mg/l 3,4-dichloroaniline is performed with each egg batch used for testing.
24. In case a solvent is used, an additional group of 20 embryos is exposed to the solvent on a separate 24-well plate, thus serving as a solvent control. To consider the test acceptable, the solvent should be demonstrated to have no significant effects on time to hatch, survival, nor produce any other adverse effects on the embryos (cf. paragraph 8c).

Start of exposure and duration of test

25. The test is initiated as soon as possible after fertilisation of the eggs and terminated after 96 hrs of exposure. The embryos should be immersed in the test solutions before cleavage of the blastodisc commences, or, at latest, by the 16 cell-stage. To start exposure with minimum delay, at least twice the number of eggs needed per treatment group are randomly selected and transferred into the respective concentrations and controls (e.g. in 100 ml crystallisation dishes; eggs should be fully covered) not later than 90 minutes post fertilisation.
26. Viable fertilised eggs should be separated from unfertilised eggs and be transferred to 24-well plates pre-conditioned for 24 hrs and refilled with 2 ml/well freshly prepared test solutions within 180 minutes post fertilisation. By means of stereomicroscopy (preferably ≥ 30 -fold magnification), fertilised eggs undergoing cleavage and showing no obvious irregularities during cleavage (e.g. asymmetry, vesicle formation) or injuries of the chorion are selected. For egg collection and separation, see Appendix 3, Fig. 1 and 3 and Appendix 4, Fig. 2.

Distribution of eggs over the 24-well plates

27. Eggs are distributed to well plates in the following numbers (see also Appendix 4, Fig. 1)
- 20 eggs on one plate for each test concentration;
 - 20 eggs as solvent control on one plate (if necessary);
 - 20 eggs as positive control on one plate;
 - 4 eggs in dilution water as internal plate control on each of the above plates;
 - 24 eggs in dilution water as negative control on one plate.

Observations

28. Apical observations performed on each tested embryo include: coagulation of embryos, lack of somite formation, non-detachment of the tail, and lack of heartbeat (Table 1). These observations are used for the determination of lethality: Any positive outcome in one of these observations means that the zebrafish embryo is dead. Additionally, hatching is recorded in treatment and control groups on a daily basis starting from 48 hrs. Observations are recorded every 24 hrs, until the end of the test.

Table 1

Apical observations of acute toxicity in zebrafish embryos 24-96 hrs post fertilisation

	Exposure times			
	24 hrs	48 hrs	72 hrs	96 hrs
Coagulated embryos	+	+	+	+
Lack of somite formation	+	+	+	+
Non-detachment of the tail	+	+	+	+
Lack of heartbeat		+	+	+

29. *Coagulation of the embryo*: Coagulated embryos are milky white and appear dark under the microscope (see Appendix 5, Fig. 1). The number of coagulated embryos is determined after 24, 48, 72 and 96 hrs.
30. *Lack of somite formation*: At 26 ± 1 °C, about 20 somites have formed after 24 hrs (see Appendix 5, Figure 2) in a normally developing zebrafish embryo. A normally developed embryo shows spontaneous movements (side-to-side contractions). Spontaneous movements indicate the formation of somites. The absence of somites is recorded after 24, 48, 72 and 96 hrs. Non-formation of somites after 24 hrs might be due to a general retardation of development. At latest after 48 hrs, the formation of somites should be developed. If not, the embryos are considered dead.
31. *Non-detachment of the tail*: In a normally developing zebrafish embryo, detachment of the tail (see Appendix 5, Figure 3) from the yolk is observed following posterior elongation of the embryonic body. Absence of tail detachment is recorded after 24, 48, 72 and 96 hrs.
32. *Lack of heartbeat*: In a normally developing zebrafish embryo at 26 ± 1 °C, the heartbeat is visible after 48 hrs (see Appendix 5, Figure 4). Particular care should be taken when recording this endpoint, since irregular (erratic) heartbeat should *not* be recorded as lethal. Moreover, visible heartbeat without circulation in aorta abdominalis is considered non-lethal. To record this endpoint, embryos showing no heartbeat should be observed under a minimum magnification of 80x for at least one minute. Absence of heartbeat is recorded after 48, 72 and 96 hrs.
33. Hatching rates of all treatment and control groups should be recorded from 48 hrs onwards and reported. Although hatching is not an endpoint used for the calculation of the LC_{50} , hatching ensures exposure of the embryo without a potential barrier function of the chorion, and as such may help data interpretation.
34. Detailed descriptions of the normal (35) and examples of abnormal development of zebrafish embryos are illustrated in Appendixes 3 and 5.

Analytical measurements

35. At the beginning and at the end of the test, pH, total hardness and conductivity in the control(s) and in the highest test chemical concentration are measured. In semi-static renewal systems the pH should be measured prior to and after water renewal. The dissolved oxygen concentration is measured at the end of the test in the negative controls and highest test concentration with viable embryos, where it should be in compliance with the test validity criteria (see paragraph 8f). If there is concern that the temperature varies across the 24-well plates, temperature is measured in three randomly selected vessels. Temperature should be recorded preferably continuously during the test or, as a minimum, daily.
36. In a static system, the concentration of the test chemical should be measured, as a minimum, in the highest and lowest test concentrations, but preferably in all treatments, at the beginning and end of the test. In semi-static (renewal) tests where the concentration of the test chemical is expected to remain within $\pm 20\%$ of the nominal values, it is recommended that, as a minimum, the highest and lowest test concentrations be analysed when freshly prepared and immediately prior to renewal. For tests where the concentration of the test chemical is not expected to remain within $\pm 20\%$ of nominal, all test concentrations must be analysed when freshly prepared and immediately prior to renewal. In case of insufficient volume for analysis, merging of test solutions, or use of surrogate chambers being of the same material and having the same volume to surface area ratios as 24-well plates, may be useful. It is strongly recommended that results be based on measured concentrations. When the concentrations do not remain within 80-120 % of the nominal concentration, the effect concentrations should be expressed relative to the geometric mean of the measured concentrations; see Chapter 5 in the OECD Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures for more details (28).

LIMIT TEST

37. Using the procedures described in this test method, a limit test may be performed at 100 mg/l of test chemical or at its limit of solubility in the test medium (whichever is the lower) in order to demonstrate that the LC_{50} is greater than this concentration. The limit test should be performed using 20 embryos in the treatment, the positive control and –if necessary- in the solvent control and 24 embryos in the negative control. If the percentage of lethality at the concentration tested exceeds the mortality in the negative control (or solvent control) by 10 %, a full study should be conducted. Any observed effects should be recorded. If mortality exceeds 10 % in the negative control (or solvent control), the test becomes invalid and should be repeated.

DATA AND REPORTING

Treatment of results

38. In this test, the individual wells are considered independent replicates for statistical analysis. The percentages of embryos for which at least one of the apical observations is positive at 48 and/or 96 hrs are plotted against test concentrations. For calculation of the slopes of the curve, LC_{50} values and the confidence limits (95 %), appropriate statistical methods should be applied (38) and the OECD Guidance Document on Current Approaches in the Statistical Analysis of Ecotoxicity Data should be consulted (39).

Test report

39. The test report should include the following information:

Test chemical:

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 physicochemical properties of the constituents.

Test organisms:

- scientific name, strain, source and method of collection of the fertilised eggs and subsequent handling.

Test conditions:

- test procedure used (e.g., semi-static renewal);
- photoperiod;
- test design (e.g., number of test chambers, types of controls);
- water quality characteristics in fish maintenance (e.g. pH, hardness, temperature, conductivity, dissolved oxygen);
- dissolved oxygen concentration, pH, total hardness, temperature and conductivity of the test solutions at the start and after 96 hrs;
- method of preparation of stock solutions and test solutions as well as frequency of renewal;
- justification for use of solvent and justification for choice of solvent, if other than water;
- the nominal test concentrations and the result of all analyses to determine the concentration of the test chemical in the test vessels; the recovery efficiency of the method and the limit of quantification (LoQ) should also be reported;
- evidence that controls met the overall survival validity criteria;
- fertilisation rate of the eggs;
- hatching rate in treatment and control groups.

Results:

- maximum concentration causing no mortality within the duration of the test;
- minimum concentration causing 100 % mortality within the duration of the test;
- cumulative mortality for each concentration at the recommended observation times;
- the LC₅₀ values at 96 hrs (and optionally at 48 hrs) for mortality with 95 % confidence limits, if possible;
- graph of the concentration-mortality curve at the end of the test;
- mortality in the controls (negative controls, internal plate controls, as well as positive control and any solvent control used);
- data on the outcome of each of the four apical observations;
- incidence and description of morphological and physiological abnormalities, if any (see examples provided in Appendix 5, Figure 2);
- incidents in the course of the test which might have influenced the results;

- statistical analysis and treatment of data (probit analysis, logistic regression model and geometric mean for LC_{50});
- slope and confidence limits of the regression of the (transformed) concentration-response curve.

Any deviation from the test method and relevant explanations.

Discussion and interpretation of results.

LITERATURE

- (1) OECD (2011) Validation Report (Phase 1) for the Zebrafish Embryo Toxicity Test: Part I and Part II. Series on Testing and Assessment No. 157, OECD, Paris.
- (2) OECD (2012) Validation Report (Phase 2) for the Zebrafish Embryo Toxicity Test: Part I and Part II (Annexes). Series on Testing and Assessment No. 179, OECD, Paris.
- (3) Braunbeck, T., Böttcher, M., Hollert, H., Kosmehl, T., Lammer, E., Leist, E., Rudolf, M. and Seitz, N. (2005) Towards an alternative for the acute fish LC_{50} test in chemical assessment: The fish embryo toxicity test goes multi-species-an update. ALTEX 22: 87-102.
- (4) ISO (2007) International Standard Water quality — Determination of the acute toxicity of waste water to zebrafish eggs (*Danio rerio*). ISO 15088:2007(E) International Organization for Standardization.
- (5) Nagel, R. (2002) DarT: The embryo test with the zebrafish (*Danio rerio*) — a general model in ecotoxicology and toxicology. ALTEX 19: 38-48.
- (6) Schulte, C. and Nagel, R. (1994) Testing acute toxicity in embryo of zebrafish, *Brachydanio rerio* as alternative to the acute fish test — preliminary results. ATLA 22, 12-19.
- (7) Bachmann, J. (2002) Development and validation of a teratogenicity screening test with embryos of the zebrafish (*Danio rerio*). PhD-thesis, Technical University of Dresden, Germany.
- (8) Lange, M., Gebauer, W., Markl, J. and Nagel, R. (1995) Comparison of testing acute toxicity on embryo of zebrafish (*Brachydanio rerio*), and RTG-2 cytotoxicity as possible alternatives to the acute fish test. Chemosphere 30/11: 2087-2102.
- (9) Knöbel, M., Busser, F.J.M., Rico-Rico, A., Kramer, N.I., Hermens, J.L.M., Hafner, C., Tanneberger, K., Schirmer, K., Scholz, S. (2012). Predicting adult fish acute lethality with the zebrafish embryo: relevance of test duration, endpoints, compound properties, and exposure concentration analysis. Environ. Sci. Technol. 46, 9690-9700.
- (10) Kammann, U., Vobach, M. and Wosniok, W. (2006) Toxic effects of brominated indoles and phenols on zebrafish embryos. Arch. Environ. Contam. Toxicol., 51: 97-102.
- (11) Groth, G., Kronauer, K. and Freundt, K.J. (1994) Effects of *N,N*-dimethylformamide and its degradation products in zebrafish embryos. Toxicol. In Vitro 8: 401-406.
- (12) Groth, G., Schreeb, K., Herdt, V. and Freundt, K.J. (1993) Toxicity studies in fertilized zebrafish fish eggs treated with *N*-methylamine, *N,N*-dimethylamine, 2-aminoethanol, isopropylamine, aniline, *N*-methylaniline, *N,N*-dimethylaniline, quinone, chloroacetaldehyde, or cyclohexanol. Bull. Environ. Contam. Toxicol. 50: 878-882.
- (13) Nguyen, L.T. and Janssen, C.R. (2001) Comparative sensitivity of embryo-larval toxicity assays with African catfish (*Clarias gariepinus*) and zebra fish (*Danio rerio*). Environ. Toxicol. 16: 566-571.
- (14) Cheng, S.H., Wai, A.W.K., So, C.H. and Wu, R.S.S. (2000) Cellular and molecular basis of cadmium-induced deformities in zebrafish embryos. Environ. Toxicol. Chem. 19: 3024-3031.
- (15) Belanger, S. E., Rawlings J. M. and Carr G. J. (2013). Use of Fish Embryo Toxicity Tests for the Prediction of Acute Fish Toxicity to Chemicals. Environmental Toxicology and Chemistry 32: 1768-1783. .

- (16) Lammer, E., Carr, G. J., Wendler, K., Rawlings, J. M., Belanger, S. E., Braunbeck, T. (2009) Is the fish embryo toxicity test (FET) with the zebrafish (*Danio rerio*) a potential alternative for the fish acute toxicity test? Comp. Biochem. Physiol. C Toxicol. Pharmacol.: 149 (2), 196-209
- (17) Chapter C.4 of this Annex: Ready Biodegradability.
- (18) Chapter C.29 of this Annex: Ready Biodegradability, CO₂ in sealed vessels.
- (19) Weigt, S., Huebler, N., Strecker, R., Braunbeck, T., Broschard, T.H. (2011) Zebrafish (*Danio rerio*) embryos as a model for testing proteratogens. Toxicology 281: 25-36.
- (20) Weigt, S., Huebler, N., Strecker, R., Braunbeck, T., Broschard, T.H. (2012) Developmental effects of coumarin and the anticoagulant coumarin derivative warfarin on zebrafish (*Danio rerio*) embryos. Reprod. Toxicol. 33: 133-141.
- (21) Incardona, J.P., Linbo, T.L., Scholz, N.L. (2011) Cardiac toxicity of 5-ring polycyclic aromatic hydrocarbons is differentially dependent on the aryl hydrocarbon receptor 2 isoform during zebrafish development. Toxicol. Appl. Pharmacol. 257: 242-249.
- (22) Kubota, A., Stegeman, J.J., Woodin, B.R., Iwanaga, T., Harano, R., Peterson, R.E., Hiraga, T., Teraoka, H. (2011) Role of zebrafish cytochrome P450 CYP1C genes in the reduced mesencephalic vein blood flow caused by activation of AHR2. Toxicol. Appl. Pharmacol. 253: 244-252.
- (23) Chapter C.48 of this Annex: Fish Short Term Reproduction Assay.. See Appendix 4a.
- (24) Lammer, E., Kamp, H.G., Hisgen, V., Koch, M., Reinhard, D., Salinas, E.R., Wendler, K., Zok, S., Braunbeck, T. (2009) Development of a flow-through system for the fish embryo toxicity test (FET) with zebrafish (*Danio rerio*). Toxicol. in vitro 23: 1436-1442.
- (25) Brown, R.S., Akhtar, P., Åkerman, J., Hampel, L., Kozin, I.S., Villerius, L.A., Klammer, H.J.C., (2001) Partition controlled delivery of hydrophobic substances in toxicity tests using poly(dimethylsiloxane) (PDMS) films. Environ. Sci. Technol. 35, 4097-4102.
- (26) Schreiber, R., Altenburger, R., Paschke, A., Küster, E. (2008) How to deal with lipophilic and volatile organic substances in microtiter plate assays. Environ. Toxicol. Chem. 27, 1676-1682.
- (27) ISO (1996) International Standards. Water quality — Determination of the acute lethal toxicity of substances to a freshwater fish [*Brachydanio rerio* Hamilton-Buchanan (Teleostei, Cyprinidae)]. ISO 7346-3: Flow-through method. Available: [<http://www.iso.org>].
- (28) OECD (2000) Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. Series on Testing and Assessment No. 23, OECD, Paris.
- (29) Laale, H.W. (1977) The biology and use of zebrafish, *Brachydanio rerio*, in fisheries research. A literature review. J. Fish Biol. 10: 121-173.
- (30) Westerfield, M. (2007) The zebrafish book: A guide for the laboratory use of zebrafish (*Brachydanio rerio*). 5th edition. Eugene, University of Oregon Press, Institute of Neuroscience, USA.
- (31) Canadian Council on Animal Care (2005) Guidelines on: the Care and Use of Fish in Research, Teaching and Testing, ISBN: 0-919087-43-4 <http://www.ccac.ca/Documents/Standards/Guidelines/Fish.pdf>
- (32) European Commission (2007) Commission recommendation 2007/526/EC of 18 June 2007 on guidelines for the accommodation and care of animals used for experimental and other scientific purposes (notified under document number C(2007) 2525) [<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:197:0001:0089:EN:PDF>]
- (33) European Union (2010) Directive 2010/63/EU of the European Parliament and Council of 22 September 2010 on the protection of animals used for scientific purposes (OJ L 276, 20.10.2010, p. 33).

- (34) Nagel, R. (1986) Untersuchungen zur Eiproduktion beim Zebrabärbling (*Brachydanio rerio*, Ham.-Buch.). J. Appl. Ichthyol. 2: 173-181.
 - (35) Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. and Schilling, T.F. (1995) Stages of embryonic development of the zebrafish. Dev. Dyn. 203: 253-310.
 - (36) Chapter C.2 of this Annex: *Daphnia* sp., Acute Immobilisation Test.
 - (37) Weil, M., Scholz, S., Zimmer, M., Sacher, F., Duis, K. (2009) Gene expression analysis in zebrafish embryos: a potential approach to predict effect concentrations in the fish early life stage test. Environ. Toxicol. Chem. 28: 1970-1978
 - (38) ISO (2006) International Standard. Water quality — Guidance on statistical interpretation of ecotoxicity data ISO TS 20281. Available: [<http://www.iso.org>].
 - (39) OECD (2006) Guidance Document on Current Approaches in the Statistical Analysis of Ecotoxicity Data: a Guidance to Application. Series on Testing and Assessment No. 54. OECD, Paris.
 - (40) Braunbeck, T., Lammer, E., 2006. Detailed review paper "Fish embryo toxicity assays". UBA report under contract no. 20385422 German Federal Environment Agency, Berlin. 298 pp.
-

Appendix 1

DEFINITIONS

Apical endpoint: Causing effect at population level.

Blastula: A cellular formation around the animal pole that covers a certain part of the yolk.

Chemical: A substance or a mixture

Epiboly: is a massive proliferation of predominantly epidermal cells in the gastrulation phase of the embryo and their movement from the dorsal to the ventral side, by which entodermal cell layers are internalised in an invagination-like process and the yolk is incorporated into the embryo.

Flow-through test: A test with continued flow of test solutions through the test system during the duration of exposure.

Internal Plate Control: Internal control consisting of 4 wells filled with dilution water per 24-well plate to identify potential contamination of the plates by the manufacturer or by the researcher during the procedure, and any plate effect possibly influencing the outcome of the test (e.g. temperature gradient).

IUPAC: International Union of Pure and Applied Chemistry

Maintenance water: Water in which the husbandry of the adult fish is performed.

Median Lethal Concentration (LC₅₀): The concentration of a test chemical that is estimated to be lethal to 50 % of the test organisms within the test duration.

Semi-static renewal test: A test with regular renewal of the test solutions after defined periods (e.g., every 24 hrs).

SMILES: Simplified Molecular Input Line Entry Specification

Somite: In the developing vertebrate embryo, somites are masses of mesoderm distributed laterally to the neural tube, which will eventually develop dermis (dermatome), skeletal muscle (myotome), and vertebrae (sclerotome).

Static test: A test in which test solutions remain unchanged throughout the duration of the test.

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

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

Appendix 2

MAINTENANCE, BREEDING AND TYPICAL CONDITIONS FOR ZEBRAFISH EMBRYO ACUTE TOXICITY TESTS

Zebrafish (<i>Danio rerio</i>)		
Origin of species		India, Burma, Malakka, Sumatra
Sexual dimorphism		Females: protruding belly, when carrying eggs Males: more slender, orange tint between blue longitudinal stripes (particularly evident at the anal fin)
Feeding regime		Dry flake food (max. 3 % fish weight per day) 3 - 5 times daily; additionally brine shrimp (<i>Artemia spec.</i>) nauplii and/or small daphnids of appropriate size obtained from an uncontaminated source. Feeding live food provides a source of environmental enrichment and therefore live food should be given wherever possible. To guarantee for optimal water quality, excess food and faeces should be removed approx. one hour after feeding.
Approximate weight of adult fish		Females: $0,65 \pm 0,13$ g Males: $0,5 \pm 0,1$ g
Maintenance of parental fish	Illumination	Fluorescent bulbs (wide spectrum); $10-20 \mu\text{E}/\text{m}^2/\text{s}$, 540-1 080 lux, or 50-100 ft-c (ambient laboratory levels); 12-16 hrs photoperiod
	Water temperature	26 ± 1 °C
	Water quality	$\text{O}_2 \geq 80$ % saturation, hardness: e.g. $\sim 30-300$ mg/l CaCO_3 , $\text{NO}_3^- \leq 48$ mg/l, NH_4^+ and $\text{NO}_2^- < 0,001$ mg/l, residual chlorine $< 10 \mu\text{g}/\text{l}$, total organic chlorine < 25 ng/l, pH = 6,5 – 8,5
	Further water quality criteria	Particulate matter < 20 mg/l, total organic carbon < 2 mg/l, total organophosphorus pesticides < 50 ng/l, total organochlorine pesticides plus polychlorinated biphenyls < 50 ng/l
	Tank size for maintenance	e.g. 180 l, 1 fish/l
	Water purification	Permanent (charcoal filtered); other possibilities include combinations with semi-static renewal maintenance or flow-through system with continuous water renewal
	Recommended male to female ratio for breeding	2:1 (or mass spawning)
Spawning tanks		e.g. 4 l tanks equipped with steel grid bottom and plant dummy as spawning stimulant; external heating mats, or mass spawning within the maintenance tanks
Egg structure and appearance		Stable chorion (i.e. highly transparent, non-sticky, diameter $\sim 0,8-1,5$ mm)
Spawning rate		A single mature female spawns at least 50-80 eggs per day. Depending on the strain, spawning rates may be considerably higher. The fertilisation rate should be ≥ 70 %. For first time spawning fish, fertilisation rates of the eggs may be lower in the first few spawns.
Test type		Static, semi-static renewal, flow-through, 26 ± 1 °C, 24 hrs conditioned test chambers (e.g. 24-well plates 2,5-5 ml per cavity)

Appendix 3

NORMAL ZEBRAFISH DEVELOPMENT AT 26 °C

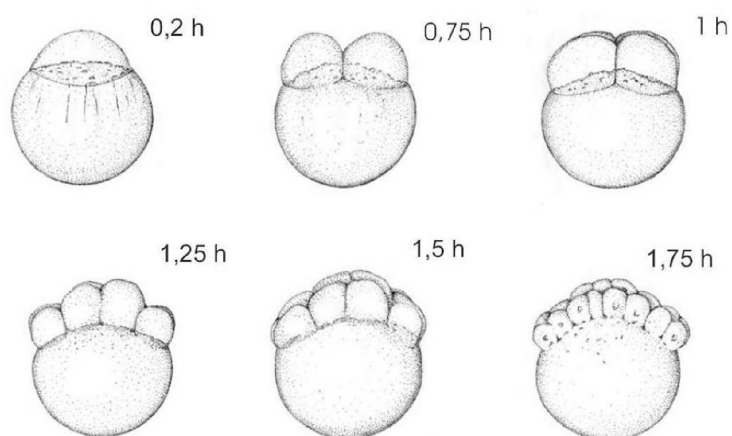


Fig. 1: Selected stages of early zebrafish (*Danio rerio*) development: 0,2 – 1,75 hrs post-fertilisation (from Kimmel *et al.*, 1995 (35)). The time sequence of normal development may be taken to diagnose both fertilisation and viability of eggs (see paragraph 26: Selection of fertilised eggs).

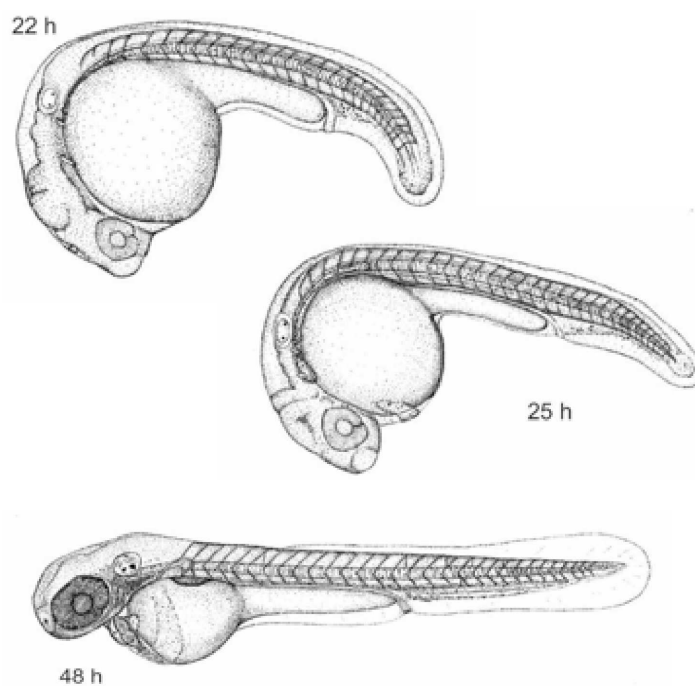


Fig. 2: Selected stages of late zebrafish (*Danio rerio*) development (de-chorionated embryo to optimise visibility): 22 – 48 hrs after fertilisation (from Kimmel *et al.*, 1995 (35)).

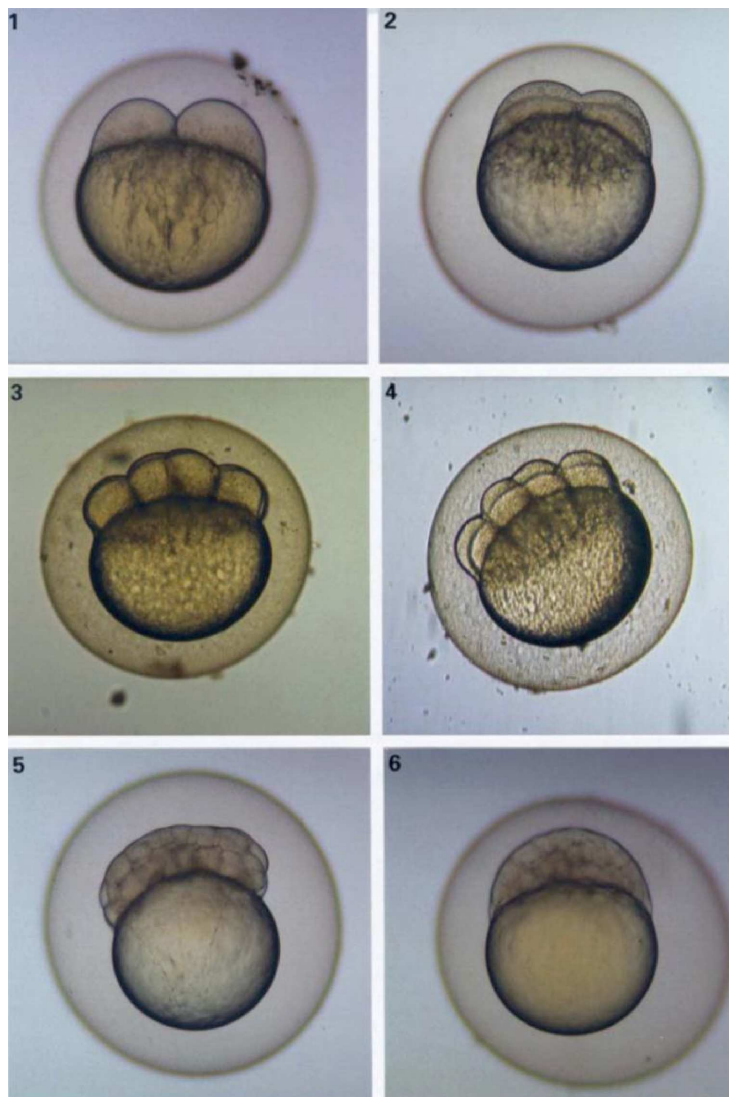
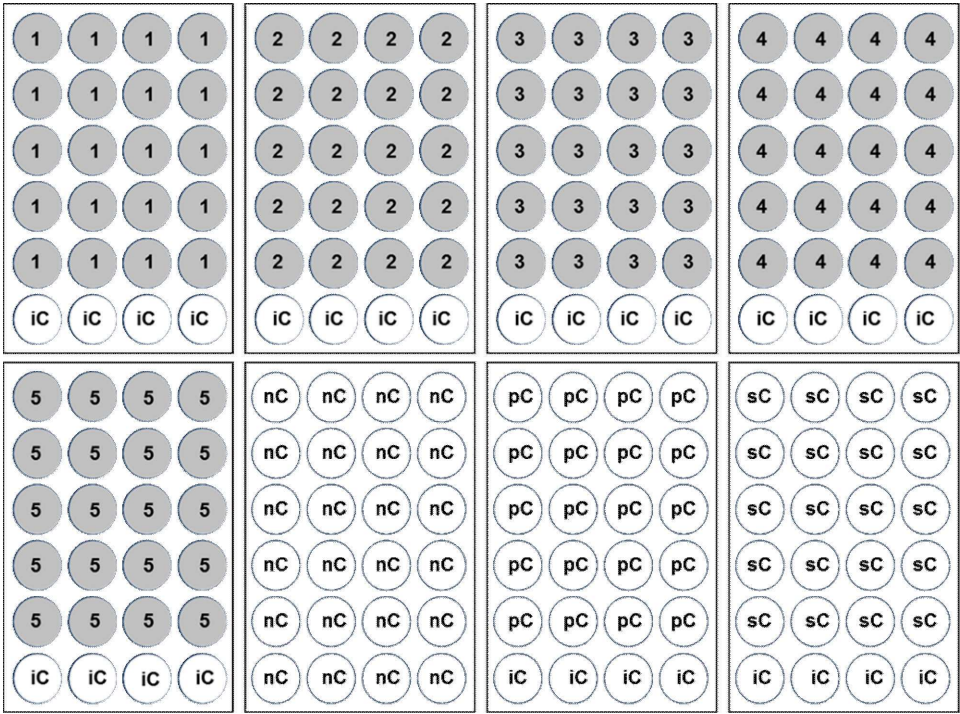


Fig. 3: Normal development of zebrafish (*Danio rerio*) embryos: (1) 0,75 hrs, 2-cell stage; (2) 1 hr, 4-cell stage; (3) 1,2 hrs, 8-cell stage; (4) 1,5 hrs, 16-cell stage; (5) 4,7 hrs, beginning epiboly; (6) 5,3 hrs, approx. 50 % epiboly (from Braunbeck & Lammer 2006 (40)).

Appendix 4

Figure 1

Layout of 24-well plates



1-5 = five test concentrations/chemical;

nC = negative control (dilution water);

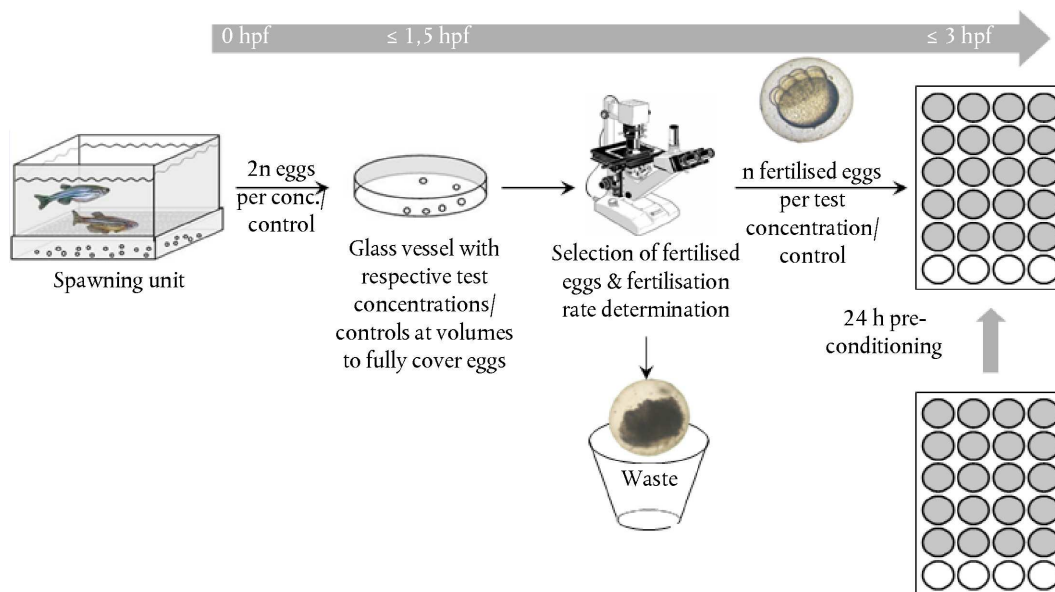
iC = internal plate control (dilution water);

pC = positive control (3,4-DCA 4 mg/l);

sC = solvent control.

Figure 2

Scheme of the zebrafish embryo acute toxicity test procedure (from left to right): production of eggs, collection of the eggs, pre-exposure immediately after fertilisation in glass vessels, selection of fertilised eggs with an inverted microscope or binocular and distribution of fertilised eggs into 24-well plates prepared with the respective test concentrations/controls, n = number of eggs required per test concentration/control (here 20), hpf = hours post-fertilisation.



Appendix 5

ATLAS OF LETHAL ENDPOINTS FOR THE ZEBRAFISH EMBRYO ACUTE TOXICITY TEST

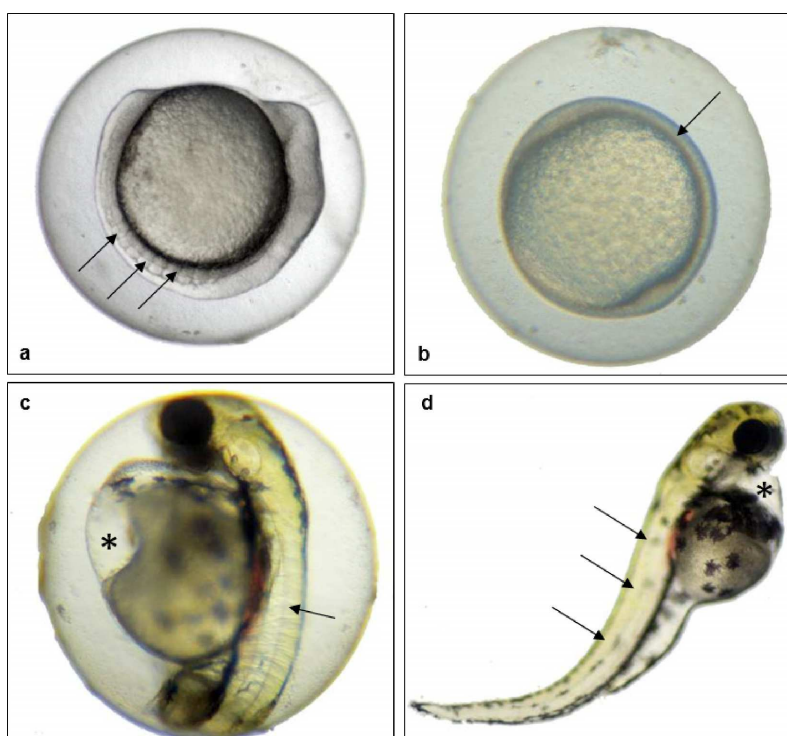
The following apical endpoints indicate acute toxicity and, consequently, death of the embryos: *coagulation of the embryo*, *non-detachment of the tail*, *lack of somite formation* and *lack of heartbeat*. The following micrographs have been selected to illustrate these endpoints.

Figure 1

Coagulation of the embryo:

Under bright field illumination, coagulated zebrafish embryos show a variety of intransparent inclusions.

Figure 2

Lack of somite formation:

Although retarded in development by approx. 10 hrs, the 24 hrs old zebrafish embryo in (a) shows well-developed somites (\rightarrow), whereas the embryo in (b) does not show any sign of somite formation (\rightarrow). Although showing a pronounced yolk sac oedema (*), the 48 hrs old zebrafish embryo in (c) shows distinct formation of somites (\rightarrow), whereas the 96 hrs old zebrafish embryo depicted in (d) does not show any sign of somite formation (\rightarrow). Note also the spinal curvature (scoliosis) and the pericardial oedema (*) in the embryo shown in (d).

Figure 3

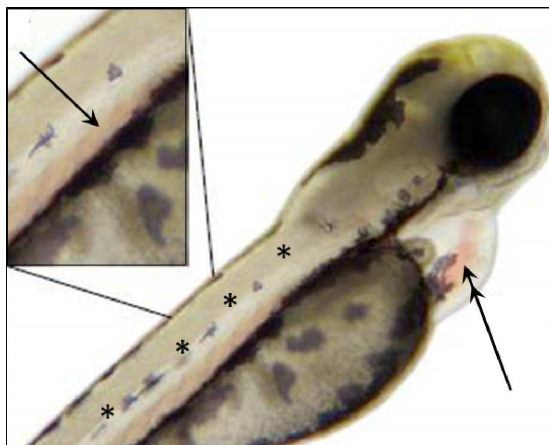
Non-detachment of the tail



Bud in lateral view (a: \rightarrow ; 96 hrs old zebrafish embryo). Note also the lack of the eye bud (*).

Figure 4

Lack of heartbeat



Lack of heartbeat is, by definition, difficult to illustrate in a micrograph. Lack of heartbeat is indicated by non-convulsion of the heart (double arrow). Immobility of blood cells in, e.g. the aorta abdominalis (\rightarrow in insert) is not an indicator for lack of heartbeat. Note also the lack of somite formation in this embryo (*, homogenous rather than segmental appearance of muscular tissues). The observation time to record an absence of heartbeat should be at least of one minute with a minimum magnification of 80 \times .

C.50 Sediment-free *Myriophyllum spicatum* toxicity test

INTRODUCTION

1. This test method is equivalent to OECD test guideline 238 (2014). It is designed to assess the toxicity of chemicals to *Myriophyllum spicatum*, a submersed aquatic dicotyledon, a species of the water milfoils family. It is based on an ASTM existing test method (1) modified as a sediment-free test system (2) to estimate the intrinsic ecotoxicity of test chemicals (independent of the distribution-behaviour of the test chemical between water and sediment). A test system without sediment has a low analytical complexity (only in the water phase) and the results can be analysed in parallel and/or comparison with those obtained in *Lemna sp.* test (3); in addition, the required sterile conditions allow to keep the effects of microorganisms and algae (chemical uptake/ degradation, etc.) as low as possible. This test does not replace other aquatic toxicity tests; it should rather complement them so that a more complete aquatic plant hazard and risk assessment is possible. The test method has been validated by a ring-test (4).
2. Details of testing with renewal (semi-static) and without renewal (static) of the test solution are described. Depending on the objectives of the test and the regulatory requirements, the use of semi-static method is recommended, e.g. for substances that are rapidly lost from solution as a result of volatilisation, adsorption, photodegradation, hydrolysis, precipitation or biodegradation. Further guidance is given in (5). This test method applies to substances, for which the test method has been validated, (see details in the ring-test report (4)) or to formulations, or known mixtures; if a mixture is tested, its constituents should be as far as possible identified and quantified. The sediment-free *Myriophyllum spicatum* test method complements the water-sediment *Myriophyllum spicatum* Toxicity Test (6). Before use of the test method for the testing of a mixture intended for a 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

3. Continuously growing plant cultures of *Myriophyllum spicatum* (only in modified Andrews' medium, see Appendix 2) are allowed to grow as monocultures in different concentrations of the test chemical over a period of 14 days in a sediment-free test system. The objective of the test is to quantify chemical-related effects on vegetative growth over this period based on assessments of selected measurement variables. Growth of shoot length, of lateral branches and roots as well as development of fresh and dry weight and increase of whorls are the measurement variables. In addition, account is taken of distinctive qualitative changes in test organisms, such as disfigurement or chlorosis and necrosis indicated by yellowing or white and brown colouring. To quantify chemical-related effects, growth in the test solutions is compared with that of the controls and the concentration bringing about a specified x % inhibition of growth is determined and expressed as the EC_x; "x" can be any value depending on the regulatory requirements, e.g. EC₁₀, EC₂₀, EC₅₀. It should be noted that estimates of EC₁₀ and EC₂₀ values are only reliable and appropriate in tests where coefficients of variation in control plants fall below the effect level being estimated, i.e. coefficients of variation should be < 20 % for robust estimation of an EC₂₀.
4. Both average specific growth rate (estimated from assessments of main shoot length and three additional measurement variables) and yield (estimated from the increase in main shoot length and three additional measurement variables) of untreated and treated plants should be determined. Specific growth rate (r) and yield (y) are subsequently used to determine the E_rC_x (e.g. E_rC₁₀, E_rC₂₀, E_rC₅₀) and E_yC_x (e.g. E_yC₁₀, E_yC₂₀, E_yC₅₀), respectively.
5. In addition, the lowest observed effect concentration (LOEC) and the no observed effect concentration (NOEC) may be statistically determined.

INFORMATION ON THE TEST CHEMICAL

6. An analytical method, with adequate sensitivity for quantification of the test chemical in the test medium, should be available. Information on the test chemical which may be useful in establishing the test conditions includes the structural formula, purity and impurities, water solubility, stability in water and light, acid dissociation constant (pK_a), partition coefficient octanol-water (K_{ow}), vapour pressure and biodegradability. Water solubility and vapour pressure can be used to calculate Henry's Law constant, which will indicate if significant losses of the test chemical during the test period are likely. This will help indicate whether particular steps to control such losses should be taken. Where information on the solubility and stability of the test chemical are uncertain, it is recommended that these be assessed under the conditions of the test, i.e. growth medium, temperature, lighting regime to be used in the test.
7. The pH control of the test medium is particularly important, e.g. when testing metals or substances which are hydrolytically unstable. Further guidance for testing chemicals with physical-chemical properties that make them difficult to test is provided in a OECD Guidance Document (5).

VALIDITY OF THE TEST

8. For the test to be valid, the doubling time of main shoot length in the control must be less than 14 days. Using the media and test conditions described in this test method, this criterion can be attained using a static or semi-static test regime.
9. The mean coefficient of variation for yield based on measurements of shoot fresh weight (i.e. from test initiation to test termination) and the additional measurement variables (see paragraph 37) in the control cultures do not exceed 35 % between replicates.
10. More than 50 % of the replicates of the control group are kept sterile over the exposure period of 14 days, which means visibly free of contamination by other organisms such as algae, fungi and bacteria (clear solution). *Note:* Guidance on how to assess sterility is provided in the ring-test report (4).

REFERENCE CHEMICAL

11. Reference chemical(s), such as 3,5-dichlorophenol used in the ring test (4), may be tested as a mean of checking the test procedure; from the ring test data, the mean EC_{50} -values of 3,5-DCP for the different response variables (see paragraphs 37-41 of this test method) are between 3,2 mg/l and 6,9 mg/l (see ring test report for details about confidence interval for these values). It is advisable to test a reference chemical at least twice a year or, where testing is carried out at a lower frequency, in parallel to the determination of the toxicity of a test chemical.

DESCRIPTION OF THE METHOD

Apparatus

12. All equipment in contact with the test media should be made of glass or other chemically inert material. Glassware used for culturing and testing purposes should be cleaned of chemical contaminants that might leach into the test medium and should be sterile. The test vessels should be long enough for the shoot in the control vessels to grow in the water phase without reaching the surface of the test medium at the end of the test. Thick-walled borosilicate glass test tubes without lip, inner diameter approximately 20 mm, length approximately 250 mm, with aluminium caps are recommended.
13. Since the modified Andrews' medium contains sucrose (which stimulates the growth of fungi and bacteria), the test solutions have to be prepared under sterile conditions. All liquids as well as equipment are sterilised before use. Sterilisation is carried out via heated air treatment (210 °C) for 4 hours or autoclaving for 20 minutes at 121 °C. In addition, all flasks, dishes, bowls etc. and other equipment undergo flame treatment at a sterile workbench just prior to use.

14. The cultures and test vessels should not be kept together. This is best achieved using separate environmental growth chambers, incubators, or rooms. Illumination and temperature should be controllable and maintained at a constant level.

Test organism

15. *Myriophyllum spicatum* — a submersed aquatic dicotyledon — is a species of the water milfoils family. Between June and August, inconspicuous pink-white flowers protrude above the water surface. The plants are rooted in the ground by a system of robust rhizomes and can be found in the entire northern hemisphere in eutrophic, however non-polluted and more calciferous still waters with muddy substrate. *Myriophyllum spicatum* prefers fresh water, but is found in brackish water as well.
16. For the sediment-free toxicity test, sterile plants are required. If the testing laboratory does not have regular cultures of *Myriophyllum spicatum*, sterile plant material may be obtained from another laboratory or (unsterile) plant material might be taken from the field or provided by a commercial supplier; if plants come from the field a taxonomic verification of the species should be envisaged. If collected from the field or provided by a commercial supplier, plants should be sterilised (1) and maintained in culture in the same medium as used for testing for a minimum of eight weeks prior to use. Field sites used for collecting starting cultures have to be free of obvious sources of contamination. Great care should be taken to ensure that the correct species is obtained when collecting *Myriophyllum spicatum* from the field, especially in regions where it can hybridise with other *Myriophyllum* species. If obtained from another laboratory they should be similarly maintained for a minimum of three weeks. The source of plant material and the species used for testing should always be reported.
17. The quality and uniformity of the plants used for the test will have a significant influence on the outcome of the test and should therefore be selected with care. Young, rapidly growing plants without visible lesions or discoloration (chlorosis) should be used. Details about preparation of the test organism are given in Appendix 4.

Cultivation

18. To reduce the frequency of culture maintenance (e.g. when no *Myriophyllum* tests are planned for a period), cultures can be held under reduced illumination and temperature ($50 \mu\text{E m}^{-2} \text{ s}^{-1}$, $20 \pm 2 \text{ }^{\circ}\text{C}$). Details of culturing are given in Appendix 3.
19. At least 14 to 21 days before testing, sufficient test organisms are transferred aseptically into fresh sterile medium and cultured for 14 to 21 days under the conditions of the test as a pre culture. Details for preparation of a pre-culture are given in Appendix 4.

Test medium

20. Only one nutrient medium is recommended for *Myriophyllum spicatum* in a sediment-free test system, as described in Appendix 2. A modification of the Andrews' medium is recommended for culturing and testing with *Myriophyllum spicatum* as described in (1). From five separately prepared nutrient stock solutions with addition of 3 % sucrose the modified Andrews' medium will be arranged. Details about preparation of the medium are given in Appendix 2.
21. A tenfold concentrated, modified Andrews' medium is needed for obtaining the test solutions (by dilution as appropriate). The composition of this medium is given in Appendix 2.

Test solutions

22. Test solutions are usually prepared by dilution of a stock solution. Stock solutions of the test chemical are normally prepared by dissolving the chemical in demineralised (i.e. distilled or deionised) water. The addition of the nutrients will be achieved by using the tenfold concentrated, modified Andrews' medium.

23. The stock solutions of the test chemical can be sterilised by autoclave at 121 °C for 20 minutes or by sterile filtration, provided that the sterilisation technique used does not denaturise the test chemical. Test solutions can also be prepared in sterile demineralised water or medium, under sterile conditions. The thermo-stability and the adsorption on different surfaces should be taken into account in the selection of the sterilisation procedure of the stock solutions of the test chemical. Because of that, it is recommended that the stock solutions be prepared under sterile conditions, i.e. using sterile material for dissolving the test chemical under sterile conditions (e.g. flame sterilisation, laminar-flow hoods, etc.) into sterile water. This technique of preparation of sterile stock solutions is valid for both substances and mixtures.
24. The highest tested concentration of the test chemical should normally not exceed its water solubility under the test conditions. For test chemicals of low water solubility it may be necessary to prepare a concentrated stock solution or dispersion of the chemical using an organic solvent or dispersant in order to facilitate the addition of accurate quantities of the test chemical to the test medium and aid in its dispersion and dissolution. Every effort should be made to avoid the use of such materials. There should be no phytotoxicity resulting from the use of auxiliary solvents or dispersants. For example, commonly used solvents which do not cause phytotoxicity at concentrations up to 100 µl/l, include acetone and dimethylformamide. If a solvent or dispersant is used, its final concentration should be reported and kept to a minimum (≤ 100 µl/l), and all treatments and controls should contain the same concentration of solvent or dispersant. Further guidance on the use of dispersants is given in (5).

Test and control groups

25. Prior knowledge of the toxicity of the test chemical to *Myriophyllum spicatum* from a range-finding test will help in selecting suitable test concentrations. In the definitive toxicity test, there should normally be five (like in the *Lemna* growth inhibition test, Chapter C.26 of this Annex) to seven test concentrations arranged in a geometric series; they should be chosen in order that the NOEC and EC₅₀ values are bracketed by the concentration range (see below). Preferably the separation factor between test concentrations should not exceed 3,2; however, a larger value may be used where the concentration-response curve is flat. Justification should be provided when fewer than five concentrations are used. At least five replicates should be used at each test concentration.
26. In setting the range of test concentrations (for range-finding and/or for the definitive toxicity test), the following should be considered:

To determine an EC_x, test concentrations should bracket the EC_x value to ensure an appropriate level of confidence. For example, if estimating the EC₅₀, the highest test concentration should be greater than the EC₅₀ value. If the EC₅₀ value lies outside of the range of test concentrations, associated confidence intervals will be large and a proper assessment of the statistical fit of the model may not be possible.

If the aim is to estimate the LOEC/NOEC, the lowest test concentration should be low enough so that growth is not significantly less than that of the control. In addition, the highest test concentration should be high enough so that growth is significantly lower than that in the control. If this is not the case, the test will have to be repeated using a different concentration range (unless the highest concentration is at the limit of solubility or the maximum required limit concentration, e.g. 100 mg/l).

27. Every test should include controls consisting of the same nutrient medium, test organism (choosing plant material as homogeneous as possible, fresh lateral branches from pre-cultures, shortened to 2,5 cm from base), environmental conditions and procedures as the test vessels but without the test chemical. If an auxiliary solvent or dispersant is used, an additional control treatment with the solvent/dispersant present at the same concentration as that in the vessels with the test chemical should be included. The number of replicate control vessels (and solvent vessels, if applicable) should be at least ten.

28. If determination of NOEC is not required, the test design may be altered to increase the number of concentrations and reduce the number of replicates per concentration. However, in any case the number of control replicates should be at least ten.

Exposure

29. Fresh lateral branches from pre-culture shortened to 2,5 cm from base are assigned randomly to the test vessels under aseptic conditions; each test vessel should contain one 2,5 cm lateral branch that should have an apical meristem on one end. The chosen plant material should be the same quality in each test vessel.
30. A randomised design for location of the test vessels in the incubator is required to minimise the influence of spatial differences in light intensity or temperature. A blocked design or random repositioning of the vessels (or repositioning more frequently) when observations are made is also required.
31. If a preliminary stability test shows that the test chemical concentration cannot be maintained (*i.e.* the measured concentration falls below 80 % of the measured initial concentration) over the test duration (14 days), a semi-static test regime is recommended. In this case, the plants should be exposed to freshly prepared test and control solutions on at least one occasion during the test (*e.g.* day 7). The frequency of exposure to fresh medium will depend on the stability of the test chemical; a higher frequency may be needed to maintain near-constant concentrations of highly unstable or volatile chemicals.
32. The exposure scenario through a foliar application (spray) is not covered in this test method.

Test conditions

33. Warm and/or cool white fluorescent lighting should be used to provide light irradiance in the range of about of 100-150 $\mu\text{E m}^{-2} \text{ s}^{-1}$ when measured as a photosynthetically active radiation (400-700 nm) at points the same distance from the light source as the bottom of the test vessels (equivalent *ca.* 6 000 to 9 000 lux) and using a light-dark cycle of 16:8 h. The method of light detection and measurement, in particular the type of sensor, will affect the measured value. Spherical sensors (which respond to light from all angles above and below the plane of measurement) and “cosine” sensors (which respond to light from all angles above the plane of measurement) are preferred to unidirectional sensors, and will give higher readings for a multi-point light source of the type described here.
34. The temperature in the test vessels should be 23 ± 2 °C. Additional care is needed on pH drift in special cases such as when testing unstable chemicals or metals; the pH should remain in a range of 6-9. See (5) for further guidance.

Duration

35. The test is terminated 14 days after the plants are transferred into the test vessels.

Measurements and analytical determinations

36. At the start of the test, the main shoot length of test organism is 2,5 cm (see paragraph 29); it is measured with a ruler (see Appendix 4) or by photography and image analysis. The main shoot length of test organism appearing normal or abnormal needs to be determined at the beginning of the test, at least once during the 14-day exposure period and at test termination. Note: As an alternative for those who do not have image analysis, if the workbench is sterilised prior to addition of plants to test vessels, a sterile ruler can also be used to measure the length of the main shoot at test initiation and during the test. Changes in plant development, *e.g.* in deformation in the shoots, appearance, indication of necrosis, chlorosis, break-up or loss of buoyancy and in root length and appearance, should be noted. Significant features of the test medium (*e.g.* presence of undissolved material, growth of algae, fungi and bacteria in the test vessel) should also be noted.

37. In addition to determinations of main shoot length during the test, effects of the test chemical on three (or more) of the following measurement variables should be also assessed:

- (i) Total lateral branches length
- (ii) Total shoot length
- (iii) Total root length
- (iv) Fresh weight
- (v) Dry weight
- (vi) Number of whorls

Note 1: The observations made during the range-finding test could help in selecting relevant additional measurements among the six variables listed above.

Note 2: The determination of the fresh and dry weights (parameters iv and v) is highly desirable.

Note 3: Due to the fact that sucrose and light (exposure of roots to light during the test) may have an influence on auxin (plant growth hormone) transport carriers, and that some chemicals may have an auxin-type mode of action, the inclusion of root endpoints (parameter iii) is questionable.

Note 4: The ring test results show high coefficients of variation ($> 60\%$) for the total lateral branch length (parameter i). Total lateral branch length is in any case encompassed within the total shoot length measurement (parameter ii) which shows more acceptable coefficients of variation of $< 30\%$.

Note 5: Resulting from the above considerations, the recommended main measurement endpoints are: total shoot length, fresh weight and dry weight (parameters ii, iv and v); parameter vi — number of whorls — is left to the experimenter's judgment.

38. Main shoot length and number of whorls have an advantage, in that they can be determined for each test and control vessel at the start, during, and at the end of the test by photography and image analysis, although a (sterile) ruler can also be used.
39. Total lateral branches length, total root length (as a sum of all lateral branches or roots) and total shoot length (as a sum of main shoot length and total lateral branches length) can be measured with a ruler at the end of exposure.
40. The fresh and/or dry weight should be determined at the start of the test from a sample of the pre-culture representative of what is used to begin the test, and at the end of the test with the plant material from each test and control vessel.
41. Total lateral branches length, total shoot length, total root length, fresh weight, dry weight and number of whorls may be determined as follows:
- (i) Total lateral branches length: The lateral branch length may be determined by measuring all lateral branches with a ruler at the end of exposure. The total lateral branches length is the sum of all lateral branches of each test and control vessel.
 - (ii) Total shoot length: The main shoot length may be determined by image analysis or using a ruler. The total shoot length is the sum of the total lateral branches length and the main shoot length of each test and control vessel at the end of exposure.

- (iii) Total root length: The root length may be determined by measuring all roots with a ruler at the end of exposure. The total root length is the sum of all roots of each test and control vessel.
- (iv) Fresh weight: The fresh weight may be determined by weighing the test organisms at the end of exposure. All plant material of each test and control vessel will be rinsed with distilled water, dabbed dry with cellulose paper. After this preparation the fresh weight will be determined by weighing. The starting biomass (fresh weight) is determined on the basis of a sample of test organisms taken from the same batch used to inoculate the test vessels.
- (v) Dry weight: After the preparations for the determination of the fresh weight the test organisms will be dried at 60 °C to a constant weight. This mass is the dry weight. The starting biomass (dry weight) is determined on the basis of a sample of test organisms taken from the same batch used to inoculate the test vessels.
- (vi) Number of whorls: All whorls will be counted out along the main shoot.

Frequency of measurement and analytical determinations

- 42. If a static test design is used, the pH of each treatment should be measured at the beginning and at the end of the test. If a semi-static test design is used, the pH should be measured in each batch of 'fresh' test solution prior to each renewal and also in the corresponding 'spent' solutions.
- 43. Light intensity should be measured in the growth chamber, incubator or room at points in the same distance from the light source as the test organisms. Measurements should be made at least once during the test. The temperature of the medium in a surrogate vessel held under the same conditions in the growth chamber, incubator or room should be recorded at least daily (or continuously with a data logger).
- 44. During the test, the concentrations of the test chemical(s) are determined at appropriate intervals. In static tests, the minimum requirement is to determine the concentrations at the beginning and at the end of the test.
- 45. In semi-static tests where the concentrations of the test chemical(s) are not expected to remain within $\pm 20\%$ of the nominal concentration, it is necessary to analyse all freshly prepared test solutions and the same solutions at each renewal. However, for those tests where the measured initial concentrations of the test chemical(s) are not within $\pm 20\%$ of nominal but where sufficient evidence can be provided to show that the initial concentrations are repeatable and stable (i.e. within the range 80 - 120 % of the initial concentration), chemical determinations may be carried out on only the highest and lowest test concentrations. In all cases, determination of test concentrations prior to renewal need only be performed on one replicate vessel at each test concentration (or the contents of the vessels pooled by replicate).
- 46. If there is evidence that the test concentration has been satisfactorily maintained within $\pm 20\%$ of the nominal or measured initial concentration throughout the test, analysis of the results can be based on nominal or measured initial values. If the deviation from the nominal or measured initial concentration is not within $\pm 20\%$, analysis of the results should be based on the geometric mean concentration during exposure or models describing the decline of the concentration of the test chemical (5).

Limit test

- 47. Under some circumstances, e.g. when a preliminary test indicates that the test chemical has no toxic effects at concentrations up to 100 mg/l or up to its limit of solubility in the test medium or in case of a formulation up to its limit of dispersibility, a limit test involving a comparison of responses in a control group and one treatment group (100 mg/l or a concentration equal to the limit of solubility), may be undertaken. It is strongly recommended that this is supported by analysis of the exposure concentration. All previously described test conditions and validity criteria apply to a limit test, with the exception that the number of treatment replicates should be doubled. Growth in the control and treatment group may be analysed using a statistical test to compare means, e.g. a Student's t-test.

DATA AND REPORTING

Response variables

48. The purpose of the test is to determine the effects of a test chemical on the vegetative growth of *Myriophyllum spicatum*. This test method describes two response variables.
- (a) Average specific growth rate: This response variable is calculated on the basis of changes in the logarithms of main shoot length, and in addition, on the basis of changes in the logarithms of other measurement parameters, i.e. total shoot length, fresh weight, dry weight or number of whorls over time (expressed per day) in the controls and each treatment group. Note: For the measurement parameter total lateral branches length and total root length a calculation of the average specific growth rate is not possible. At the beginning of the test, the test organism has no lateral branches and no roots (based on the preparation from the pre-culture); starting from the value zero, the calculation of the average specific growth rate is not defined.
- (b) Yield: This response variable is calculated on the basis of changes in main shoot length, and in addition, on the basis of changes in other measurement parameters — i.e. preferably total shoot length, fresh weight, dry weight or number of whorls, and other parameters if deemed useful — in the controls and in each treatment group until the end of the test.
49. Toxicity estimates should be based on main shoot length and three additional measurement variables (i.e. preferably total shoot length, fresh weight, dry weight or number of whorls, see paragraph 37 and Notes 2, 4 and 5 to this paragraph), because some chemicals may affect other measurement variables much more than the main shoot length. This effect would not be detected by calculating main shoot length only.

Average specific growth rate

50. The average specific growth rate for a specific period is calculated as the logarithmic increase in the growth variables — main shoot length and three additional measurement variables (i.e. total shoot length, fresh weight, dry weight or number of whorls) — using the formula below for each replicate of control and treatments:

$$\mu_{i-j} = \frac{\ln(N_j) - \ln(N_i)}{t}$$

where:

μ_{i-j} : average specific growth rate from time i to j

N_i : measurement variable in the test or control vessel at time i

N_j : measurement variable in the test or control vessel at time j

t: time period from i to j

For each treatment group and control group, calculate a mean value for growth rate along with variance estimates.

51. The average specific growth rate should be calculated for the entire test period (time “i” in the above formula is the beginning of the test and time “j” is the end of the test). For each test concentration and control, calculate a mean value for average specific growth rate along with the variance estimates. In addition, the section-by-section growth rate should be assessed in order to evaluate effects of the test chemical occurring during the exposure period (e.g. by inspecting log-transformed growth curves).

52. Percent inhibition of growth rate (I_r) may then be calculated for each test concentration (treatment group) according to the following formula:

$$\%I_r = \frac{(\mu_c - \mu_T)}{\mu_c} \times 100$$

where:

% I_r : percent inhibition in average specific growth rate

μ_c : mean value for μ in the control

μ_T : mean value for μ in the treatment group

Yield

53. Effects on yield are determined on the basis of the measurement variable main shoot length and three additional measurement variables (*i.e.* preferably total shoot length, fresh weight, dry weight or number of whorls) present in each test vessel at the start and at the end of the test. For fresh weight or dry weight, the starting biomass is determined on the basis of a sample of test organisms taken from the same batch used to inoculate the test vessels. For each test concentration and control, calculate a mean value for yield along with variance estimates. The mean percent inhibition in yield (% I_y) may be calculated for each treatment group as follows:

$$\%I_y = \frac{(b_c - b_T)}{b_c}$$

where:

% I_y : percent reduction in yield

b_c : final biomass minus starting biomass for the control group

b_T : final biomass minus starting biomass in the treatment group

Doubling time

54. To determine the doubling time (T_d) of main shoot length and adherence to this validity criterion (see paragraph 8), the following formula is used with data obtained from the control vessels:

$$T_d = \ln 2 / \mu$$

Where μ is the average specific growth rate determined as described in paragraphs 50-52.

Plotting concentration-response curves

55. Concentration-response curves relating mean percentage inhibition of the response variable (I_r or I_y calculated as shown in paragraph 53) and the log concentration of the test chemical should be plotted.

EC_x estimation

56. Estimates of the EC_x should be based upon both average specific growth rate ($E_r C_x$) and yield ($E_y C_x$), each of which should in turn be based upon main shoot length, and possibly additional measurement variables (*i.e.* preferably total shoot length, fresh weight, dry weight or number of whorls). This is because there are chemicals that impact main shoot length and other measurement variables differently. The desired toxicity parameters are therefore four EC_x values for each inhibition level x calculated: $E_r C_x$ (main shoot length); $E_r C_x$ (*i.e.* preferably total shoot length, fresh weight, dry weight, or number of whorls); $E_y C_x$ (main shoot length); and $E_y C_x$ (*i.e.* preferably total shoot length, fresh weight, dry weight or number of whorls).

57. It should be noted that EC_x values calculated using these two response variables are not comparable and this difference is recognised when using the results of the test. EC_x values based upon average specific growth rate (E_rC_x) will in most cases be higher than results based upon yield (E_yC_x) — if the test conditions of this test method are adhered to — due to the mathematical basis of the respective approaches. This difference should not be interpreted as a difference in sensitivity between the two response variables, simply the values are different mathematically.

Statistical procedures

58. The aim is to obtain a quantitative concentration-response relationship by regression analysis. It is possible to use a weighted linear regression after having performed a linearising transformation of the response data, for instance with probit or logit or Weibull models (7), but non-linear regression procedures are preferred techniques that better handle unavoidable data irregularities and deviations from smooth distributions. Approaching either zero or total inhibition such irregularities may be magnified by the transformation, interfering with the analysis (7). It should be noted that standard methods of analysis using probit, logit, or Weibull transforms are intended for use on quantal (e.g. mortality or survival) data, and should be modified to accommodate growth rate or yield data. Specific procedures for determination of EC_x values from continuous data can be found in (8) (9) (10).
59. For each response variable to be analysed, use the concentration-response relationship to calculate point estimates of EC_x values. When possible, the 95 % confidence limits for each estimate should be determined. Goodness of fit of the response data to the regression model should be assessed either graphically or statistically. Regression analysis should be performed using individual replicate responses, not treatment group means.
60. EC_{50} estimates and confidence limits may also be obtained using linear interpolation with bootstrapping (10), if available regression models/methods are unsuitable for the data.
61. For estimation of the LOEC and hence the NOEC, it is necessary to compare treatment means using analysis of variance (ANOVA) techniques. The mean for each concentration is then compared with the control mean using an appropriate multiple comparison or trend test method. Dunnett's or Williams' test may be useful (12) (13) (14) (15) (16). It is necessary to assess whether the ANOVA assumption of homogeneity of variance holds. This assessment may be performed graphically or by a formal test (15). Suitable tests are Levene's or Bartlett's. Failure to meet the assumption of homogeneity of variances can sometimes be corrected by logarithmic transformation of the data. If heterogeneity of variance is extreme and cannot be corrected by transformation, analysis by methods such as step-down Jonkheere trend tests should be considered. Additional guidance on determining the NOEC can be found in (10).
62. Recent scientific developments have led to a recommendation of abandoning the concept of NOEC and replacing it with regression based point estimates EC_x . An appropriate value for x has not been established for this *Myriophyllum* test. However, a range of 10 to 20 % appears to be appropriate (depending on the response variable chosen), and preferably both the EC_{10} and EC_{20} and their confidence limits should be reported.

Reporting

63. The test report includes the following:

Test chemical

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 or mixture:

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

Test species

- Scientific name and source.

Test conditions

- Test procedure used (static or semi-static).
- Date of start of the test and its duration.
- Test medium.
- Description of the experimental design: test vessels and covers, solution volumes, main shoot length per test vessel at the beginning of the test.
- Test concentrations (nominal and measured as appropriate) and number of replicates per concentration.
- Methods of preparation of stock and test solutions including the use of any solvents or dispersants.
- Temperature during the test.
- Light source, light intensity and homogeneity.
- pH values of the test and control media.
- The method of analysis of test chemical with appropriate quality assessment data (validation studies, standard deviations or confidence limits of analyses).
- Methods for determination of main shoot length and other measurement variables, e.g. total lateral branches length, total shoot length, total root length, fresh weight, dry weight or number of whorls.
- State of the culture (sterile or non-sterile) of each test and control vessel at each observation.
- All deviations from this test method.

Results

- Raw data: main shoot length and other measurement variables in each test and control vessel at each observation and occasion of analysis.
- Means and standard deviations for each measurement variable.
- Growth curves for each measurement variable.
- Calculated response variables for each treatment replicate, with mean values and coefficient of variation for replicates.
- Graphical representation of the concentration/effect relationship.
- Estimates of toxic endpoints for response variables e.g. EC₅₀, EC₁₀, EC₂₀, and associated confidence intervals. If calculated, LOEC and/or NOEC and the statistical methods used for their determination.
- If ANOVA has been used, the size of the effect which can be detected (e.g. the least significant difference).
- Any stimulation of growth found in any treatment.
- Any visual signs of phytotoxicity as well as observations of test solutions.
- Discussion of the results, including any influence on the outcome of the test resulting from deviations from this test method.

LITERATURE:

- (1) ASTM Designation E 1913-04, Standard Guide for Conducting Static, Axenic, 14-Day Phytotoxicity Tests in Test Tubes with the Submersed Aquatic Macrophyte, *Myriophyllum sibiricum* Komarov.
 - (2) Maletzki, D. et al. (2010), *Myriophyllum spicatum* als ökotoxikologischer Testorganismus: Methodenentwicklung eines sedimentfreien Testsystems und erste Ergebnisse mit 3,5-Dichlorphenol, *Umweltwiss Schadst Forsch*, No. 22, pp. 702–710.
 - (3) Chapter C.26 of this Annex: *Lemna* sp. Growth Inhibition Test,
 - (4) OECD (2014), “*Myriophyllum spicatum* Toxicity Test: Results of an inter-laboratory ring test using a sediment-free test system”, OECD Environment, Health and Safety Publications (EHS), Series Testing and Assessment, No. 205, OECD Publishing, Paris.
 - (5) OECD (2000), “Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures”, OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 23, OECD Publishing, Paris.
 - (6) Chapter C.51 of this Annex: Water-Sediment *Myriophyllum spicatum* Toxicity Test
 - (7) Christensen, E.R., N. Nyholm (1984), Ecotoxicological Assays with Algae: Weibull Dose-Response Curves, *Environmental Science & Technology*, Vol. 18/9, 713-718.
 - (8) Nyholm, N. et al. (1992), Statistical treatment of data from microbial toxicity tests, *Environmental Toxicology and Chemistry*, Vol. 11/2, pp. 157-167.
 - (9) Bruce, R.D., D.J. Versteeg (1992), A statistical procedure for modelling continuous toxicity data, *Environmental Toxicology and Chemistry*, Vol. 11/10, pp. 1485-1494.
 - (10) OECD (2006), “Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application”, OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 54, OECD Publishing, Paris.
 - (11) Norberg-King, T.J. (1988), An interpolation estimate for chronic toxicity: The ICp approach, National Effluent Toxicity Assessment Center Technical Report 05-88, US EPA, Duluth, MN.
 - (12) Dunnett, C.W. (1955), A multiple comparisons procedure for comparing several treatments with a control, *Journal of the American Statistical Association*, Vol. 50/272, pp. 1096-1121.
 - (13) Dunnett, C.W. (1964), New tables for multiple comparisons with a control, *Biometrics*, Vol. 20/3, pp. 482-491.
 - (14) Williams, D.A. (1971), A test for differences between treatment means when several dose levels are compared with a zero dose control, *Biometrics*, Vol. 27/1, pp. 103-117.
 - (15) Williams, D.A. (1972), The comparison of several dose levels with a zero dose control, *Biometrics*, Vol. 28/2, pp. 519-531.
 - (16) Brain, P., R. Cousens (1989), An equation to describe dose-responses where there is stimulation of growth at low doses, *Weed Research*, Vol. 29/2, pp. 93-96.
-

Appendix 1

DEFINITIONS

Biomass is the fresh and/or dry weight of living matter present in a population. In this test the biomass is the sum of main shoot, all lateral branches and all roots.

Chemical is a substance or a mixture.

Chlorosis is the change of the color from green to yellowing of test organism especially of the whorls.

EC_x is the concentration of the test chemical dissolved in test medium that results in a x % (e.g. 50 %) reduction in growth of *Myriophyllum spicatum* within a stated exposure period (to be mentioned explicitly if deviating from full or normal test duration). To unambiguously denote an EC value deriving from growth rate or yield the symbol " E_rC " is used for growth rate and " E_yC " is used for yield, followed by the measurement variable used, e.g. E_rC (main shoot length).

Growth is an increase in the measurement variable, e.g. main shoot length, total lateral branches length, total shoot length, total root length, fresh weight, dry weight or number of whorls, over the test period.

Growth rate (average specific growth rate) is the logarithmic increase in the measurement variable during the exposure period. *Note:* Growth rate related response variables are independent of the duration of the test as long as the growth pattern of unexposed control organisms is exponential.

Lowest Observed Effect Concentration (LOEC) is the lowest tested concentration at which the chemical is observed to have a statistically significant reducing effect on growth (at $p < 0,05$) when compared with the control, within a given exposure time. 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.

Measurement variables are any type of variables which are measured to express the test endpoint using one or more different response variables. In this test method main shoot length, total lateral branches length; total shoot length, total root length, fresh weight, dry weight and number of whorls are measurement variables.

Monoculture is a culture with one plant species.

Necrosis is dead (i.e. white or dark brown) tissue of the test organism.

No Observed Effect Concentration (NOEC) is the test concentration immediately below the LOEC.

Response variable is a variable for the estimation of toxicity derived from any measured variable describing biomass by different methods of calculation. For this test method growth rate and yield are response variables derived from measurement variables like main shoot length, total shoot length, fresh weight, dry weight, or number of whorls.

Semi-static (renewal) test is a test in which the test solution is periodically replaced at specific intervals during the test.

Static test is a test method without renewal of the test solution during the test.

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

Test endpoint describes the general factor that will be changed relative to control by the test chemical as aim of the test. In this test method the test endpoint is inhibition of growth which may be expressed by different response variables which are based on one or more measurement variables.

Test medium is the complete synthetic growth medium on which test plants grow when exposed to the test chemical. The test chemical will normally be dissolved in the test medium.

UVCB is a substance of unknown or variable composition, complex reaction product or biological material

Yield is value of a measurement variable to express biomass at the end of the exposure period minus the measurement variable at the start of the exposure period. Note: When the growth pattern of unexposed organisms is exponential, yield-based response variables will decrease with the test duration.

Appendix 2

MODIFIED ANDREWS' MEDIUM FOR STOCK CULTURE AND PRE-CULTURE

From five separately prepared nutrient stock solutions the modified Andrews' medium required for stock culture and pre culture will be prepared, with addition of 3 % sucrose.

Table 1

Composition of Andrews' nutrient solution: (ASTM Designation E 1913-04)

Production of nutrient stock solutions			Production of nutrient solution
Stock solution	Chemical	Initial weight per 1 000 ml	ml per 5 l nutrient solution
1	KCl	74,6 mg	50
	KNO ₃	8,08 g	
	Ca(NO ₃) ₂ × 4 H ₂ O	18,88 g	
2	MgSO ₄ × 7 H ₂ O	9,86 g	50
3	See below stock solution 3.1		50
4	KH ₂ PO ₄	2,72 g	50
5	FeSO ₄ × 7 H ₂ O	0,278 g	50
	Na ₂ EDTA × 2 H ₂ O	0,372 g	

Stock solutions can be kept in a refrigerator for 6 months (at 5-10 °C). Only stock solution No. 5 has a reduced shelf life (two months).

Table 2

Production of stock solution 3.1 for preparing stock solution 3

Chemical	Initial weight g/100 ml
MnSO ₄ × 4 H ₂ O	0,223
ZnSO ₄ × 7 H ₂ O	0,115
H ₃ BO ₃	0,155
CuSO ₄ × 5 H ₂ O	0,0125
(NH ₄) ₆ Mo ₇ O ₂₄ × 4 H ₂ O	0,0037

After having produced stock solution 3.1 (Table 2), deep-freeze this solution in approximately 11 ml-aliquots (at – 18 °C at least). The deep-frozen portions have a shelf life of five years.

To produce stock solution 3, defrost stock solution 3.1, fill 10 ml of it into a 1 l volumetric flask and add ultra-pure water up to the flask's mark.

To obtain modified Andrews' medium, fill approximately 2 500 ml ultra-pure water into a 5 l volumetric flask. After adding 50 ml of each stock solution, fill 90 % of the volumetric flask with ultra-pure water and set pH to 5,8.

After this, add 150 g dissolved sucrose (3 % per 5 l); then, fill the volumetric flask with ultra-pure water up to the mark. Finally, the nutrient solution is filled into 1 l Schott flasks and autoclaved at 121 °C for 20 minutes.

The nutrient solution thus yielded can be kept sterile in a refrigerator (at 5-10 °C) for three months.

Modified Andrews' medium for Sediment-free toxicity test

From the five nutrient stock solutions already mentioned in Tables 1 and 2, a tenfold concentrated, modified Andrews' medium required for obtaining the test solutions will be prepared, with addition of 30 % sucrose. To do so, fill approximately 100 ml ultra-pure water into a 1 l volumetric flask. After adding 100 ml of each of the stock solutions, set pH to 5,8. After this, add 30 % dissolved sucrose (300 g per 1 000 ml); then, fill the volumetric flask with ultra-pure water up to the mark.

Finally, the nutrient solution is filled into 0,5 l Schott flasks and autoclaved at 121 °C for 20 minutes.

The tenfold concentrated modified nutrient solution thus yielded can be kept sterile in a refrigerator (at 5-10 °C) for three months.

Appendix 3

MAINTENANCE OF STOCK CULTURE

In this Appendix 3 the stock culture of *Myriophyllum spicatum* L.⁽¹⁾, a submersed aquatic dicotyledon, a species of the water milfoils family is described. Between June and August, inconspicuous pink-white flowers protrude above the water surface. The plants are rooted in the ground by a system of robust rhizomes and can be found in the entire northern hemisphere in eutrophic, however non-polluted and more calciferous still waters with muddy substrate. *Myriophyllum spicatum* prefers fresh water, but is found in brackish water as well.

For sediment-free stock culture under laboratory conditions, sterile plants are required. Sterile plants are available from the ecotoxicology laboratory of the German Umweltbundesamt (Federal Environment Agency of Germany).

Alternatively, test organisms can be prepared from non-sterile plants in accordance with ASTM designation E 1913-04. See below — extracted from the ASTM Standard Guide — the procedure for culturing *Myriophyllum sibiricum* collected from field:

"If starting from field collected, non-sterile plants, collect *M. sibiricum* turions in the autumn. Place the turions into a 20-l aquarium containing 5 cm of sterile sediment that is covered with silica sand or for example by Turface® and 18 l of reagent water. Aerate the aquarium and maintain at a temperature of 15 °C and a fluence rate of 200 to 300 $\mu\text{mol m}^{-2} \text{s}^{-1}$ for 16 h per day. The plant culture in the aquarium may be maintained as a backup source of plants in case the sterile plant cultures are destroyed by mechanical malfunction in the growth cabinet, contamination, or other reason. The plants grown in the aquarium are not sterile and sterile cultures cannot be maintained in a batch culturing system. To sterilize the culture, plants are removed from the aquarium and rinsed under flowing deionized water for about 0,5 h. Under aseptic conditions in a laminar airflow cabinet, the plants are disinfected for less than 20 min (until most of the plant tissue is bleached and just the growing apex is still green) in a 3 % (w/v) sodium hypochlorite solution containing 0,01 % of a suitable surfactant. Agitate the disinfectant and plant material. Segments with several nodes are transferred into sterile culture tubes containing 45 ml of sterilized modified Andrews' medium and capped with plain culture tube closures. Only one plant segment is placed into each test chamber. Laboratory sealant film is used to secure the closure to the culture vessel. Once a sterile culture has been established, plant segments containing several nodes should be transferred to new test chambers containing fresh liquid nutrient media every ten to twelve days. As demonstrated by culturing on agar plates, the plants must be sterile and remain sterile for eight weeks before testing can be initiated."

Since the modified Andrews' medium contains sucrose (which stimulates the growth of fungi and bacteria), all material, solutions and culturing be conducted under sterile conditions. All liquids as well as equipment are sterilised before use. Sterilisation is carried out via heated air treatment (210 °C) for 4 hours or autoclaving for 20 minutes at 121 °C. In addition, all flasks, dishes, bowls etc and other equipment undergo flame treatment at the sterile workbench just prior to use.

Stock cultures can be maintained under reduced illumination and temperature ($50 \mu\text{E m}^{-2} \text{s}^{-1}$, 20 ± 2 °C) for longer times without needing to be re-established. The *Myriophyllum* growth medium should be the same as that used for testing but other nutrient rich media can be used for stock cultures.

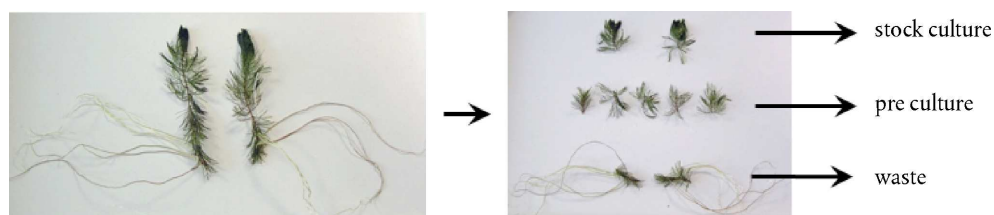
The plant segments are distributed axenically over several 500 ml Erlenmeyer or/and 2 000 ml Fernbach flasks, each filled with approximately 450 respectively 1 000 ml modified Andrews' medium. Then, the flasks are axenically cellulose plug stoppered.

In addition, thorough flame treatment of equipment at the sterile workbench just prior to use is absolutely necessary. Dependent on number and size, the plants are to be transferred into fresh nutrient solution approximately every three weeks.

Apices as well as segments of the stem middle part for this renewed culture can be used. Number and size of transferred plants (or segments of plants) are dependent on how many plants are needed. For example, you can transfer five shoot segments into one Fernbach flask and three shoot segments into one Erlenmeyer flask, each with a length of 5 cm. Discard any rooted, flowering, dead or otherwise conspicuous parts.

⁽¹⁾ Carl von Linné (* May, 23th, 1707 in Råshult/Älmhult; † January, 10th, 1778 in Uppsala).

Figure 1

Cutting of plants for the stock and pre culture after 3 weeks of cultivation

Culturing of plants is to be performed in 500 ml Erlenmeyer and 2 000 ml Fernbach flasks in a cooling incubator at 20 ± 2 °C with continuously light at approximately $100\text{--}150 \mu\text{E m}^{-2} \text{s}^{-1}$ or 6 000–9 000 Lux (emitted by chamber illumination with colour temperature “warm white light”).

Figure 2

Culturing of plants in a cooling incubator with chamber illumination

Chemically clean (acid-washed) and sterile glass culture vessels should be used and aseptic handling techniques employed. In the event of contamination of the stock culture e.g. by algae, fungi and/or bacteria a new culture should be prepared or a stock culture from another laboratory should be used to renewal of the one culture.

Appendix 4

MAINTENANCE OF PRE-CULTURE AND PREPARATION OF TEST ORGANISM FOR TESTING

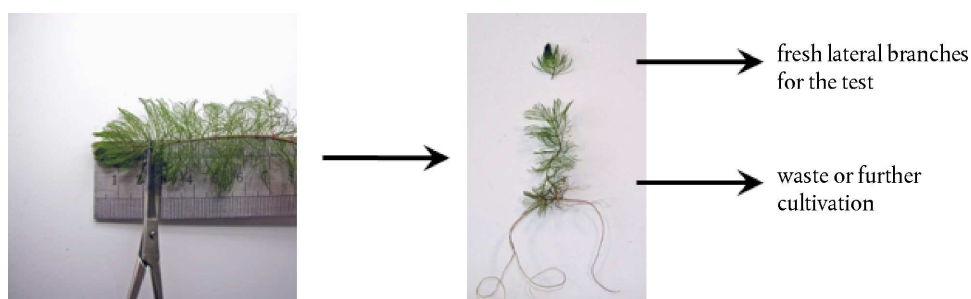
To obtain pre-culture, cut shoots of stock culture into segments with two whorls each; put segments into Fernbach flasks filled with modified Andrews' medium (with 3 % sucrose). Each flask can contain up to 50 shoot segments. However, care is to be taken that the segments are vital and do not have any roots and lateral branches or their buds (see figure 1 in Appendix 3).

The pre-culture organisms are cultured for 14 to 21 days under sterile conditions in an environmental chamber with alternating 16/8 hour light/dark phases. Light intensity selected from the range of 100-150 $\mu\text{E m}^{-2} \text{s}^{-1}$. The temperature in the test vessels should be $23 \pm 2^\circ\text{C}$.

Since the modified Andrews' medium contains sucrose (which stimulates the growth of algae, fungi and bacteria), test chemical solutions should be prepared and culturing be conducted under sterile conditions. All liquids as well as equipment are sterilised before use. Sterilisation is carried out via heated air treatment (210°C) for 4 hours or autoclaving for 20 minutes at 121°C . In addition, all flasks, dishes, bowls etc. and other equipment undergo flame treatment at the sterile workbench just prior to use.

Shoots are axenically removed from the pre-culture flasks, choosing material that is as homogeneous as possible. Each testing requires at least 60 test organisms (testing with eight test chemical concentrations). For testing, take fresh lateral branches from pre-cultures, shorten them to 2,5 cm from base (measured with ruler) and transfer them into a beaker containing sterile modified Andrews' medium. These fresh lateral branches can be used for the sediment-free *Myriophyllum spicatum* toxicity test.

Figure 2

Cutting of plants from the pre culture for the sediment-free *Myriophyllum spicatum* toxicity test

C.51 water-sediment *Myriophyllum spicatum* toxicity test**INTRODUCTION**

1. This test method is equivalent to the OECD test guideline 239 (2014). Test methods are available for the floating, monocotyledonous aquatic plant, *Lemna* species (1) and for algal species (2). These methods are routinely used to generate data to address the risk of test chemicals, in particular chemicals with herbicidal activity, to non-target aquatic plant species. However, in some cases, data for additional macrophyte species may be required. Recent guidance published from the Society of Environmental Toxicology and Chemistry (SETAC) workshop on Aquatic Macrophyte Risk Assessment for Pesticides (AMRAP) proposed that data for a rooted macrophyte species may be required for test chemicals where *Lemna* and algae are known not to be sensitive to the mode of action or if partitioning to sediment is a concern, leading to exposure via root uptake (3). Based on current understanding and experience, *Myriophyllum* spp were selected as the preferred species in cases where data are required for a submerged, rooted dicotyledonous species (4) (5) (6). This test does not replace other aquatic toxicity tests; it should rather complement them so that a more complete aquatic plant hazard and risk assessment is possible. The water-sediment *Myriophyllum spicatum* test method complements the sediment-free *Myriophyllum spicatum* Toxicity Test (7).
2. This document describes a test method, which allows assessment of the effects of a test chemical on the rooted, aquatic plant species *Myriophyllum spicatum*, growing in a water-sediment system. The test method is based partly on existing methods (1) (2) (8) and takes account of recent research related to the risk assessment of aquatic plants (3). The water-sediment method has been validated by an international ring-test conducted with *Myriophyllum* species grown under static conditions, which were exposed to the test chemical through applications made via the water column (9). However, the test system is readily adapted to allow for exposure via spiked sediment or exposure via the water phase in semi-static or pulsed-dose scenarios, although these scenarios have not been formally ring tested. Furthermore, the general method can be used for other rooted, submerged and emergent species including other *Myriophyllum* species (e.g. *Myriophyllum aquaticum*) and *Glyceria maxima* (10). Modifications of test conditions, design and duration may be required for alternative species. In particular, more work is needed to define appropriate procedures for *Myriophyllum aquaticum*. These options are not presented in detail in this test method, which describes the standard approach for exposure of *Myriophyllum spicatum* in a static system via the water phase.
3. This test method applies to substances, for which the test method has been validated, (see details in the ring test report (9)) or to formulations or known mixtures. A *Myriophyllum* test may be conducted to fulfil a Tier 1 data requirement triggered by potential test chemical partitioning to sediment or mode of action/selectivity issues. Equally, a laboratory-based *Myriophyllum* test may be required as part of a higher-tier strategy to address concerns over the risk to aquatic plants. The specific reason for conducting a test will determine the route of exposure (i.e. via water or sediment). Before use of this test method for the testing of a mixture intended for a 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

4. The test is designed to assess chemical-related effects on the vegetative growth of *Myriophyllum* plants growing in standardised media (water, sediment and nutrients). For this purpose, shoot apices of healthy, non-flowering plants are potted in standardised, artificial sediment, which is supplemented with additional nutrients to ensure adequate plant growth, and then maintained in Smart and Barko medium (Appendix 1). After an establishment period to allow for root formation, plants are exposed to a series of test concentrations added to the water column. Alternatively, exposure via the sediment may be simulated by spiking the artificial sediment with the test chemical and transplanting plants into this spiked sediment. In both cases, plants are subsequently maintained under controlled environmental conditions for 14 days. Effects on growth are determined from quantitative assessments of shoot length, fresh weight and dry weight, as well as qualitative observations of symptoms such as chlorosis, necrosis or growth deformities.

5. To quantify chemical-related effects, growth in the test solutions is compared with the growth of the control plants, and the concentration causing a specified x % inhibition of growth is determined and expressed as the EC_x ; " x " can be any value depending on the regulatory requirements, e.g. EC_{10} , EC_{20} and EC_{50} . It should be noted that estimates of EC_{10} and EC_{20} values are only reliable and appropriate in tests where coefficients of variation in control plants fall below the effect level being estimated, i.e. coefficients of variation should be < 20 % for robust estimation of an EC_{20} .
6. Both average specific growth rate (estimated from assessments of shoot length, shoot fresh weight and shoot dry weight) and yield (estimated from the increase in shoot length, shoot fresh weight and shoot dry weight) of untreated and treated plants should be determined. Specific growth rate (r) and yield (y) are subsequently used to determine the $E_r C_x$ (e.g. $E_r C_{10}$, $E_r C_{20}$, $E_r C_{50}$) and $E_y C_x$ (e.g. $E_y C_{10}$, $E_y C_{20}$, $E_y C_{50}$), respectively.
7. If required, the lowest observed effect concentration (LOEC) and the no observed effect concentration (NOEC) may be statistically determined from estimates of average specific growth rates and yield.

INFORMATION ON THE TEST CHEMICAL

8. An analytical method with adequate sensitivity for quantification of the chemicals in the test medium should be available.
9. Information on the test chemical which may be useful in establishing the test conditions includes the structural formula, composition in the case of multi-constituent substances, UVCBs, mixtures or formulations, purity, water solubility, stability in water and light, acid dissociation constant (pK_a), partition coefficient octanol-water (K_{ow}), if available K_d to sediments, vapour pressure and biodegradability. Water solubility and vapour pressure can be used to calculate Henry's Law constant, which will indicate whether significant losses of the test chemical during the test period are likely. If losses of the test chemicals are likely, the losses should be quantified and the subsequent steps to control such losses should be documented. Where information on the solubility and stability of the test chemical(s) is uncertain, it is recommended that these properties are assessed under the conditions of the test, i.e. growth medium, temperature, lighting regime to be used in the test. *Note:* when light dependent peroxidising herbicides are tested, the laboratory lighting used should contain the equivalent presence of solar ultraviolet light found in natural sunlight.
10. The pH should be measured and adjusted in the test medium as appropriate. The pH control of the test medium is particularly important, e.g. when testing metals or chemicals which are hydrolytically unstable. Further guidance for testing chemicals with physical-chemical properties that make them difficult to test is provided in a OECD Guidance Document (11).

VALIDITY OF THE TEST

11. For the test results to be valid, the mean total shoot length and mean total shoot fresh weight in control plants at least double during the exposure phase of the test. In addition, control plants must not show any visual symptoms of chlorosis and should be visibly free from contamination by other organisms such as algae and/or bacterial films on the plants, at the surface of the sediment and in the test medium.
12. The mean coefficient of variation for yield based on measurements of shoot fresh weight (i.e. from test initiation to test termination) in the control cultures does not exceed 35 % between replicates.

REFERENCE CHEMICAL

13. A reference chemical(s), such as 3,5-dichlorophenol used in the ring test (9), should be periodically tested in order to check the performance of the test procedure over time. The ring test data indicate that the mean EC_{50} values of 3,5-DCP for the different response variables were between 4,7 and 6,1 mg/l (see the ring-test report for details of anticipated confidence interval around these values). It is advisable to test a reference chemical at least twice a year or, where testing is carried out infrequently, in parallel with the definitive toxicity tests. A guide to expected EC_{50} values for 3,5-DCP is provided in the Statistical Report of the International Ring-test (9).

DESCRIPTION OF THE METHOD

Test apparatus

14. The test should be conducted under controlled environmental conditions, i.e. in a growth chamber, growth room or laboratory, with controllable day length, lighting and temperature (see section “Test conditions”, paragraphs 56-58). Stock cultures should be maintained separately from test vessels.
15. The study should be conducted using glass test vessels such as aquaria or beakers; 2-l glass beakers (approximately 24 cm high and 11 cm in diameter) are commonly used. However, other (i.e. larger) vessels may be suitable provided that there is sufficient depth of water to allow unlimited growth and keep the plants submerged throughout the test duration.
16. Plastic or glass plant pots (approximately 9 cm diameter and 8 cm high and 500 ml volume) may be used as containers for potting the plants into the sediment. Alternatively, glass beakers may be used and are preferred in some cases (e.g. testing hydrophobic chemicals or chemicals with high K_{ow}).
17. The choice of pot/beaker size needs to be considered alongside the choice of test vessels and the preferred test design (see below). If using Test Design A (one shoot per pot with three pots per vessel) then smaller pots or larger vessels may be needed. If using Test Design B (three shoots per pot and one pot per vessel) then the stated pot and vessel sizes should be adequate. In all cases, the minimum overlaying water depth should be 12 cm above the top of the sediment and the ratio of sediment surface area/volume to water surface area/volume should be recorded.

Test organism

18. The general approaches described in this test method can be used to test a range of aquatic plant species. However, the conditions outlined in this test method have been tailored for testing the water milfoil species, *Myriophyllum spicatum*. This species belongs to the dicotyledonous family, Haloragaceae.
19. *Myriophyllum spicatum* (Eurasian water milfoil) is a submerged, rooted species which tolerates a wide range of conditions and is found in both static and flowing water bodies. *M. spicatum* is a perennial which dies back to the roots over winter. Plants usually flower and set seed freely although vegetative propagation from axillary buds or stem fragments that detach naturally or after disturbance, is often the primary method of colonisation.

Cultivation of the test organism

20. Plants may be obtained from natural populations or via aquatic plant suppliers. In both cases, the source of the plants should be documented and species identity should be verified. Great care should be taken to ensure that the correct species is obtained when collecting *Myriophyllum spicatum* from the field, especially in regions where it can hybridise with other *Myriophyllum* species. If in doubt, use of verified laboratory cultures from known sources is recommended. Plants that have been exposed to any chemical contaminants, or collected from sites known to be contaminated, should not be used in this test.
21. In regions where *M. spicatum* is not readily available during the winter months, long-term maintenance of stock cultures may be necessary under glasshouse or laboratory conditions. Stock cultures should be maintained under conditions similar to the test conditions although irradiance and temperature may be reduced in order to reduce the frequency of culture maintenance (e.g. when *Myriophyllum* tests are not planned for a period). Use of larger aquaria and plant pots, than would be used in tests, is recommended in order to allow space for proliferation. Sediment and water-media composition should be the same as would be used for tests although alternative methods of sediment fertilisation may be adopted (e.g. use of commercial slow-release fertiliser formulations)

22. Stock plants should be visibly free of contamination with any other organisms, including snails, filamentous algae, fungi and insects, e.g. eggs or larvae of the moth *Paraponyx stratiotata* and larve or adults of the curculionidae *Eubrychius velutus*. Rinsing plant material in fresh water may be necessary to eliminate visible contamination. In addition, efforts should be made to minimise the development of unicellular algae and bacterial contamination although complete sterility of the plant material is not necessary. Stock cultures should be monitored and transplanted as necessary to avoid development of algal and bacterial contamination. Aeration of stock cultures may be beneficial should algal or bacterial contamination become problematic.
23. In all cases, plants are cultured/ acclimatised under conditions that are similar, but not necessarily identical, to those used in the test for an adequate period (i.e. > 2 weeks) before their use in a test.
24. Flowering stock cultures should not be used in a test as vegetative growth rates generally decline during and after flowering.

Sediment

25. The following formulated sediment, based on the artificial sediment used in Chapter C.28 of this Annex (8), is recommended for use in this test. The sediment is prepared as described in TM C.28, except for the addition of nutrients as described below:
 - (a) 4-5 % peat (dry weight, according to $2 \pm 0,5$ % organic carbon) as close to pH 5,5 to 6,0 as possible; it is important to use peat in powder form, finely ground (preferably particle size < 1 mm) and only air dried.
 - (b) 20 % (dry weight) kaolin clay (kaolinite content preferably above 30 %).
 - (c) 75-76 % (dry weight) quartz sand (fine sand should predominate with more than 50 % of the particles between 50 and 200 μm).
 - (d) An aqueous nutrient medium is added such that the final sediment batch contains 200 mg/Kg dry sediment of both ammonium chloride and sodium phosphate and the moisture content of the final mixture is in a range of 30-50 %.
 - (e) Calcium carbonate of chemically pure quality (CaCO_3) is added to adjust the pH of the final mixture of the sediment to $7,0 \pm 0,5$.
26. The source of peat, kaolin clay and sand should be known and documented. If the origin is unknown or gives some level of concern, then the respective components should be checked for the absence of chemical contamination (e.g. heavy metals, organochlorine compounds, organophosphorous compounds).
27. The dry constituents of the sediment should be mixed homogenously prior to mixing the aqueous nutrient solution thoroughly into the sediment. The moist sediment should be prepared at least two days before use to allow thorough soaking of the peat and to prevent hydrophobic peat particles floating to the surface when the sediment is overlaid with media; before use, the moist sediment may be stored in the dark.
28. For the test, the sediment is transferred into a suitable size containers, such as plant pots of a diameter which fit into the glass vessels (the sediment surface area should cover approximately 70 % or more of the vessel surface area). In cases where the container has holes at the bottom, a piece of filter paper in the bottom of the container will help to keep the sediment within the container. The pots are filled with the sediment such that the sediment surface is level, prior to covering with a thin layer (~ 2 to 3 mm) of an inert material such as sand, fine horticultural grit (or crushed coral) to keep the sediment in place.

Test medium

29. Smart and Barko medium (12) is recommended for culturing and testing *Myriophyllum spicatum*. Preparation of this media is described in the Appendix 1. The pH of the media (water phase) at test initiation should be between 7,5 and 8,0 for optimum plant growth.

Experimental design

30. The test should incorporate a minimum of six replicate test vessels for the untreated control and a minimum of four replicate test vessels for each of a minimum of five concentration levels.
31. If determination of NOEC is not required, the test design may be altered to increase the number of concentrations and reduce the number of replicates per concentration.
32. Each test vessel represents a replicate containing three shoots. There are two options for growing three shoots in each test vessel:
 - Test Design A: one shoot per pot and three pots per vessel.
 - Test Design B: three shoots per pot and one pot per vessel.
 - Alternative test designs of one shoot per pot per test vessel are acceptable provided that replication is adjusted as required to achieve the required validity criteria.
33. The individual test vessels should be randomly assigned to the treatment groups. A randomised design for the location of the test vessels in the test area is required to minimise the influence of spatial differences in light intensity or temperature.

Test chemical concentrations and control groups

34. Concentrations should typically follow a geometric series; the separation factor between test concentrations should not exceed 3,2. Prior knowledge of the toxicity of the test chemical from a range-finding test will help to select suitable test concentrations.
35. To determine an EC_x , test concentrations should bracket the EC_x to ensure an appropriate level of confidence. For example, if estimating the EC_{50} , the highest test concentration should be greater than the EC_{50} value. If the EC_{50} value lies outside of the range of test concentrations, associated confidence intervals will be large and a proper assessment of the statistical fit of the model may not be possible. The use of more test concentrations will improve the confidence interval around the resulting EC_x value.
36. To determine the LOEC/NOEC (optional endpoint), the lowest test concentration should be sufficiently low such that growth is not significantly different from growth in control plants. In addition, the highest test concentration should be sufficiently high such that growth is significantly lower than that in the control. The use of more replicates will enhance the statistical power of the no effect-concentration/ ANOVA design.

Limit test

37. In cases where a range-finding test indicates that the test chemical does not have an adverse effect at concentrations up to 100 mg/l or up to its limit of solubility in the test medium, or in the case of a formulation up to the limit of dispersibility, a limit test may be undertaken to facilitate comparison of responses in a control group and one treatment group — 100 mg/l or a concentration equal to the limit of solubility, or 1 000 mg/kg dry sediment. This test should follow the general principles of a standard dose-response test, with the exception that an increase in the minimum number of replicates to six test vessels per control and concentration is advised. Growth in the control and treatment group may be analysed using a statistical test to compare means, e.g. a Student's t-test.

Test solutions

38. Test solutions are usually prepared by dilution of a stock solution, prepared by dissolving or dispersing the test chemical in Smart and Barko media, using demineralised (i.e. distilled or deionised) water (see Appendix 1).

39. The highest test concentration should normally not exceed the water solubility of the test chemical or, in the case of formulations, the dispersibility under the test conditions.
40. For test chemicals of low water solubility, it may be necessary to prepare a concentrated stock solution or dispersion of the chemical using an organic solvent or dispersant in order to facilitate the addition of accurate quantities of the test chemical to the test medium and aid in its dispersion and dissolution. Every effort should be made to avoid the use of such solvents or dispersants. There should be no phytotoxicity resulting from the use of auxiliary solvents or dispersants. For example, commonly used solvents, which do not cause phytotoxicity at concentrations up to 100 µl/l, include acetone and dimethylformamide. If a solvent or dispersant is used, its final concentration should be reported and kept to a minimum (≤ 100 µl/l). Under these circumstances all treatments and (solvent) controls should contain the same concentration of solvent or dispersant. Untreated control replicates that do not contain a solvent or dispersant are also incorporated into the test design. Further guidance on the use of dispersants is given in an OECD Guidance Document (11).

TEST PROCEDURE

41. The test procedure varies according to the application route of the test chemical (i.e. via the water or sediment phase). The likely behaviour of the test chemical in a water-sediment system should be considered to inform the choice of exposure regime used in the test (i.e. static or static renewal, spiked water or spiked sediment). Spiked sediment tests may be preferred in some cases for chemicals that are predicted to significantly partition to sediment.

Establishment phase

42. Healthy shoot apices/tips, i.e. without side shoots, are cut from the culture plants to give a shoot length of 6 cm (± 1 cm). For Test Design A (one shoot per pot and three pots per vessel) single shoot tips are planted into each pot. For Test Design B (three shoots per pot and one pot per vessel) four to five shoot apices are planted into each pot containing the sediment.
43. In both cases surplus pots should be planted to allow for selection of uniform plants at test initiation, as well as to provide spare plants to be used for inspection of root growth immediately prior to treatment and spare plants to be harvested for shoot biomass and length measurements on Day 0.
44. Shoots are inserted such that approximately three cm, covering at least two nodes, are beneath the sediment surface.
45. Pots are then transferred to test vessels under the same environmental conditions as for the exposure phase and maintained for seven days in Smart and Barko medium to induce root development.
46. After this time, several plants in spare pots should be removed for inspection of root growth. If root growth is not visible (i.e. root tips are not visible), then the establishment phase should be extended until root growth is visible. This step is recommended to ensure that plants are actively growing at the time of test initiation.

Selection of uniform plant material

47. For Test Design A (one shoot per pot and three pots per vessel) pots are selected for uniformity prior to test initiation. For Test Design B (three shoots per pot and one pot per vessel), surplus plants are removed to leave three plants that are uniform in size and appearance.

Exposure via the water phase

48. Pots, selected for uniformity, are placed into the test vessels as required for the experimental design. Smart and Barko medium is then added to the test vessels. Care should be taken to avoid disturbance of the sediment. For this purpose, media may be added using a funnel or a plastic disc to cover the sediment while the medium is poured into the test vessels provided that the disc is removed immediately afterwards. Alternatively, plant pots may be placed in the test vessels after the addition of the media. In both cases, fresh media may be used at the beginning of the exposure phase, if necessary to minimise the potential build-up of algae and bacteria or to allow preparation of single batches of test solution across replicates.
49. The shoot length above sediment is measured, either prior to or after the addition of the medium.
50. The relevant amounts of the test chemical may be added to the test medium before it is added to the test vessels. Alternatively, the test chemical may be introduced into the medium after it has been added to the test vessels. In this case, care should be taken to ensure that the test chemical is homogeneously distributed throughout the test system without disturbing the sediment.
51. In all cases, the appearance (e.g. clear, cloudy, etc.) of the test media is recorded at test initiation.

Exposure via sediment

52. Spiked sediments of the chosen concentration are prepared by addition of a solution of the test chemical directly to fresh sediment. A stock solution of the test chemical dissolved in deionised water is mixed with the formulated sediment by rolling mill, feed mixer or hand mixing. If poorly soluble in water, the test chemical can be dissolved in as small a volume as possible of a suitable organic solvent (e.g. hexane, acetone or chloroform). This solution is then mixed with ca. 10 g of fine quartz sand for one test vessel. The solvent is allowed to evaporate and the sand is then mixed with the suitable amount of sediment per test beaker. Only agents that volatilise readily can be used to solubilise, disperse or emulsify the test chemical. It should be borne in mind that the volume/weight of sand spiked with the test chemical has to be taken into account in the final preparation of the sediment (i.e. the sediment should thus be prepared with less sand). Care should be taken to ensure that the test chemical added to sediment is thoroughly and evenly distributed within the sediment.
53. The spiked sediment is filled into the pots (as described above). Plants, selected for uniformity and an adequate root system, are removed from the pots used during the establishment phase and transplanted into the spiked sediment as described above.
54. Pots are placed into the test vessels as required for the experimental design. Smart and Barko medium is then added carefully (i.e. using a funnel) in order to avoid disturbance of the sediment. The shoot length above sediment is measured, either prior to or after the addition of the media.

Maintenance of water levels over the test duration

55. The final water volume must be recorded and the water level marked on each test vessel. If water evaporates during the test by more than 10 %, the water level should be adjusted with distilled water. If necessary, beakers may be loosely covered by a transparent cover such as transparent plastic lids to minimise evaporation and contamination with algal spores.

Test conditions

56. Warm and/or cool white fluorescent lighting are used to provide light irradiance in the range of about 140 (± 20) $\mu\text{E}\cdot\text{m}^{-2}\text{ s}^{-1}$ when measured as a photosynthetically active radiation (400-700 nm) at the water surface and using a light:dark ratio of 16:8 h. Any differences from the selected light irradiance over the test area should not exceed the range of ± 15 %.

57. The temperature in the test vessels is 20 ± 2 °C.
58. The pH of the control medium should not increase by more than 1,5 units during the test. However, deviation of more than 1,5 units would not invalidate the test when it can be shown that the validity criteria specified previously are met.

Test duration

59. The exposure period is 14 days.

Measurements and analytical determinations

60. After the establishment phase and immediately prior to treatment (i.e. on Day 0), spare plants from five randomly selected pots for the three plants per pot design or 15 pots for the one plant per pot design, are harvested for assessment of shoot length and fresh and dry weight as described below.
61. For plants transferred into the exposure phase, the following assessments are made as shown in Table 1:
- Assessments of main shoot length, side shoot number and side shoot length are recorded at least at the end of the exposure period (e.g. on day 14).
 - Visual assessments of plant health are recorded at least three times during the exposure period (e.g. on days 0, 7 and 14).
 - Assessments of shoot fresh weight and dry weight are made at the end of the test (i.e. on Day 14).
62. Shoot length is determined using a ruler. If side shoots are present, their numbers and length should also be measured.
63. Visual assessments of plant health are made by recording the appearance of plants and the general condition of the test medium. Observations to be noted include:
- Necrosis, chlorosis or other discoloration such as excessive reddening relative to control plants.
 - Development of bacterial or algal contamination;
 - Growth abnormalities such as stunting, altered internodal length, distorted shoots/leaves, the proliferation of side shoots, leaf loss, loss of turgor and stem fragmentation.
 - Visual assessments of root health are made at test termination, by carefully washing sediment from roots to enable observation of the root system. A proposed scale for assessment, relative to control plants, is shown below:
- (1) roots absent
 - (2) few roots
 - (3) moderate root development
 - (4) very good root development, similar to controls.
64. Assessments of fresh weight are made at the beginning and end of the test by cutting the shoot at sediment level and then blotting dry prior to weighing. Care should be taken to remove sediment particles that may adhere to the base of the shoot. Shoot material is then placed in a drying oven at ca. 60 °C and dried to a constant weight, prior to re-weighing and recording the dry weight.
65. A summary of the minimum biological assessments required over the test duration is provided in Table 1.

Table 1

Assessment schedule

Day after treatment (DAT)	<i>Myriophyllum spicatum</i>			
	Shoot length, side shoot length and number	Visual assessment of shoots	Shoot fresh and dry weight, Visual assessment of roots	pH O ₂
0	A	A	A	A
4	—	—	—	—
7	—	A	—	A
14	A	A	A	A

A: indicates that assessments are required on these occasions

—: indicates that measurements are not required.

Frequency of measurements and analytical determinations

66. The temperature of the medium in a supplementary vessel held under the same conditions in the growth chamber, incubator or room should be recorded at least daily (or continuously with a data logger).
67. The pH and dissolved oxygen concentration of the test medium should be checked at test initiation, at least once during the study and at the end of the study in all replicate vessels. On each occasion, measurements should be taken at the same time of the day. If bulk solutions are used to prepare all replicates at each test concentration, then a single measurement of each bulk solution is acceptable on Day 0.
68. Irradiance should be measured in the growth chamber, incubator or room at points equivalent to level of the water surface. Measurements should be made at least once at test initiation or during the test. The method of light detection and measurement, in particular the type of sensor, will affect the measured value. Spherical sensors (which respond to light from all angles above and below the plane of measurement) and “cosine” sensors (which respond to light from all angles above the plane of measurement) are preferred to unidirectional sensors, and will give higher readings for a multi-point light source of the type described here.

Analytical measurements of test chemical

69. The correct application of the test chemical should be supported by analytical measurements of test chemical concentrations.
70. Water samples should be collected for test chemical analysis shortly after test initiation (i.e. on the day of application for stable test chemicals or one hour after application for chemicals that are not stable) and at test termination for all test concentrations.
71. Concentrations in sediment and sediment pore-water should be determined at test initiation and test termination, at least in the highest test concentration, unless the test chemicals are known to be stable in water (> 80 % of nominal). Measurements in sediment and pore-water might not be necessary if the partitioning of the test chemical between water and sediment has been clearly determined in a water/sediment study under comparable conditions (e.g. sediment to water ratio, application method, sediment type).

72. Sampling of sediment at test initiation is likely to disrupt the test system. Hence, additional treated test vessels may be required to facilitate analytical determinations at test initiation and test termination. Similarly, where intermediate assessments are considered necessary, i.e. on day 7, and analyses require large samples of sediment that cannot be easily removed from the test system, analytical determinations should be performed using additional test vessels treated in the same way as those used for biological assessments.
73. Centrifugation at, for example, 10 000 g and 4 °C for 30 minutes is recommended to isolate interstitial water. However, if the test chemical is demonstrated not to absorb to filters, filtration may also be acceptable. In some cases, it might not be possible to analyse concentrations in the pore water if the sample size is too small.
74. In semi-static tests (i.e. exposure via the water phase) where the concentration of the relevant test chemical(s) is not expected to remain within 20 % of the nominal concentration over the test duration without renewal of test solutions, used and freshly prepared test solutions should be sampled for analyses of test chemical concentration at each renewal.
75. In cases where the measured initial concentration of the test chemical is not within 20 % of nominal but where sufficient evidence can be provided to show that the initial concentrations are repeatable and stable (i.e. within the range of 80-120 % of the initial concentration), chemical determinations may be carried out on only the highest and lowest test concentrations.
76. In all cases, determination of test chemical concentrations need only be performed on one replicate vessel at each test concentration. Alternatively, the test solutions of all replicates for each concentration may be pooled for analyses.
77. If there is evidence that the test chemical concentration has been maintained within 20 % of the nominal or measured initial concentration throughout the test, then analysis of the results and subsequent derivation of endpoints can be based on nominal or measured initial values.
78. In these cases, effect concentrations should be based on nominal or measured water concentrations at the beginning of the test.
79. However, if there is evidence that the concentration has declined (i.e. is not maintained within 20 % of the nominal or measured initial concentration in the treated compartment) throughout the test, then analysis of the results should be based on the geometric mean concentration during exposure or models describing the decline of the concentration of the test chemical in the treated compartment (11).

DATA EVALUATION

80. In cases where use of a solvent / dispersant is required, data from solvent and untreated controls may be pooled for the purposes of statistical analyses provided that the responses of the solvent and untreated controls are not statistically significantly different.

Response variables

81. The purpose of the test is to determine the effects of the test chemical on the vegetative growth of the test species, using two response variables, average specific growth rate and yield, as follows:

Average specific growth rate

82. This response variable is based on changes in the logarithms of total shoot length, total shoot fresh weight and total shoot dry weight, over time in the controls and each treatment group. This variable is calculated for each replicate of each control and treatment group. The mean length and weight of the three plants per test vessel (replicate) and, subsequently, the growth rate for each replicate, should be calculated using the following formula:

$$\mu_{i-j} = \frac{\ln(N_j) - \ln(N_i)}{t}$$

where:

μ_{i-j} : average specific growth rate from time i to j

N_i : measurement variable in the test or control vessel at time i

N_j : measurement variable in the test or control vessel at time j

t: time period from i to j

83. From the replicate responses, a mean value for growth rate along with variance estimates should be calculated for each treatment and control group.
84. The average specific growth rate should be calculated for the entire test period (time “i” in the above formula is the beginning of the test and time “j” is the end of the test). For each test concentration and control, calculate a mean value for average specific growth rate along with the variance estimates.
85. Percent inhibition of growth rate (I_r) may then be calculated for each test concentration (treatment group) according to the following formula:

$$\%I_r = \frac{(\mu_c - \mu_r)}{\mu_c} \times 100$$

where:

% I_r : percent inhibition in average specific growth rate

μ_c : mean value for μ in the control

μ_r : mean value for μ in the treatment group

Yield

86. This response variable is based on changes in total shoot length, total shoot fresh weight and total shoot dry weight, over time in the controls and each treatment group. The mean percent inhibition in yield (% I_y) may be calculated for each treatment group as follows:

$$\%I_y = \frac{(b_c - b_r)}{b_c}$$

where:

% I_y : percent reduction in yield

b_c : final biomass minus starting biomass for the control group

b_r : final biomass minus starting biomass in the treatment group

Plotting concentration-response curves

87. Concentration-response curves relating mean percentage inhibition of the response variable (I_r or I_y), calculated as shown above and the log concentration of the test chemical should be plotted.

EC_x estimation

88. Estimates of the EC_x (e.g. EC₅₀) should be based upon both average specific growth rate (E_rC_x) and yield (E_yC_x), each of which should in turn be based upon total shoot fresh weight, total shoot dry weight and total shoot length.
89. It should be noted that EC_x values calculated using these two response variables are not comparable and this difference is recognised when using the results of the test. EC_x values based upon average specific growth rate (E_rC_x) will in most cases be higher than results based upon yield (E_yC_x) — if the test conditions of this test method are adhered to — due to the mathematical basis of the respective approaches. This difference should not be interpreted as a difference in sensitivity between the two response variables, simply the values are different mathematically.

Statistical procedures

90. The aim is to obtain a quantitative concentration-response relationship by regression analysis. It is possible to use a weighted linear regression after having performed a linearising transformation of the response data, for instance into probit or logit or Weibull units (13), but non-linear regression procedures are preferred techniques that better handle unavoidable data irregularities and deviations from smooth distributions. Approaching either zero or total inhibition such irregularities may be magnified by the transformation, interfering with the analysis (13). It should be noted that standard methods of analysis using probit, logit, or Weibull transforms are intended for use on quantal (e.g. mortality or survival) data, and should be modified to accommodate growth rate or yield data. Specific procedures for determination of EC_x values from continuous data can be found in (14) (15) (16) (17).
91. For each response variable to be analysed, use the concentration-response relationship to calculate point estimates of EC_x values. The 95 % confidence limits for each estimate are determined and goodness of fit of the response data to the regression model should be assessed either graphically or statistically. Regression analysis should be performed using individual replicate responses, not treatment group means.
92. EC₅₀ estimates and confidence limits may also be obtained using linear interpolation with bootstrapping (18), if available regression models/methods are unsuitable for the data.
93. For estimation of the LOEC and hence the NOEC, it is necessary to compare treatment means using analysis of variance (ANOVA) techniques. The mean for each concentration is then compared with the control mean using an appropriate test method (e.g. Dunnett's, Williams' tests) (19) (20) (21) (22). It is necessary to assess whether the ANOVA assumption of normal distribution (ND) and variance homogeneity (VH) of variance holds. This assessment should be performed by Shapiro-Wilks-test (ND) or Levene's test (VH). Failure to meet the assumption of ND and homogeneity of variances can sometimes be corrected by logarithmic transformation of the data. If heterogeneity of variance and/or deviation from ND is extreme and cannot be corrected by transformation, analysis by methods such as Bonferroni-Welch-t-test, step-down Jonkheere Terpstra test and Bonferroni-Median-Test should be considered. Additional guidance on determining the NOEC can be found in (16).

REPORTING

94. The test report includes the following details:

Test chemical

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 species

- scientific name and source.

Test conditions

- duration and conditions of establishment phase;
- test procedure used (static, semi-static, pulsed);
- date of start of the test and its duration;
- test medium, *i.e.* sediment and liquid nutrient medium;
- description of the experimental design: growth chamber/room or laboratory, test vessels and covers, solution volumes, length and weight of test plants per test vessel at the beginning of the test, ratio of sediment surface to water surface, sediment and water volume ratio;
- test concentrations (nominal and measured as appropriate) and number of replicates per concentration;
- methods of preparation of stock and test solutions including the use of any solvents or dispersants;
- temperature during the test;
- light source, irradiance ($\mu\text{E}\cdot\text{m}^{-2}\text{ s}^{-1}$)
- pH values of the test and control media as well as appearance of test media at test initiation and end;
- oxygen concentrations;
- the method of analysis with appropriate quality assessment data (validation studies, standard deviations or confidence limits of analyses);
- methods for determination of measurement variables, *e.g.*, length, dry weight, fresh weight;
- all deviations from this test method.

Results

- raw data: shoot length and shoot weight of plants/pot and other measurement variables in each test and control vessel at each observation and occasion of analysis according to the assessment schedule provided in Table 1;
- means and standard deviations for each measurement variable;
- growth curves for each concentration;
- doubling time/growth rate in the control based on shoot length and fresh weight including the coefficient of variation for yield of fresh weight;
- calculated response variables for each treatment replicate, with mean values and coefficient of variation for replicates;
- graphical representation of the concentration/effect relationship;
- estimates of toxic endpoints for response variables *e.g.* EC_{50} , and associated confidence intervals. If calculated, LOEC and/or NOEC and the statistical methods used for their determination;

- if ANOVA has been used, the size of the effect which can be detected (e.g. the minimum significant difference);
- any stimulation of growth found in any treatment;
- any visual signs of phytotoxicity as well as observations of test solutions;
- discussion of the results, including any influence on the outcome of the test resulting from deviations from this test method.

LITERATURE:

- (1) Chapter C.26 of this Annex: *Lemna* sp. Growth Inhibition Test.
- (2) Chapter C.3 of this Annex: Freshwater Alga and Cyanobacteria, Growth Inhibition Test.
- (3) Maltby, L. et al. (2010), Aquatic Macrophyte Risk Assessment for Pesticides, Guidance from the AMRAP Workshop in Wageningen (NL), 14-16 January 2008.
- (4) Arts, G.H.P. et al. (2008), Sensitivity of submersed freshwater macrophytes and endpoints in laboratory toxicity tests, *Environmental Pollution*, Vol. 153, pp. 199-206.
- (5) ISO 16191:2013 Water quality — Determination of the toxic effect of sediment on the growth behaviour of *Myriophyllum aquaticum*.
- (6) Knauer, K. et al. (2006), Methods for assessing the toxicity of herbicides to submersed aquatic plants, *Pest Management Science*, Vol. 62/8, pp. 715-722.
- (7) Chapter C.50 of this Annex: Sediment-free *Myriophyllum spicatum* Toxicity Test.
- (8) Chapter C.28 of this Annex: Sediment-Water Chironomid Toxicity Using Spiked Water.
- (9) Ratte, M., H. Ratte (2014), "Myriophyllum Toxicity Test: Result of a ring test using *M. aquaticum* and *M. spicatum* grown in a water-sediment system", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 206, OECD Publishing, Paris.
- (10) Davies, J. et al. (2003), Herbicide risk assessment for non-target aquatic plants: sulfosulfuron — a case study, *Pest Management Science*, Vol. 59/2, pp. 231 - 237.
- (11) OECD (2000), "Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 23, OECD Publishing, Paris.
- (12) Smart, R.M., J.W. Barko (1985), Laboratory culture of submersed freshwater macrophytes on natural sediments, *Aquatic Botany*, Vol. 21/3, pp. 251-263.
- (13) Christensen, E.R., N. Nyholm (1984), Ecotoxicological Assays with Algae: Weibull Dose-Response Curves, *Environmental Science Technology*, Vol. 18/9, pp. 713-718.
- (14) Nyholm, N. et al. (1992), Statistical treatment of data from microbial toxicity tests, *Environmental Toxicology and Chemistry*, Vol. 11/2, pp. 157-167.
- (15) Bruce, R.D., D.J. Versteeg (1992), A statistical procedure for modelling continuous toxicity data, *Environmental Toxicology and Chemistry*, Vol. 11/10, 1485-1494.
- (16) OECD (2006), "Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application", OECD Environment, Health and Safety Publications (EHS), Series on Testing and Assessment, No. 54, OECD Publishing, Paris.
- (17) Brain, P., R. Cousens (1989), An equation to describe dose-responses where there is stimulation of growth at low doses, *Weed Research*, Vol. 29/2, pp/ 93-96.

-
- (18) Norberg-King, T.J. (1988), An interpolation estimate for chronic toxicity: The ICp approach, National Effluent Toxicity Assessment Center Technical Report 05-88. US EPA, Duluth, MN.
- (19) Dunnett, C.W. (1955), A multiple comparisons procedure for comparing several treatments with a control, Journal of the American Statistical Association, Vol. 50/272, pp. 1096-1121.
- (20) Dunnett, C.W. (1964), New tables for multiple comparisons with a control, Biometrics, Vol. 20/3, pp. 482-491.
- (21) Williams, D.A. (1971), A test for differences between treatment means when several dose levels are compared with a zero dose control, Biometrics, Vol. 27/1, pp. 103-117.
- (22) Williams, D.A. (1972), The comparison of several dose levels with a zero dose control, Biometrics, Vol. 28/2, pp. 519-531.
-

*Appendix 1***SMART AND BARKO MEDIUM COMPOSITION**

Component	Amount of reagent added to water (*) (mg/l)
$\text{CaCl}_2 \cdot 2 \text{H}_2\text{O}$	91,7
$\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$	69,0
NaHCO_3	58,4
KHCO_3	15,4
pH (air equilibrium)	7,9
(*) Demineralised (i.e. distilled or deionised) water.	

Appendix 2

DEFINITIONS

Biomass is the fresh and/or dry weight of living matter present in a population. In this test the biomass is the sum of main shoot, all lateral branches and all roots.

Chemical is a substance or a mixture.

Chlorosis is the change of the color from green to yellowing of test organism especially of the whorls.

EC_x is the concentration of the test chemical dissolved in test medium that results in a x % (e.g. 50 %) reduction in growth of *Myriophyllum spicatum* within a stated exposure period (to be mentioned explicitly if deviating from full or normal test duration). To unambiguously denote an EC value deriving from growth rate or yield the symbol " E_rC " is used for growth rate and " E_yC " is used for yield, followed by the measurement variable used, e.g. E_rC (main shoot length).

Growth is an increase in the measurement variable, e.g. main shoot length, total lateral branches length, total shoot length, total root length, fresh weight, dry weight or number of whorls, over the test period.

Growth rate (average specific growth rate) is the logarithmic increase in the measurement variable during the exposure period. *Note:* Growth rate related response variables are independent of the duration of the test as long as the growth pattern of unexposed control organisms is exponential.

Lowest Observed Effect Concentration (LOEC) is the lowest tested concentration at which the chemical is observed to have a statistically significant reducing effect on growth (at $p < 0,05$) when compared with the control, within a given exposure time. 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.

Measurement variables are any type of variables which are measured to express the test endpoint using one or more different response variables. In this test method main shoot length, total lateral branches length; total shoot length, total root length, fresh weight, dry weight and number of whorls are measurement variables.

Monoculture is a culture with one plant species.

Necrosis is dead (i.e. white or dark brown) tissue of the test organism.

No Observed Effect Concentration (NOEC) is the test concentration immediately below the LOEC.

Response variable is a variable for the estimation of toxicity derived from any measured variable describing biomass by different methods of calculation. For this test method growth rate and yield are response variables derived from measurement variables like main shoot length, total shoot length, fresh weight, dry weight, or number of whorls.

Semi-static (renewal) test is a test in which the test solution is periodically replaced at specific intervals during the test.

Static test is a test method without renewal of the test solution during the test.

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

Test endpoint describes the general factor that will be changed relative to control by the test chemical as aim of the test. In this test method the test endpoint is inhibition of growth which may be expressed by different response variables which are based on one or more measurement variables.

Test medium is the complete synthetic growth medium on which test plants grow when exposed to the test chemical. The test chemical will normally be dissolved in the test medium.

UVCB is a substance of unknown or variable composition, complex reaction product or biological material

Yield is value of a measurement variable to express biomass at the end of the exposure period minus the measurement variable at the start of the exposure period. *Note:* When the growth pattern of unexposed organisms is exponential, yield-based response variables will decrease with the test duration.'
