070182/EU XXIV.GP
Eingelangt am 20/01/12

COUNCIL OF THE EUROPEAN UNION

Brussels, 20 January 2012

5596/12
ADD 1

COMPET 31
ENV 35
CHIMIE 3
MI 37
ENT 15

COVER NOTE

from:	European Commission
date of receipt:	19 January 2012
to:	General Secretariat of the Council
No Cion doc.:	D018298/01 Annex
Subject:	Annex to Draft Commission Regulation (EU) No … amending, for the purpose of its adaptation to technical progress, 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)

Delegations will find attached Commission document D018298/01 Annex.

_____

Encl.: D018298/01 Annex

ANNEX

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

1. Chapter B.42 is replaced by the following:

"B.42 Skin Sensitisation: Local Lymph Node Assay

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals and EU Test Methods based on them are periodically reviewed in light of scientific progress, changing regulatory needs, and animal welfare considerations. The original Test Method (TM) for the determination of skin sensitisation in the mouse, the Local Lymph Node Assay (LLNA; OECD Test Guideline 429; chapter B.42 of this Annex) was adopted previously (1). The details of the validation of the LLNA and a review of the associated work have been published (2) (3) (4) (5) (6) (7) (8) (9) (10) (11). The updated LLNA is based on the evaluation of experience and scientific data (12). This is the second TM to be designed for assessing skin sensitisation potential of chemicals (substances and mixtures) in animals. The other TM (*i.e.* OECD Test Guideline 406; chapter B.6 of this Annex) utilises guinea pig tests, notably the guinea pig maximisation test and the Buehler test (13). The LLNA provides advantages over B.6 and OECD Test Guideline 406 (13) with regard to animal welfare. This updated LLNA TM includes a set of Performance Standards (PS) (Appendix 1) that can be used to evaluate the validation status of new and/or modified test methods that are functionally and mechanistically similar to the LLNA, in accordance with the principles of OECD Guidance Document No. 34 (14).

2. The LLNA studies the induction phase of skin sensitisation and provides quantitative data suitable for dose-response assessment. It should be noted that the mild/moderate sensitisers which are recommended as suitable positive control chemicals (PC) for guinea pig test methods (*i.e.* B. 6; OECD Test Guideline 406) (13) are also appropriate for use with the LLNA (6) (8) (15). A reduced LLNA (rLLNA) approach, which could use up to 40% fewer animals is also described as an option in this TM (16) (17) (18). The rLLNA may be used when there is a regulatory need to confirm a negative prediction of skin sensitising potential, provided there is adherence to all other LLNA protocol specifications, as described in this TM. Prediction of a negative outcome should be made based on all available information as described in paragraph 4. Before applying the rLLNA approach, clear justifications and scientific rationale for its use should be provided. If, against expectations, a positive or equivocal result is obtained in the rLLNA, additional testing may be needed in order to interpret or clarify the finding. The rLLNA should not be used for the hazard identification of skin sensitising test substances when dose-response information is needed such as sub-categorisation for Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures and UN Globally Harmonized System of Classification and Labelling of Chemicals.

DEFINITIONS

3. Definitions used are provided in Appendix 2.

INITIAL CONSIDERATIONS AND LIMITATIONS

4. The LLNA provides an alternative method for identifying potential skin sensitising chemicals. This does not necessarily imply that in all instances the LLNA should be used in place of guinea pig tests (*i.e.* B.6; OECD Test Guideline 406) (13), but rather that the assay is of equal merit and may be employed as an alternative in which positive and negative results generally no longer require further confirmation. The testing laboratory should consider all available information on the test substance prior to conducting the study. Such information will include the identity and chemical structure of the test substance; its physicochemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the test substance; and toxicological data on structurally related chemicals. This information should be considered in order to determine whether the LLNA is appropriate for the substance (given the incompatibility of limited types of chemicals with the LLNA- see paragraph 5) and to aid in dose selection.

5. The LLNA is an *in vivo* method and, as a consequence, will not eliminate the use of animals in the assessment of allergic contact sensitising activity. It has, however, the potential to reduce the number of animals required for this purpose. Moreover, the LLNA offers a substantial refinement (less pain and distress) of the way in which animals are used for allergic contact sensitisation testing. The LLNA is based upon consideration of immunological events stimulated by chemicals during the induction phase of sensitisation. Unlike guinea pig tests (*i.e.* B.6; OECD Test Guideline 406) (13) the LLNA does not require that challenge-induced dermal hypersensitivity reactions be elicited. Furthermore, the LLNA does not require the use of an adjuvant, as is the case for the guinea pig maximisation test (13). Thus, the LLNA reduces animal pain and distress. Despite the advantages of the LLNA over B.6 and OECD Test Guideline 406, it should be recognised that there are certain limitations that may necessitate the use of B.6 or OECD Test Guideline 406 (13) (*e.g.* false negative findings in the LLNA with certain metals, false positive findings with certain skin irritants [such as some surfactant type chemicals] (19) (20), or solubility of the test substance). In addition, chemical classes or substances containing functional groups shown to act as potential confounders (21) may necessitate the use of guinea pig tests (*i.e.* B.6; OECD Test Guideline 406) (13). Further, based on the limited validation database, which consisted primarily of pesticide formulations, the LLNA is more likely than the guinea pig test to yield a positive result for these types of test substances (22). However, when testing formulations, one could consider including similar substances with known results as benchmark substances to demonstrate that the LLNA is functioning properly (see paragraph 16). Other than such identified limitations, the LLNA should be applicable for testing any substances unless there are properties associated with these substances that may interfere with the accuracy of the LLNA.

PRINCIPLE OF THE TEST

6. The basic principle underlying the LLNA is that sensitisers induce proliferation of lymphocytes in the lymph nodes draining the site of test substance application. This proliferation is proportional to the dose and to the potency of the applied allergen and provides a simple means of obtaining a quantitative measurement of sensitisation. Proliferation is measured by comparing the mean proliferation in each test group to the mean proliferation in the vehicle treated control (VC) group. The ratio of the mean proliferation in each treated group to that in the concurrent VC group, termed the Stimulation Index (SI), is determined, and should be ≥ 3 before classification of the test substance as a potential skin

sensitiser is warranted. The procedures described here are based on the use of *in vivo* radioactive labelling to measure an increased number of proliferating cells in the draining auricular lymph nodes. However, other endpoints for assessment of the number of proliferating cells may be employed provided the PS requirements are fully met (Appendix 1).

DESCRIPTION OF THE ASSAY

Selection of animal species

7. The mouse is the species of choice for this test. Young adult female mice of CBA/Ca or CBA/J strain, which are nulliparous and non-pregnant, are used. At the start of the study, animals should be between 8-12 weeks old, and the weight variation of the animals should be minimal and not exceed 20% of the mean weight. Alternatively, other strains and males may be used when sufficient data are generated to demonstrate that significant strain and/or gender-specific differences in the LLNA response do not exist.

Housing and feeding conditions

8. Mice should be group-housed (23), unless adequate scientific rationale for housing mice individually is provided. The temperature of the experimental animal room should be 22 ± 3ºC. Although the relative humidity should be at least 30% and preferably not exceed 70%, other than during room cleaning, the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water.

Preparation of animals

9. The animals are randomly selected, marked to permit individual identification (but not by any form of ear marking), and kept in their cages for at least five days prior to the start of dosing to allow for acclimatisation to the laboratory conditions. Prior to the start of treatment all animals are examined to ensure that they have no observable skin lesions.

Preparation of dosing solutions

10. Solid chemicals should be dissolved or suspended in solvents/vehicles and diluted, if appropriate, prior to application to an ear of the mice. Liquid chemicals may be applied neat or diluted prior to dosing. Insoluble chemicals, such as those generally seen in medical devices, should be subjected to an exaggerated extraction in an appropriate solvent to reveal all extractable constituents for testing prior to application to an ear of the mice. Test substances should be prepared daily unless stability data demonstrate the acceptability of storage.

Reliability check

11. Positive control chemicals (PC) are used to demonstrate appropriate performance of the assay by responding with adequate and reproducible sensitivity as a sensitising test substance for which the magnitude of the response is well characterised. Inclusion of a concurrent PC is recommended because it demonstrates competency of the laboratory to successfully conduct each assay and allows for an assessment of intra-, and inter-laboratory reproducibility and comparability. A PC for each study is also required by some regulatory authorities and therefore users are encouraged to consult the relevant authorities prior to conducting the LLNA. Accordingly, the routine use of a concurrent PC is encouraged to avoid

the need for additional animal testing to meet such requirements that might arise from the use of a periodic PC (see paragraph 12). The PC should produce a positive LLNA response at an exposure level expected to give an increase in the SI > 3 over the negative control (NC) group. The PC dose should be chosen such that it does not cause excessive skin irritation or systemic toxicity and the induction is reproducible but not excessive (*i.e.* a SI > 20 would be excessive). Preferred PC are 25% hexyl cinnamic aldehyde (Chemical Abstracts Service [CAS] No 101-86-0) in acetone: olive oil (4:1, v/v) and 5% mercaptobenzothiazole (CAS No 149-30-4) in *N,N*-dimethylformamide (see Appendix 1, Table 1). There may be circumstances in which, given adequate justification, other PC, meeting the above criteria, may be used.

12. While inclusion of a concurrent PC group is recommended, there may be situations in which periodic testing (*i.e.* at intervals ≤ 6 months) of the PC may be adequate for laboratories that conduct the LLNA regularly *(i.e.* conduct the LLNA at a frequency of no less than once per month) and have an established historical PC database that demonstrates the laboratory's ability to obtain reproducible and accurate results with PCs. Adequate proficiency with the LLNA can be successfully demonstrated by generating consistent positive results with the PC in at least 10 independent tests conducted within a reasonable period of time (*i.e.* less than one year).

13. A concurrent PC group should always be included when there is a procedural change to the LLNA (*e.g.* change in trained personnel, change in test method materials and/or reagents, change in test method equipment, change in source of test animals), and such changes should be documented in laboratory reports. Consideration should be given to the impact of these changes on the adequacy of the previously established historical database in determining the necessity for establishing a new historical database to document consistency in the PC results.

14. Investigators should be aware that the decision to conduct a PC study on a periodic basis instead of concurrently has ramifications on the adequacy and acceptability of negative study results generated without a concurrent PC during the interval between each periodic PC study. For example, if a false negative result is obtained in the periodic PC study, negative test substance results obtained in the interval between the last acceptable periodic PC study and the unacceptable periodic PC study may be questioned. Implications of these outcomes should be carefully considered when determining whether to include concurrent PCs or to only conduct periodic PCs. Consideration should also be given to using fewer animals in the concurrent PC group when this is scientifically justified and if the laboratory demonstrates, based on laboratory-specific historical data, that fewer mice can be used (12).

15. Although the PC should be tested in the vehicle that is known to elicit a consistent response (*e.g.* acetone: olive oil; 4:1, v/v), there may be certain regulatory situations in which testing in a non-standard vehicle (clinically/chemically relevant formulation) will also be necessary (24). If the concurrent PC is tested in a different vehicle than the test substance, then a separate VC for the concurrent PC should be included.

16. In instances where test substances of a specific chemical class or range of responses are being evaluated, benchmark substances may also be useful to demonstrate that the test method is functioning properly for detecting the skin sensitisation potential of these types of test substances. Appropriate benchmark substances should have the following properties:

- structural and functional similarity to the class of the test substance being tested;
- known physical/chemical characteristics;

- supporting data from the LLNA;
- supporting data from other animal models and/or from humans.

TEST PROCEDURE

Number of animals and dose levels

17. A minimum of four animals is used per dose group, with a minimum of three concentrations of the test substance, plus a concurrent NC group treated only with the vehicle for the test substance, and a PC (concurrent or recent, based on laboratory policy in considering paragraphs 11-14). Testing multiple doses of the PC should be considered, especially when testing the PC on an intermittent basis. Except for absence of treatment with the test substance, animals in the control groups should be handled and treated in a manner identical to that of animals in the treatment groups.

18. Dose and vehicle selection should be based on the recommendations given in references (3) and (5). Consecutive doses are normally selected from an appropriate concentration series such as 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, 0.5%, etc. Adequate scientific rationale should accompany the selection of the concentration series used. All existing toxicological information (*e.g.* acute toxicity and dermal irritation) and structural and physicochemical information on the test substance of interest (and/or structurally related substances) should be considered where available, in selecting the three consecutive concentrations so that the highest concentration maximises exposure while avoiding systemic toxicity and/or excessive local skin irritation (3) (25). In the absence of such information, an initial pre-screen test may be necessary (see paragraphs 21-24).

19. The vehicle should not interfere with or bias the test result and should be selected on the basis of maximising the solubility in order to obtain the highest concentration achievable while producing a solution/suspension suitable for application of the test substance. Recommended vehicles are acetone: olive oil (4:1, v/v), *N,N*-dimethylformamide, methyl ethyl ketone, propylene glycol, and dimethyl sulphoxide (19) but others may be used if sufficient scientific rationale is provided. In certain situations it may be necessary to use a clinically relevant solvent or the commercial formulation in which the test substance is marketed as an additional control. Particular care should be taken to ensure that hydrophilic test substances are incorporated into a vehicle system, which wets the skin and does not immediately run off, by incorporation of appropriate solubilisers (*e.g.* 1% Pluronic® L92). Thus, wholly aqueous vehicles are to be avoided.

20. The processing of lymph nodes from individual mice allows for the assessment of inter-animal variability and a statistical comparison of the difference between test substance and VC group measurements (see paragraph 35). In addition, evaluating the possibility of reducing the number of mice in the PC group is feasible when individual animal data are collected (12). Further, some regulatory authorities require the collection of individual animal data. Nonetheless, pooled animal data may be considered acceptable by some regulatory authorities and in such situations, users may have the option of collecting either individual or pooled animal data.

Pre-screen test

21. In the absence of information to determine the highest dose to be tested (see paragraph 18), a pre-screen test should be performed in order to define the appropriate dose level to test in the LLNA. The purpose of the pre-screen test is to provide guidance for selecting the maximum dose level to use in the main LLNA study, where information on the concentration that induces systemic toxicity (see paragraph 24) and/or excessive local skin irritation (see paragraph 23) is not available. The maximum dose level tested should be 100% of the test substance for liquids or the maximum possible concentration for solids or suspensions.

22. The pre-screen test is conducted under conditions identical to the main LLNA study, except there is no assessment of lymph node proliferation and fewer animals per dose group can be used. One or two animals per dose group are suggested. All mice will be observed daily for any clinical signs of systemic toxicity or local irritation at the application site. Body weights are recorded pre-test and prior to termination (Day 6). Both ears of each mouse are observed for erythema and scored using Table 1 (25). Ear thickness measurements are taken using a thickness gauge (*e.g.* digital micrometer or Peacock Dial thickness gauge) on Day 1 (pre-dose), Day 3 (approximately 48 hours after the first dose), and Day 6. Additionally, on Day 6, ear thickness could be determined by ear punch weight determinations, which should be performed after the animals are humanely killed. Excessive local skin irritation is indicated by an erythema score ≥3 and/or an increase in ear thickness of ≥25% on any day of measurement (26) (27). The highest dose selected for the main LLNA study will be the next lower dose in the pre-screen concentration series (see paragraph 18) that does not induce systemic toxicity and/or excessive local skin irritation.

Table 1: Erythema Scores

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

23. In addition to a 25% increase in ear thickness (26) (27), a statistically significant increase in ear thickness in the treated mice compared to control mice has also been used to identify irritants in the LLNA (28) (29) (30) (31) (32) (33) (34). However, while statistically significant increases can occur when ear thickness is less than 25% they have not been associated specifically with excessive irritation (30) (32) (33) (34)

24. The following clinical observations may indicate systemic toxicity (35) (36) when used as part of an integrated assessment and therefore may indicate the maximum dose level

to use in the main LLNA: changes in nervous system function (*e.g.* pilo-erection, ataxia, tremors, and convulsions); changes in behaviour (*e.g.* aggressiveness, change in grooming activity, marked change in activity level); changes in respiratory patterns (*i.e.* changes in frequency and intensity of breathing such as dyspnea, gasping, and rales), and changes in food and water consumption. In addition, signs of lethargy and/or unresponsiveness and any clinical signs of more than slight or momentary pain and distress, or a >5% reduction in body weight from Day 1 to Day 6, and mortality should be considered in the evaluation. Moribund animals or animals obviously in pain or showing signs of severe and enduring distress should be humanely killed (37).

Main study experimental schedule

25. The experimental schedule of the assay is as follows:

- *Day 1:*

  Individually identify and record the weight of each animal and any clinical observation. Apply 25 µL of the appropriate dilution of the test substance, the vehicle alone, or the PC (concurrent or recent, based on laboratory policy in considering paragraphs 11-15), to the dorsum of each ear.

- *Days 2 and 3:*

  Repeat the application procedure carried out on Day 1.

- *Days 4 and 5:*

  No treatment.

- *Day 6:*

  Record the weight of each animal. Inject 250 µL of sterile phosphate-buffered saline (PBS) containing 20 µCi (7.4×10^5 Bq) of tritiated (^{3}H)-methyl thymidine into all test and control mice via the tail vein. Alternatively, inject 250 µL sterile PBS containing 2 µCi (7.4×10^4 Bq) of ^{125}I-iododeoxyuridine and 10^{-5}M fluorodeoxyuridine into all mice via the tail vein. Five hours (5 h) later, humanely kill the animals. Excise the draining auricular lymph nodes from each mouse ear and process together in PBS for each animal (individual animal approach); alternatively excise and pool the lymph nodes from each ear in PBS for each treatment group (pooled treatment group approach). Details and diagrams of the lymph node identification and dissection can be found in reference (12). To further monitor the local skin response in the main study, additional parameters such as scoring of ear erythema or ear thickness measurements (obtained either by using a thickness gauge, or ear punch weight determinations at necropsy) may be included in the study protocol.

Preparation of cell suspensions

26. A single-cell suspension of lymph node cells (LNC) excised bilaterally using the individual animal approach or alternatively, the pooled treatment group approach is prepared by gentle mechanical disaggregation through 200 micron-mesh stainless steel gauze or

another acceptable technique for generating a single-cell suspension. The LNC are washed twice with an excess of PBS and the DNA is precipitated with 5% trichloroacetic acid (TCA) at 4°C for 18h (3). Pellets are either re-suspended in 1mL TCA and transferred to scintillation vials containing 10mL of scintillation fluid for ^{3}H-counting, or transferred directly to gamma counting tubes for ^{125}I-counting.

Determination of cellular proliferation (incorporated radioactivity)

27. Incorporation of ^{3}H-methyl thymidine is measured by β-scintillation counting as disintegrations per minute (DPM). Incorporation of ^{125}I-iododeoxyuridine is measured by ^{125}I-counting and also is expressed as DPM. Depending on the approach used, the incorporation is expressed as DPM/mouse (individual animal approach) or DPM/treatment group (pooled treatment group approach).

Reduced LLNA

28. In certain situations, when there is a regulatory need to confirm a negative prediction of skin sensitising potential, an optional rLLNA protocol (16) (17) (18) using fewer animals may be used, provided there is adherence to all other LLNA protocol specifications in this TM. Before applying the rLLNA approach, clear justifications and scientific rationale for its use should be provided. If a positive or equivocal result is obtained, additional testing may be needed in order to interpret or clarify the finding.

29. The reduction in number of dose groups is the only difference between the LLNA and the rLLNA test method protocols and for this reason the rLLNA does not provide dose-response information. Therefore, the rLLNA should not be used when dose-response information is needed. Like the multi-dose LLNA, the test substance concentration evaluated in the rLLNA should be the maximum concentration that does not induce overt systemic toxicity and/or excessive local skin irritation in the mouse (see paragraph 18).

OBSERVATIONS

Clinical observations

30. Each mouse should be carefully observed at least once daily for any clinical signs, either of local irritation at the application site or of systemic toxicity. All observations are systematically recorded with records being maintained for each mouse. Monitoring plans should include criteria to promptly identify those mice exhibiting systemic toxicity, excessive local skin irritation, or corrosion of skin for euthanasia (37).

Body weights

31. As stated in paragraph 25, individual animal body weights should be measured at the start of the test and at the scheduled humane kill.

CALCULATION OF RESULTS

32. Results for each treatment group are expressed as the SI. When using the individual animal approach, the SI is derived by dividing the mean DPM/mouse within each test substance group, and the PC group, by the mean DPM/mouse for the solvent/VC group. The average SI for the VCs is then one. When using the pooled treatment group approach, the SI is

obtained by dividing the pooled radioactive incorporation for each treatment group by the incorporation of the pooled VC group; this yields a mean SI.

33. The decision process regards a result as positive when SI ≥ 3. However, the strength of the dose-response, the statistical significance and the consistency of the solvent/vehicle and PC responses may also be used when determining whether a borderline result is declared positive (4)(5)(6).

34. If it is necessary to clarify the results obtained, consideration should be given to various properties of the test substance, including whether it has a structural relationship to known skin sensitisers, whether it causes excessive local skin irritation in the mouse, and the nature of the dose-response relationship seen. These and other considerations are discussed in detail elsewhere (7).

35. Collecting radioactivity data at the level of the individual mouse will enable a statistical analysis for presence and degree of dose-response relationship in the data. Any statistical assessment could include an evaluation of the dose-response relationship as well as suitably adjusted comparisons of test groups (*e.g.* pair-wise dosed group versus concurrent VC comparisons). Statistical analyses may include, *e.g.* linear regression or William's test to assess dose-response trends, and Dunnett's test for pair-wise comparisons. In choosing an appropriate method of statistical analysis, the investigator should maintain an awareness of possible inequalities of variances and other related problems that may necessitate a data transformation or a non-parametric statistical analysis. In any case the investigator may need to carry out SI calculations and statistical analyses with and without certain data points (sometimes called "outliers").

DATA AND REPORTING

Data

36. Data should be summarised in tabular form. When using the individual animal approach, show the individual animal DPM values, the group mean DPM/animal, its associated error term (*e.g.* SD, SEM), and the mean SI for each dose group compared against the concurrent VC group. When using the pooled treatment group approach, show the mean/median DPM and the mean SI for each dose group compared against the concurrent VC group.

Test report

37. The test report should contain the following information:

Test and control substances:

- identification data (*e.g.* CAS and EC numbers, if available; source; purity; known impurities; lot number);
- physical nature and physicochemical properties (*e.g.* volatility, stability, solubility);
- if mixture, composition and relative percentages of components;

Solvent/vehicle:

- identification data (purity; concentration, where appropriate; volume used);

– justification for choice of vehicle;

Test animals:

– source of CBA mice;
– microbiological status of the animals, when known;
– number and age of animals;
– source of animals, housing conditions, diet, etc;

Test conditions:

– details of test substance preparation and application;
– justification for dose selection (including results from pre-screen test, if conducted);
– vehicle and test substance concentrations used, and total amount of test substance applied;
– details of food and water quality (including diet type/source, water source);
– details of treatment and sampling schedules;
– methods for measurement of toxicity;
– criteria for considering studies as positive or negative;
– details of any protocol deviations and an explanation on how the deviation affects the study design and results;

Reliability check:

– summary of results of latest reliability check, including information on test substance, concentration and vehicle used;
– concurrent and/or historical PC and concurrent NC data for testing laboratory;
– if a concurrent PC was not included, the date and laboratory report for the most recent periodic PC and a report detailing the historical PC data for the laboratory justifying the basis for not conducting a concurrent PC;

Results:

– individual weights of mice at start of dosing and at scheduled kill; as well as mean and associated error term (*e.g.* SD, SEM) for each treatment group;
– time course of onset and signs of toxicity, including dermal irritation at site of administration, if any, for each animal;
– a table of individual mouse (individual animal approach) or mean/median (pooled treatment group approach) DPM values and SI values for each treatment group;
– mean and associated error term (*e.g.* SD, SEM) for DPM/mouse for each treatment group and the results of outlier analysis for each treatment group when using the individual animal approach;
– calculated SI and an appropriate measure of variability that takes into account the inter-animal variability in both the test substance and control groups when using the individual animal approach;
– dose-response relationship;
– statistical analyses, where appropriate;

Discussion of results:

- a brief commentary on the results, the dose-response analysis, and statistical analyses, where appropriate, with a conclusion as to whether the test substance should be considered a skin sensitiser.

LITERATURE

(1) OECD (2002), Skin Sensitisation: Local Lymph Node Assay. OECD Guideline for the Testing of Chemicals No 429, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(2) Kimber, I. and Basketter, D.A. (1992), The murine local lymph node assay; collaborative studies and new directions: A commentary, *Food Chem. Toxicol*., 30, 165-169.

(3) Kimber, I., Dearman, R.J., Scholes, E.W. and Basketter, D.A. (1994), The local lymph node assay: developments and applications, *Toxicol*., 93, 13-31.

(4) Kimber, I., Hilton, J., Dearman, R.J., Gerberick, G.F., Ryan, C.A., Basketter, D.A., Lea, L., House, R.V., Ladies, G.S., Loveless, S.E. and Hastings, K.L. (1998), Assessment of the skin sensitisation potential of topical medicaments using the local lymph node assay: An interlaboratory exercise, *J. Toxicol. Environ. Health,* 53, 563-79.

(5) Chamberlain, M. and Basketter, D.A. (1996), The local lymph node assay: status of validation, *Food Chem. Toxicol*., 34, 999-1002.

(6) Basketter, D.A., Gerberick, G.F., Kimber, I. and Loveless, S.E. (1996), The local lymph node assay: A viable alternative to currently accepted skin sensitisation tests, *Food Chem. Toxicol*., 34, 985-997.

(7) Basketter, D.A., Gerberick, G.F. and Kimber, I. (1998), Strategies for identifying false positive responses in predictive sensitisation tests, *Food Chem. Toxicol*., 36, 327-33.

(8) Van Och, F.M.M., Slob, W., De Jong, W.H., Vandebriel, R.J. and Van Loveren, H. (2000), A quantitative method for assessing the sensitising potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of uncertainty margins, *Toxicol*., 146, 49-59.

(9) Dean, J.H., Twerdok, L.E., Tice, R.R., Sailstad, D.M., Hattan, D.G., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: II. Conclusions and recommendations of an independent scientific peer review panel, *Reg. Toxicol. Pharmacol,* 34: 258-273.

(10) Haneke, K.E., Tice, R.R., Carson, B.L., Margolin, B.H., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: III. Data analyses completed by the national toxicology program interagency center for the evaluation of alternative toxicological methods, *Reg. Toxicol. Pharmacol,* 34, 274-286.

(11) Sailstad, D.M., Hattan, D., Hill, R.N., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: I. The ICCVAM review process, *Reg. Toxicol. Pharmacol,* 34: 249-257.

(12) ICCVAM (2009), Recommended Performance Standards: Murine Local Lymph Node Assay, NIH Publication Number 09-7357, Research Triangle Park, NC:

National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf]

(13) OECD (1992), *Skin Sensitisation*. OECD Guideline for Testing of Chemicals No 406, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(14) OECD (2005), *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment*, Environment, Health and Safety Monograph, Series on Testing and Assessment No. 34, ENV/JM/MONO(2005)14, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(15) Dearman, R.J., Hilton, J., Evans, P., Harvey, P., Basketter, D.A. and Kimber, I. (1998), Temporal stability of local lymph node assay responses to hexyl cinnamic aldehyde, *J. Appl. Toxicol*., 18, 281-284.

(16) Kimber, I., Dearman, R.J., Betts, C.J., Gerberick, G.F., Ryan, C.A., Kern, P.S., Patlewicz, G.Y. and Basketter, D.A. (2006), The local lymph node assay and skin sensitisation: a cut-down screen to reduce animal requirements? *Contact Dermatitis*, 54, 181-185.

(17) ESAC (2007), Statement on the Reduced Local Lymph Node Assay (rLLNA), European Commission Directorate General, Joint Research Centre, Institute for Health and Consumer Protection, European Centre for the Validation of Alternative Methods, April 2007. Available at: [http://ecvam.jrc.it/ft_doc/ESAC26_statement_rLLNA_20070525-1.pdf]

(18) ICCVAM (2009), The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Test Method Evaluation Report. The Reduced Murine Local Lymph Node Assay: An Alternative Test Method Using Fewer Animals to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products, NIH Publication Number 09-6439, Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/]

(19) ICCVAM (1999), The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds, The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), NIH Publication No. 99-4494, Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf]

(20) Kreiling, R., Hollnagel, H.M., Hareng, L., Eigler, L., Lee, M.S., Griem, P., Dreessen, B., Kleber, M., Albrecht, A., Garcia, C. and Wendel, A. (2008), Comparison of the skin sensitising potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT), Food Chem. Toxicol., 46, 1896-1904.

(21) Basketter, D., Ball, N., Cagen, S., Carrilo, J.C., Certa, H., Eigler, D., Garcia, C., Esch, H., Graham, C., Haux, C., Kreiling, R. and Mehling, A. (2009), Application of a weight of evidence approach to assessing discordant sensitisation datasets: implications for REACH, Reg. Toxicol. Pharmacol., 55, 90-96.

(22) ICCVAM (2009), ICCVAM Test Method Evaluation Report. Assessment of the Validity of the LLNA for Testing Pesticide Formulations and Other Products, Metals, and Substances in Aqueous Solutions, NIH Publication Number 10-7512, Research Triangle Park, NC: National Institute of Environmental Health Sciences, Available at: [http://iccvam.niehs.nih.gov/]

(23) ILAR (1996), Institute of Laboratory Animal Research (ILAR) Guide for the Care and Use of Laboratory Animals, 7th ed. Washington, DC: National Academies Press.

(24) McGarry, H.F. (2007), The murine local lymph node assay: regulatory and potency considerations under REACH, Toxicol., 238, 71-89.

(25) OECD (2002), Acute Dermal Irritation/Corrosion. OECD Guideline for Testing of Chemicals No. 404, Paris, France. Available at: http://www.oecd.org/document/40/0,3343,en_2649_34377_37051368_1_1_1_1,00.html]

(26) Reeder, M.K., Broomhead, Y.L., DiDonato, L. and DeGeorge, G.L. (2007), Use of an enhanced local lymph node assay to correctly classify irritants and false positive substances, Toxicologist, 96, 235.

(27) ICCVAM (2009), Non-radioactive Murine Local Lymph Node Assay: Flow Cytometry Test Method Protocol (LLNA: BrdU-FC) Revised Draft Background Review Document, Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/immunotox/fcLLNA/BRDcomplete.pdf]

(28) Hayes, B.B., Gerber, P.C., Griffey, S.S. and Meade, B.J. (1998), Contact hypersensitivity to dicyclohexylcarbodiimide and diisopropylcarbodiimide in female B6C3F1 mice, Drug. Chem. Toxicol., 21, 195-206.

(29) Homey, B., von Schilling, C., Blumel, J., Schuppe, H.C., Ruzicka, T., Ahr, H.J., Lehmann, P. and Vohr, V.W. (1998), An integrated model for the differentiation of chemical-induced allergic and irritant skin reactions, *Toxicol. Appl. Pharmacol*., 153, 83-94.

(30) Woolhiser, M.R., Hayes, B.B. and Meade, B.J. (1998), A combined murine local lymph node and irritancy assay to predict sensitisation and irritancy potential of chemicals, *Toxicol. Meth*., 8, 245-256.

(31) Hayes, B.B. and Meade, B.J. (1999), Contact sensitivity to selected acrylate compounds in B6C3F1 mice: relative potency, cross reactivity, and comparison of test methods, *Drug. Chem. Toxicol*., 22, 491-506.

(32) Ehling, G., Hecht, M., Heusener, A., Huesler, J., Gamer, A.O., van Loveren, H., Maurer, T., Riecke, K., Ullmann, L., Ulrich, P., Vandebriel, R. and Vohr, H.W. (2005), A European inter-laboratory validation of alternative endpoints of the murine local lymph node assay: first round. *Toxicol*., 212, 60-68.

(33) Vohr, H.W. and Ahr, H.J. (2005), The local lymph node assay being too sensitive? *Arch. Toxicol*., 79, 721-728.

(34) Patterson, R.M., Noga, E. and Germolec, D. (2007), Lack of evidence for contact sensitisation by Pfiesteria extract, *Environ. Health Perspect*., 115, 1023-1028.

(35) OECD (1987*), Acute Dermal Toxicity,* OECD Guideline for Testing of Chemicals No 402, Paris, France. Available at: [http://www.oecd.org/env/testguidelines]

(36) ICCVAM (2009), Report on the ICCVAM-NICEATM/ECVAM/JaCVAM Scientific Workshop on Acute Chemical Safety Testing: Advancing In Vitro Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Research Triangle Park, NC: National Institute of Environmental Health Sciences, Available at: http://iccvam.niehs.nih.gov/methods/acutetox/Tox_workshop.htm]

(37) OECD (2000), *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation,* Environmental Health and Safety Monograph, Series on Testing and Assessment No. 19, ENV/JM/MONO(2000)7, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

Appendix 1

PERFORMANCE STANDARDS FOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED LLNA TEST METHODS FOR SKIN SENSITISATION

INTRODUCTION

1. The purpose of Performance Standards (PS) is to communicate the basis by which new test methods, both proprietary (*i.e.* copyrighted, trademarked, registered) and non-proprietary can be determined to have sufficient accuracy and reliability for specific testing purposes. These PS, based on validated and accepted test methods, can be used to evaluate the reliability and accuracy of other similar methods (colloquially referred to as "me-too" tests) that are based on similar scientific principles and measure or predict the same biological or toxic effect (14).

2. Prior to adoption of modified methods (*i.e.* proposed potential improvements to an approved test method), there should be an evaluation to determine the effect of the proposed changes on the test's performance and the extent to which such changes affect the information available for the other components of the validation process. Depending on the number and nature of the proposed changes, the generated data and supporting documentation for those changes, they should either be subjected to the same validation process as described for a new test, or, if appropriate, to a limited assessment of reliability and relevance using established PS (14).

3. Similar or modified methods proposed for use under this TM should be evaluated to determine their reliability and accuracy using chemicals representing the full range of the LLNA scores. To avoid unwarranted animal use, it is strongly recommended that model developers consult the appropriate authorities before starting validation studies in accordance with the PS and guidance provided in this TM.

4. These PS are based on the US-ICCVAM, EC-ECVAM and Japanese-JaCVAM harmonised PS (12), for evaluating the validity of similar or modified versions of the LLNA. The PS consists of essential test method components, recommended reference chemicals and standards for accuracy and reliability that the proposed method should meet or exceed.

I. Essential test method components

5. To ensure that a similar or modified LLNA method is functionally and mechanistically analogous to the LLNA and measures the same biological effect, the following components should be included in the test method protocol:

- The test substance should be applied topically to both ears of the mouse;
- Lymphocyte proliferation should be measured in the lymph nodes draining from the site of test substance application;
- Lymphocyte proliferation should be measured during the induction phase of skin sensitisation;
- For test substances, the highest dose selected should be the maximum concentration that does not induce systemic toxicity and/or excessive local skin irritation in the

mouse. For positive reference chemicals, the highest dose should be at least as high as the LLNA EC3 values of the corresponding reference chemicals (see Table 1) without producing systemic toxicity and/or excessive local skin irritation in the mouse;

- A concurrent VC should be included in each study and, where appropriate, a concurrent PC should also be used;
- A minimum of four animals per dose group should be used;
- Either individual or pooled animal data may be collected.

If any of these criteria are not met, then these PS cannot be used for validation of the similar or modified method.

II. Minimum list of reference chemicals

6. The US-ICCVAM, EC-ECVAM and Japanese-JaCVAM harmonised PS (12) identified 18 minimum reference chemicals that should be used and four optional reference chemicals (*i.e.* substances that produced either false positive or false negative results in the LLNA, when compared to human and guinea pig results (B.6, or OECD Test Guideline 406) (13), and therefore provide the opportunity to demonstrate equal to or better performance than the LLNA) that are included in the LLNA PS. The selection criteria for identifying these chemicals were:

- The list of reference chemicals represented the types of substances typically tested for skin sensitisation potential and the range of responses that the LLNA is capable of measuring or predicting;
- The substances had well-defined chemical structures;
- LLNA data from guinea pig tests (*i.e.* B.6; OECD Test Guideline 406) (13) and (where possible) data from humans were available for each substance; and
- The substances were readily available from a commercial source.

The recommended reference chemicals are listed in Table 1. Studies using the proposed reference chemicals should be evaluated in the vehicle with which they are listed in Table 1. In situations where a listed substance may not be available, other substances that meet the selection criteria mentioned may be used, with adequate justification.

TABLE 1: RECOMMENDED REFERENCE CHEMICALS FOR THE LLNA PS.

Number	Chemicals[1]	CAS No	Form	Veh[2]	EC3 %[3]	N[4]	0.5x - 2.0x EC3	Actual EC3 Range	LLNA vs. GP	LLNA vs. Human
1	5-Chloro-2-methyl-4-isothiazolin-3-one (CMI)/ 2-methyl-4-isothiazolin-3-one (MI)[5]	26172-55-4/ 2682-20-4	Liq	DMF	0.009	1	0.0045-0.018	NC	+/+	+/+
2	DNCB	97-00-7	Sol	AOO	0.049	15	0.025-0.099	0.02-0.094	+/+	+/+
3	4-Phenylenediamine	106-50-3	Sol	AOO	0.11	6	0.055-0.22	0.07-0.16	+/+	+/+
4	Cobalt chloride	7646-79-9	Sol	DMSO	0.6	2	0.3-1.2	0.4-0.8	+/+	+/+
5	Isoeugenol	97-54-1	Liq	AOO	1.5	47	0.77-3.1	0.5-3.3	+/+	+/+
6	2-Mercaptobenzothiazole	149-30-4	Sol	DMF	1.7	1	0.85-3.4	NC	+/+	+/+
7	Citral	5392-40-5	Liq	AOO	9.2	6	4.6-18.3	5.1-13	+/+	+/+
8	HCA	101-86-0	Liq	AOO	9.7	21	4.8-19.5	4.4-14.7	+/+	+/+
9	Eugenol	97-53-0	Liq	AOO	10.1	11	5.05-20.2	4.9-15	+/+	+/+
10	Phenyl benzoate	93-99-2	Sol	AOO	13.6	3	6.8-27.2	1.2-20	+/+	+/+
11	Cinnamic alcohol	104-54-1	Sol	AOO	21	1	10.5-42	NC	+/+	+/+
12	Imidazolidinyl urea	39236-46-9	Sol	DMF	24	1	12-48	NC	+/+	+/+
13	Methyl methacrylate	80-62-6	Liq	AOO	90	1	45-100	NC	+/+	+/+
14	Chlorobenzene	108-90-7	Liq	AOO	25	1	NA	NA	-/-	-/*
15	Isopropanol	67-63-0	Liq	AOO	50	1	NA	NA	-/-	-/+
16	Lactic acid	50-21-5	Liq	DMSO	25	1	NA	NA	-/-	-/*
17	Methyl salicylate	119-36-8	Liq	AOO	20	9	NA	NA	-/-	-/-
18	Salicylic acid	69-72-7	Sol	AOO	25	1	NA	NA	-/-	-/-

Number	Chemicals[1]	CAS No	Form	Veh[2]	EC3 %[3]	N[4]	0.5x - 2.0x EC3	Actual EC3 Range	LLNA vs. GP	LLNA vs. Human
Optional Substances to Demonstrate Improved Performance Relative to the LLNA										
19	Sodium lauryl sulphate	151-21-3	Sol	DMF	8.1	5	4.05-16.2	1.5-17.1	+/-	+/-
20	Ethylene glycol dimethacrylate	97-90-5	Liq	MEK	28	1	14-56	NC	+/-	+/+
21	Xylene	1330-20-7	Liq	AOO	95.8	1	47.9-100	NC	+/**	+/-
22	Nickel chloride	7718-54-9	Sol	DMSO	5	2	NA	NA	-/+	-/+

Abbreviations: AOO = acetone: olive oil (4:1, v/v); CAS No = Chemical Abstracts Service Number; DMF = *N,N*-dimethylformamide; DMSO = dimethyl sulfoxide; DNCB = 2,4-dinitrochlorobenzene; EC3 = estimated concentration needed to produce a stimulation index of 3; GP = guinea pig test result (*i.e.* B. 6 or OECD Test Guideline 406) (13); HCA = hexyl cinnamic aldehyde; Liq = liquid; LLNA = murine local lymph node assay result (*i.e.* B. 42 or OECD Test Guideline 429) (1); MEK = methyl ethyl ketone; NA = not applicable since stimulation index <3; NC = not calculated since data was obtained from a single study; Sol = solid; Veh = test vehicle.

[1] Chemicals should be prepared daily unless stability data demonstrate the acceptability of storage.

[2] Because of the potential impact of different vehicles on the performance of the LLNA, the recommended vehicle for each reference chemical should be used (24) (32).

[3] Mean value where more than one EC3 value was available. For negative substances (*i.e.* with stimulation index <3, the highest concentration tested is provided)

[4] Number of LLNA studies from which data were obtained.

[5] Commercially available as Kathon CG (CAS No 55965-84-9), which is a 3:1 mixture of CMI and MI. The relative concentrations of each component range from 1.1% to 1.25% (CMI) and 0.3% to 0.45% (MI). The inactive components are magnesium salts (21.5% to 24%) and copper nitrate (0.15% to 0.17%), with the remaining formulation 74% to 77% water. Kathon CG is readily available through Sigma-Aldrich and Rohm and Haas (now Dow Chemical Corporation).

* = Presumed to be a non-sensitiser in humans based on the fact that no clinical patch test results were located, it is not included as a patch test kit allergen, and no case reports of human sensitisation were located.

** = GP data not available.

III. Defined reliability and accuracy standards

7. The accuracy of a similar or modified LLNA method should meet or exceed that of the LLNA PS when it is evaluated using the 18 minimum reference chemicals that should be used. The new or modified method should result in the correct classification based on a "yes/no" decision. However, the new or modified method might not correctly classify all of the minimum reference chemicals that should be used. If, for example, one of the weak sensitisers were misclassified, a rationale for the misclassification and appropriate additional data (*e.g.* test results that provide correct classifications for other substances with physical, chemical, and sensitising properties similar to those of the misclassified reference chemical) could be considered to demonstrate equivalent performance. Under such circumstances, the validation status of the new or modified LLNA test method would be evaluated on a case-by-case basis.

Intra-laboratory reproducibility

8. To determine intra-laboratory reproducibility, a new or modified LLNA method should be assessed using a sensitising substance that is well characterised in the LLNA. Therefore, the LLNA PS are based on the variability of results from repeated tests of hexyl cinnamic aldehyde (HCA). To assess intra-laboratory reliability, threshold estimated concentration (ECt) values for HCA should be derived on four separate occasions with at least one week between tests. Acceptable intra-laboratory reproducibility is indicated by a laboratory's ability to obtain, in each HCA test, ECt values between 5% and 20%, which represents the range of 0.5-2.0 times the mean EC3 specified for HCA (10%) in the LLNA (see Table 1).

Inter-laboratory reproducibility

9. Inter-laboratory reproducibility of a new or modified LLNA method should be assessed using two sensitising substances that are well characterised in the LLNA. The LLNA PS are based on the variability of results from tests of HCA and 2,4-dinitrochlorobenzene (DNCB) in different laboratories. ECt values should be derived independently from a single study conducted in at least three separate laboratories. To demonstrate acceptable inter-laboratory reproducibility, each laboratory should obtain ECt values of 5% to 20% for HCA and 0.025% to 0.1% for DNCB, which represents the range of 0.5-2.0 times the mean EC3 concentrations specified for HCA (10%) and DNCB (0.05%), respectively, in the LLNA (see Table 1).

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 (14).

Benchmark substance: A sensitising or non-sensitising substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties; (i) consistent and reliable source(s); (ii) structural and functional similarity to the class of substances being tested; (iii) known physicochemical characteristics; (iv) supporting data on known effects, and (v) known potency in the range of the desired response.

Estimated concentration threshold (ECt): Estimated concentration of a test substance needed to produce a stimulation index that is indicative of a positive response.

Estimated concentration three (EC3): Estimated concentration of a test substance needed to produce a stimulation index of three.

False negative: A test substance incorrectly identified as negative or non-active by a test method, when in fact it is positive or active.

False positive: A test substance incorrectly identified as positive or active by a test, when in fact it is negative or non-active.

Hazard: The potential for an adverse health or ecological effect. The adverse effect is manifested only if there is an exposure of sufficient level.

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same test substances, can produce qualitatively and quantitatively similar results. Inter-laboratory reproducibility is determined during the pre-validation and validation processes, and indicates the extent to which a test can be successfully transferred between laboratories, also referred to as between-laboratory reproducibility (14).

Intra-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Also referred to as within-laboratory reproducibility (14).

Me-too test: A colloquial expression for a test method that is structurally and functionally similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation. Interchangeably used with similar test method (14).

Outlier: An outlier is an observation that is markedly different from other values in a random sample from a population.

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is functionally and

mechanistically similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (14).

Proprietary test method: A test method for which manufacture and distribution is restricted by patents, copyrights, trademarks, etc.

Quality assurance: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures, and the accuracy of data transfer, are assessed by individuals who are independent from those performing the testing.

Reference chemicals: Chemicals selected for use in the validation process, for which responses in the *in vitro* or *in vivo* reference test system or the species of interest are already known. These chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative. Different sets of reference chemicals may be required for the different stages of the validation process, and for different test methods and test uses (14).

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 (14).

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 (14).

Skin sensitisation: An immunological process that results when a susceptible individual is exposed topically to an inducing chemical allergen, which provokes a cutaneous immune response that can lead to the development of contact sensitisation.

Stimulation Index (SI): A value calculated to assess the skin sensitisation potential of a test substance that is the ratio of the proliferation in treated groups to that in the concurrent vehicle control group.

Test substance (also referred to as test chemical): Any substance or mixture tested using this TM.

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 (14)."

2. Chapter B.46 is replaced by the following:

"B.46 *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method

INTRODUCTION

1. Skin irritation refers to the production of reversible damage to the skin following the application of a test chemical for up to 4 hours [as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) and 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 (1)(3)]. This Test Method (TM) provides an *in vitro* procedure that may be used for the hazard identification of irritant chemicals (substances and mixtures) in accordance with UN GHS and EU CLP Category 2 (1) (2) (3). In the EU and other regions, that have not adopted the optional UN GHS Category 3 (mild irritants), this TM can also be used to identify non-classified chemicals, i.e. UN GHS and EU CLP "No Category" (1)(3). This TM may be used to determine the skin irritancy of chemicals as a stand-alone replacement test for *in vivo* skin irritation testing within a tiered testing strategy (4 and chapter B.4 in this Annex).

2. The assessment of skin irritation has typically involved the use of laboratory animals [OECD Test Guideline 404; chapter B.4 in this Annex] (4). In relation to animal welfare concerns, B.4 was revised in 2004 allowing for the determination of skin corrosion/irritation by applying a tiered testing strategy, using validated *in vitro* or *ex vivo* test methods, thus avoiding pain and suffering of animals. Three validated *in vitro* test methods have been adopted as OECD Test Guidelines 430, 431 and 435 (5) (6) (7) and two of them as chapters B.40 and B.40bis of this Annex, to be used for the corrosivity part of the tiered testing strategy of B.4 or OECD Test Guideline 404 (4).

3. This TM addresses the human health endpoint skin irritation. It is based on reconstructed human epidermis (RhE), which in its overall design (the use of human derived non-transformed epidermis keratinocytes as cell source and use of representative tissue and cytoarchitecture) closely mimics the biochemical and physiological properties of the upper parts of the human skin, i.e. the epidermis. This TM also includes a set of Performance Standards (PS) (Appendix 2) for the assessment of similar and modified RhE-based methods developed by EC-ECVAM (8), in accordance with the principles of OECD Guidance Document No. 34 (9).

4. There are three validated methods that adhere to this TM. Prevalidation, optimisation and validation studies have been completed for an *in vitro* method (10) (11) (12) (13) (14) (15) (16) (17) (18) (19) (20), using a RhE model, commercially available as EpiSkin™ (designated the Validated Reference Method – VRM). Two other commercially available *in vitro* skin irritation RhE methods have shown similar results to the VRM according to PS-based validation (21), and these are the EpiDerm™ SIT (EPI-200) and the SkinEthic™ RHE methods (22).

5. Before a proposed similar or modified *in vitro* RhE method other than the VRM, EpiDerm™ SIT (EPI-200) or SkinEthic™ RHE methods can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined in order to ensure that it can be regarded as similar to that of the VRM, in accordance with the requirements of the PS set out in this TM (Appendix 2). Moreover, it is recommended to

consult the OECD Explanatory Background Document on *in vitro* skin irritation testing before developing and validating a similar or modified *in vitro* RhE method and submitting it for regulatory adoption (23).

DEFINITIONS

6. Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

7. A limitation of the TM, as demonstrated by the validation study (16), is that it does not allow the classification of chemicals to the optional UN GHS Category 3 (mild irritants) (1). When used as a partial replacement test, follow-up *in vivo* testing may be required to fully characterize skin irritation potential (4 and chapter B.4 of this Annex). It is recognized that the use of human skin is subject to national and international ethical considerations and conditions.

8. This TM addresses the *in vitro* skin irritation component of the tiered testing strategy of B.4 (OECD Test Guideline 404) on dermal corrosion/irritation (4). While this TM does not provide adequate information on skin corrosion, it should be noted that B.40 bis (OECD Test Guideline 431) on skin corrosion is based on the same RhE test system, though using another protocol (chapter B.40 bis). This method is based on RhE-models using human keratinocytes, which therefore represent *in vitro* the target organ of the species of interest. It moreover directly covers the initial step of the inflammatory cascade/mechanism of action (cell damage and tissue damage resulting in localised trauma) that occurs during irritation *in vivo*. A wide range of chemicals has been tested in the validation underlying this TM and the empirical database of the validation study amounted to 58 chemicals in total (16)(18)(23). This is applicable to solids, liquids, semi-solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other pre-treatment of the sample is required. Gases and aerosols have not been assessed yet in a validation study (24). While it is conceivable that these can be tested using RhE technology, the current TM does not allow testing of gases and aerosols. It should also be noted that highly coloured chemicals may interfere with the cell viability measurements and need the use of adapted controls for corrections (see paragraphs 24-26).

9. A single testing run composed of three replicate tissues should be sufficient for a test chemical when the classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements and/or mean percent viability equal to 50 ± 5%, a second run should be considered, as well as a third one in case of discordant results between the first two runs.

PRINCIPLE OF THE TEST

10. The test chemical is applied topically to a three-dimensional RhE model, comprised of non-transformed human-derived epidermal keratinocytes, which have been cultured to form a multilayered, highly differentiated model of the human *epidermis*. It consists of organized basal, spinous and granular layers, and a multilayered *stratum corneum* containing intercellular lamellar lipid layers representing main lipid classes analogous to those found *in vivo*.

11. Chemical-induced skin irritation, manifested by erythema and oedema, is the result of a cascade of events beginning with penetration of the *stratum corneum* and damage to the underlying layers of keratinocytes. The dying keratinocytes release mediators that begin the inflammatory cascade which acts on the cells in the *dermis*, particularly the stromal and endothelial cells. It is the dilatation and increased permeability of the endothelial cells that produce the observed erythema and oedema (24). The RhE-based methods measure the initiating events in the cascade.

12. Cell viability in RhE models is measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue; CAS number 298-93-1 into a blue formazan salt that is quantitatively measured after extraction from tissues (25). Irritant chemicals are identified by their ability to decrease cell viability below defined threshold levels (*i.e.* ≤ 50%, for UN GHS/ EU CLP Category 2). Depending on the regulatory framework in which the results of this TM are used, chemicals that produce cell viabilities above the defined threshold level, may be considered non-irritants (*i.e.* > 50%, No Category).

DEMONSTRATION OF PROFICIENCY

13. Prior to routine use of any of the three validated methods that adhere to this TM, laboratories should demonstrate technical proficiency, using the ten Reference Chemicals listed in Table 1. For similar methods developed under this TM or for modifications of any of the three validated methods, the PS requirements described in Appendix 2 of this TM should be met prior to using the method for regulatory testing.

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

Table 1: Reference Chemicals[1]

Chemical	CAS NR	*In vivo* score[2]	Physical state	UN GHS/EU CLP Category
naphthalene acetic acid	86-87-3	0	Solid	No Cat.
isopropanol	67-63-0	0.3	Liquid	No Cat.
methyl stearate	112-61-8	1	Solid	No Cat.
heptyl butyrate	5870-93-9	1.7	Liquid	No Cat. (*Optional Cat. 3*)[3,4]
hexyl salicylate	6259-76-3	2	Liquid	No Cat. (*Optional Cat. 3*)[3,4]
cyclamen aldehyde	103-95-7	2.3	Liquid	Cat. 2
1-bromohexane	111-25-1	2.7	Liquid	Cat. 2
potassium hydroxide (5% aq.)	1310-58-3	3	Liquid	Cat. 2
1-methyl-3-phenyl-1-piperazine	5271-27-2	3.3	Solid	Cat. 2
Heptanal	111-71-7	3.4	Liquid	Cat. 2

[1] These Reference Chemicals are a subset of the Reference Chemicals used in the validation study.

[2] *In vivo* score in accordance with B.4 and OECD Test Guideline 404 (4).

[3] Under this Test Method, the UN GHS optional Category 3 (mild irritants) (1) is considered as No Category.

[4] The UN GHS optional Category3 is not applicable under the EU CLP.

PROCEDURE

15. The following is a description of the components and procedures of a RhE method for skin irritation assessment. A RhE model should be reconstructed, and can be in-house-prepared or obtained commercially. Standard Operating Procedures (SOPs) for the EpiSkin™, EpiDerm™ SIT (EPI-200) and SkinEthic™ RHE are available (26)(27)(28). Testing should be performed according to the following:

RHE TEST METHOD Components

General conditions

16. Non-transformed human keratinocytes should be used to reconstruct the epithelium. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum, stratum granulosum*) should be present under a functional *stratum corneum*. *Stratum corneum* should be multilayered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic marker chemicals, *e.g.* sodium dodecyl sulphate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET_{50}) upon application of the marker chemical at a specified, fixed concentration. The containment properties of the RhE model

should prevent the passage of material around the *stratum corneum* to the viable tissue, which would lead to poor modelling of skin exposure. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, or fungi.

Functional conditions

Viability

17. The assay used for determining the magnitude of viability is the MTT-assay (25). The RhE model users should ensure that each batch of the RhE model used meets defined criteria for the negative control (NC). The optical density (OD) of the extraction solvent alone should be sufficiently small, *i.e.* OD<0.1. An acceptability range (upper and lower limit) for the negative control OD values (in the Skin Irritation Test Method conditions) are established by the RhE model developer/supplier, and the acceptability ranges for the 3 validated methods are given in Table 2. It should be documented that the tissues treated with NC are stable in culture (provide similar viability measurements) for the duration of the test exposure period.

Table 2: Acceptability ranges for negative control OD values

	Lower acceptance limit	**Upper acceptance limit**
EpiSkin™ (SM)	≥ 0.6	≤ 1.5
EpiDerm™ SIT (EPI-200)	≥ 1.0	≤ 2.5
SkinEthic™ RHE	≥ 1.2	≤ 2.5

Barrier function

18. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of cytotoxic marker chemicals, *e.g.* SDS or Triton X-100, as estimated by IC_{50} or ET_{50} (Table 3).

Morphology

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

Reproducibility

20. The results of the positive control chemical (PC) and negative controls (NC) of the test method should demonstrate reproducibility over time.

Quality control (QC)

21. The RhE model developer/supplier should ensure and demonstrate that each batch of the RhE model used meets defined production release criteria, among which those for *viability* (paragraph 17), *barrier function* (paragraph 18) and *morphology* (paragraph 19) are the most relevant. These data should be provided to the method users, so that they are able to include this information in the test report. An acceptability range (upper and lower limit) for the IC_{50} or the ET_{50} should be established by the RhE model developer/supplier (or

investigator when using an in-house model). Only results produced with qualified tissues can be accepted for reliable prediction of irritation classification. As an example, the acceptability ranges for the three validated methods are given in Table 3.

Table 3: Examples of QC batch release criteria

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM) (18 hours treatment with SDS)(26)	IC_{50} = 1.0 mg/ml	IC_{50} = 3.0 mg/ml
EpiDerm™ SIT (EPI-200) (1% Triton X-100)(27)	ET_{50} = 4.8 hr	ET_{50} = 8.7 hr
SkinEthic™ RHE (1% Triton X-100)(28)	ET_{50} = 4.0 hr	ET_{50} = 9.0 hr

Application of the Test and Control Chemicals

22. At least three replicates should be used for each test chemical and for the controls in each run. For liquids as well as solids, sufficient amount of test chemical should be applied to uniformly cover the *epidermis* surface while avoiding an infinite dose, *i.e.* a minimum of 25 µL/cm^2 or 25 mg/cm^2 should be used. For solids, the *epidermis* surface should be moistened with deionised or distilled water before application, to improve contact between the test chemical and the *epidermis* surface. Whenever possible, solids should be tested as a fine powder. At the end of the exposure period, the test chemical should be carefully washed from the *epidermis* surface with aqueous buffer, or 0.9% NaCl. Depending on which of the three validated RhE methods is used, the exposure period varies between 15 and 60 minutes, and the incubation temperature between 20 and 37°C. These exposure periods and temperatures are optimized for each RhE method and represent the different intrinsic properties of the methods, for details, see the Standard Operating Procedures (SOPs) for the methods (26)(27)(28).

23. Concurrent NC and PC should be used in each run to demonstrate that viability (with the NC), barrier function and resulting tissue sensitivity (with the PC) of the tissues are within a defined historical acceptance range. The suggested PC is 5% aqueous SDS. The suggested NC chemicals are water or phosphate buffered saline (PBS).

Cell Viability Measurements

24. The most important element of the test procedure is that viability measurements are not performed immediately after the exposure to the test chemicals, but after a sufficiently long post-treatment incubation period of the rinsed tissues in fresh medium. This period allows both for recovery from weak cytotoxic effects and for appearance of clear cytotoxic effects. The test optimisation phase (11) (12) (13) (14) (15) demonstrated that a 42 hours post-treatment incubation period was optimal.

25. The MTT assay is a validated quantitative method which should be used to measure cell viability under this TM. It is compatible with use in a three-dimensional tissue construct. The tissue sample is placed in MTT solution of appropriate concentration (*e.g.* 0.3 - 1 mg/mL)

for 3 hours. The precipitated blue formazan product is then extracted from the tissue using a solvent (*e.g.* isopropanol, acidic isopropanol), and the concentration of formazan is measured by determining the OD at 570 nm using a filter band pass of maximum ± 30 nm.

26. Optical properties of the test chemical or its chemical action on the MTT may interfere with the assay leading to a false estimate of viability (because the test chemical may prevent or reverse the colour generation as well as cause it). This may occur when a specific test chemical is not completely removed from the tissue by rinsing or when it penetrates the *epidermis*. If a test chemical acts directly on the MTT (MTT-reducer), is naturally coloured, or becomes coloured during tissue treatment, additional controls should be used to detect and correct for test chemical interference with the viability measurement technique. Detailed description of how to correct direct MTT reduction and interferences by colouring agents is available in the SOPs for the three validated methods (26)(27)(28).

Acceptability Criteria

27. For each method using valid RhE model batches (see paragraph 21), tissues treated with the NC should exhibit OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes. Control OD values should not be below historically established boundaries. Similarly, tissues treated with the PC, *i.e.* 5% aqueous SDS, should reflect their ability to respond to an irritant chemical under the conditions of the TM(26) (27) (28). Associated and appropriate measures of variability between tissue replicates should be defined (*e.g.* if standard deviations (SD) are used they should be within the 1-sided 95% tolerance interval calculated from historical data; for the VRM SD < 18%).

Interpretation of Results and Prediction Model

28. The OD values obtained with each test chemical can be used to calculate the percentage of viability normalised to NC, which is set to 100%. The cut-off value of percentage cell viability distinguishing irritant from non-classified test chemicals and the statistical procedure(s) used to evaluate the results and identify irritant chemicals should be clearly defined, documented, and proven to be appropriate. The cut-off values for the prediction of irritation are given below:

> The test chemical is considered to be irritant to skin in accordance with UN GHS/EU CLP Category 2 if the tissue viability after exposure and post-treatment incubation is less than or equal (≤) to 50%.
>
> Depending on the regulatory framework in which the results of this TM are used, the test chemical may be considered to be non-irritant to skin in accordance with UN GHS/EU CLP No Category if the tissue viability after exposure and post-treatment incubation is more than (>) 50%.

DATA AND REPORTING

Data

29. For each run, data from individual replicate tissues (*e.g.* OD values and calculated percentage cell viability data for each test chemical, including classification) should be reported in tabular form, including data from repeat experiments as appropriate. In addition

means ± SD for each run should be reported. Observed interactions with MTT reagent and coloured test chemicals should be reported for each tested chemical.

Test Report

30. The test report should include the following information:

Test and Control Chemicals:

- Chemical name(s) such as CAS name and number, EC name and number, if known;
- Purity and composition of the chemical (in percentage(s) by weight);
- Physical/chemical properties relevant to the conduct of the study (*e.g.* physical state, stability, volatility, pH and water solubility if known);
- Treatment of the test/control chemicals prior to testing, if applicable (*e.g.* warming, grinding);
- Storage conditions;

Justification of the RhE model and protocol used

Test Conditions:

- Cell system used;
- Complete supporting information for the specific RhE model used including its performance. This should include, but is not limited to;
  - i) viability
  - ii) barrier function
  - iii) morphology
  - iv) reproducibility and predictivity
  - v) Quality controls (QC) of the model
- Details of the test procedure used;
- Test doses used, duration of exposure and post treatment incubation period;
- Description of any modifications of the test procedure;
- Reference to historical data of the model. This should include, but is not limited to:
  - i) acceptability of the QC data with reference to historical batch data
  - ii) acceptability of the positive and negative control values with reference to positive and negative control means and ranges
- Description of evaluation criteria used including the justification for the selection of the cut-off point(s) for the prediction model;
- Reference to historical control data;

Results:

- Tabulation of data from individual test chemicals for each run and each replicate measurement;
- Indication of controls used for direct MTT-reducers and/or colouring test chemicals;
- Description of other effects observed;

Discussion of the results

Conclusion

LITERATURE

(1) UN (2009), United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Third revised edition, UN New York and Geneva. Available at: [http://www.unece.org/trans/danger/publi/ghs/ghs_rev03/03files_e.html]

(2) EC-ECVAM (2009), Statement on the "Performance under UN GHS of three *in vitro* assays for skin irritation testing and the adaptation of the Reference Chemicals and Defined Accuracy Values of the ECVAM skin irritation Performance Standards", issued by the ECVAM Scientific Advisory Committee (ESAC30), 9 April 2009. Available at: [http://ecvam.jrc.ec.europa.eu]

(3) 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.

(4) OECD (2004), Acute Dermal Irritation/Corrosion, OECD Guideline for the Testing of Chemicals No. 404, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(5) OECD (2004), *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance (TER), OECD Guideline for the Testing of Chemicals No. 430, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(6) OECD (2004), *In Vitro* Skin Corrosion: Human Skin Model Test, OECD Guideline for the Testing of Chemicals No. 431, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(7) OECD (2006), *In Vitro* Membrane Barrier Test Method for Skin Corrosion, OECD Guideline for the Testing of Chemicals No. 435, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(8) EC-ECVAM (2009), Performance Standards for *in vitro* skin irritation test methods based on Reconstructed human Epidermis (RhE)? Available at: [http://ecvam.jrc.ec.europa.eu]

(9) 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. Available at: [http://www.oecd.org/env/testguidelines]

(10) Fentem, J.H., Briggs, D., Chesné, C., Elliot, G.R., Harbell, J.W., Heylings, J.R., Portes, P., Roguet, R., van de Sandt, J.J. M. and Botham, P. (2001), A prevalidation study on *in vitro* tests for acute skin irritation, Results and evaluation by the Management Team, Toxicol. in Vitro 15, 57-93.

(11) Portes, P., Grandidier, M.-H., Cohen, C. and Roguet, R. (2002), Refinement of the EPISKIN protocol for the assessment of acute skin irritation of chemicals: follow-up to the ECVAM prevalidation study, Toxicol. in Vitro 16, 765–770.

(12) Kandárová, H., Liebsch, M., Genschow, E., Gerner, I., Traue, D., Slawik, B. and Spielmann, H. (2004), Optimisation of the EpiDerm test protocol for the upcoming ECVAM validation study on *in vitro* skin irritation tests, ALTEX 21, 107–114.

(13) Kandárová, H., Liebsch, M., Gerner, I., Schmidt, E., Genschow, E., Traue, D. and Spielmann, H. (2005), The EpiDerm test protocol for the upcoming ECVAM validation study on *in vitro* skin irritation tests – An assessment of the performance of the optimised test, ATLA 33, 351-367.

(14) Cotovio, J., Grandidier, M.-H., Portes, P., Roguet, R. and Rubinsteen, G. (2005), The *in vitro* acute skin irritation of chemicals: optimisation of the EPISKIN prediction model within the framework of the ECVAM validation process, ATLA 33, 329-349.

(15) Zuang, V., Balls, M., Botham, P.A., Coquette, A., Corsini, E., Curren, R.D., Elliot, G.R., Fentem, J.H., Heylings, J.R., Liebsch, M., Medina, J., Roguet, R., van De Sandt, J.J.M., Wiemann, C. and Worth, A. (2002), Follow-up to the ECVAM prevalidation study on *in vitro* tests for acute skin irritation, The European Centre for the Validation of Alternative Methods Skin Irritation Task Force report 2, ATLA 30, 109-129.

(16) Spielmann, H., mailto:Hoffmann, S., Liebsch, M., Botham, P., Fentem, J., Eskes, C., Roguet, R., Cotovio, J., Cole, T., Worth, A., Heylings, J., Jones, P., Robles, C., Kandárová, H., mailto:Gamer, A., Remmele, M., Curren, R., Raabe, H., Cockshott, A., Gerner, I. and Zuang, V. (2007), The ECVAM international validation study on *in vitro* tests for acute skin irritation: Report on the validity of the EPISKIN and EpiDerm assays and on the skin integrity function test, ATLA 35, 559-601.

(17) Hoffmann, S. (2006), ECVAM skin irritation validation study phase II: Analysis of the primary endpoint MTT and the secondary endpoint IL1-α. Available at: [http://ecvam.jrc.ec.europa.eu]

(18) Eskes, C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I. and Zuang, V. (2007), The ECVAM international validation study on *in vitro* tests for acute skin irritation: selection of test chemicals, ATLA 35, 603-619.

(19) Cotovio, J., Grandidier, M.-H., Lelièvre, D., Roguet, R., Tinois-Tessonneaud, E. and Leclaire, J. (2007), *In vitro* acute skin irritancy of chemicals using the validated EPISKIN model in a tiered strategy - Results and performances with 184 cosmetic ingredients, AATEX, 14, 351-358.

(20) EC-ECVAM (2007), Statement on the validity of *in vitro* tests for skin irritation, issued by the ECVAM Scientific Advisory Committee (ESAC26), 27 April 2007. Available at: [http://ecvam.jrc.ec.europa.eu]

(21) EC-ECVAM (2007), Performance Standards for applying human skin models to *in vitro* skin irritation testing. Available at: [http://ecvam.jrc.ec.europa.eu]

(22) EC-ECVAM (2008), Statement on the scientific validity of *in vitro* tests for skin irritation testing, issued by the ECVAM Scientific Advisory Committee (ESAC29), 5 November 2008. Available at: [http://ecvam.jrc.ec.europa.eu]

(23) OECD (2010), Explanatory background document to the OECD draft Test Guideline on *in vitro* skin irritation testing. OECD Series on Testing and Assessment, No. 137, OECD,

Paris. Available at:
[http://www.oecd.org/document/24/0,3746,en_2649_34377_47858904_1_1_1_1,00.html]

(24) Welss, T., Basketter, D.A. and Schröder, K.R. (2004), *In vitro* skin irritation: fact and future. State of the art review of mechanisms and models, Toxicol. in Vitro 18, 231-243.

(25) Mosmann, T. (1983), Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays, J. Immunol. Methods 65, 55-63.

(26) EpiSkin™ SOP, Version 1.8 (February 2009), ECVAM Skin Irritation Validation Study: Validation of the EpiSkin™ test method 15 min - 42 hours for the prediction of acute skin irritation of chemicals. Available at: [http://ecvam.jrc.ec.europa.eu]

(27) EpiDerm™ SOP, Version 7.0 (Revised March 2009), Protocol for: *In vitro* EpiDerm™ skin irritation test (EPI-200-SIT), For use with MatTek Corporation's reconstructed human epidermal model EpiDerm (EPI-200). Available at: [http://ecvam.jrc.ec.europa.eu]

(28) SkinEthic™ RHE SOP, Version 2.0 (February 2009), SkinEthic skin irritation test-42bis test method for the prediction of acute skin irritation of chemicals: 42 minutes application + 42 hours post-incubation. Available at: [http://ecvam.jrc.ec.europa.eu]

(29) Harvell, J.D., Lamminstausta, K., and Maibach, H.I. (1995), Irritant contact dermatitis, In: Practical Contact Dermatitis, pp 7-18, (Ed. Guin J. D.). Mc Graw-Hill, New York.

(30) Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances, OJ L 225, 21.8.2001, p.1.

(31) Basketter, D.A., York, M., McFadden, J.P. and Robinson, M.K. (2004), Determination of skin irritation potential in the human 4-h patch test. Contact Dermatitis 51, 1-4.

(32) Jirova, D., Liebsch, M., Basketter, D., Spiller, E., Kejlova, K., Bendova, H., Marriott, M. and Kandarova, H. (2007), Comparison of human skin irritation and photo-irritation patch test data with cellular *in vitro* assays and animal *in vivo* data, ALTEX, 14, 359-365.

(33) Jírová, D., Basketter, D., Liebsch, M., Bendová, H., Kejlová, K., Marriott, M. and Kandárová, H. (2010), Comparison of human skin irritation patch test data with *in vitro* skin irritation assays and animal data, Contact Dermatitis, 62, 109-116.

Appendix 1

DEFINITIONS

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

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

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

ET_{50}: Can be estimated by determination of the exposure time required to reduce cell viability by 50% upon application of the marker chemical at a specified, fixed concentration, see also IC_{50}.

EU CLP (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): Implements in the European Union (EU) the UN GHS system for the classification and labelling of chemicals (substances and mixtures)(3).

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

IC_{50}: Can be estimated by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, see also ET_{50}.

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

Me-too test: A colloquial expression for a test method that is structurally and functionally similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation. Interchangeably used with similar test method (9).

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

reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (9).

Reference chemicals: Chemicals selected for use in the validation process, for which responses in the *in vitro* or *in vivo* reference test system or the species of interest are already known. These chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative. Different sets of reference chemicals may be required for the different stages of the validation process, and for different test methods and test uses (9).

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

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

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

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

Skin irritation: The production of reversible damage to the skin following the application of a test chemical for up to 4 hours. Skin irritation is a locally arising, non-immunogenic reaction, which appears shortly after stimulation (29). Its main characteristic is its reversible nature involving inflammatory reactions and most of the clinical characteristic signs of irritation (erythema, oedema, itching and pain) related to an inflammatory process.

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

Tiered testing strategy: Testing which uses test methods in a sequential manner; the test methods selected in each succeeding level are decided based on the results in the previous level of testing (9).

Test chemical (also referred to as test substance): Any substance or mixture tested using this TM.

Appendix 2

PERFORMANCE STANDARDS FOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED *IN VITRO* RECONSTRUCTED HUMAN *EPIDERMIS* (RhE) METHODS FOR SKIN IRRITATION

INTRODUCTION

1. The purpose of Performance Standards (PS) is to communicate the basis by which new methods, both proprietary (*i.e.* copyrighted, trademarked, registered) and non-proprietary can be determined to have sufficient accuracy and reliability for specific testing purposes. These PS, based on validated and accepted methods, can be used to evaluate the reliability and accuracy of other analogous methods (colloquially referred to as "me-too" tests) that are based on similar scientific principles and measure or predict the same biological or toxic effect (9).

2. Prior to adoption of modified methods, *i.e.* proposed potential improvements to an approved method, there should be an evaluation to determine the effect of the proposed changes on the test's performance and the extent to which such changes affect the information available for the other components of the validation process. Depending on the number and nature of the proposed changes, the generated data and supporting documentation for those changes, they should either be subjected to the same validation process as described for a new test, or, if appropriate, to a limited assessment of reliability and relevance using established PS (9).

3. Similar (me-too) or modified methods of any of the three validated methods [EpiSkin™ (Validated Reference Method – VRM), EpiDerm™ SIT (EPI-200) and SkinEthic™ RHE] proposed for use under this TM should be evaluated to determine their reliability and accuracy using chemicals representing the full range of the Draize irritancy scores. When evaluated using the 20 recommended Reference Chemicals of the PS (Table 1), the proposed similar or modified methods should have reliability and accuracy values which are comparable or better than those derived from the VRM (Table 2) (2) (16). The reliability and accuracy values that should be achieved are provided in paragraphs 8 to 12 of this Appendix. Non-classified (UN GHS/EU CLP No Category) and classified (UN GHS/EU CLP Category 2) (1) chemicals, representing different chemical classes are included, so that the reliability and accuracy (sensitivity, specificity and overall accuracy) of the proposed method can be compared to that of the VRM. The reliability of the method, as well as its ability to correctly identify UN GHS/EU CLP Category 2 irritant chemicals and, depending on the regulatory framework for which data are produced, also its ability to correctly identify UN GHS/EU CLP No Category chemicals (should be determined prior to its use for testing new test chemicals.

4. These PS are based on the EC-ECVAM PS (8), updated according to the UN GHS and EU CLP systems on classification and labelling (1) (3). The original PS were defined after the completion of the validation study (21) and were based on the EU classification system as laid down in Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances[1]. Due to the adoption of the UN GHS system for classification and

[1] OJ L 225, 21.8.2001, p. 1.

labelling in EU (EU CLP) (3), which took place between the finalisation of the validation study and the completion of this TM, the PS have been updated (8). This update concerns mainly changes (i), in the set of the PS Reference Chemicals; and (ii), the defined reliability and accuracy values (2) (23).

PERFORMANCE STANDARDS FOR *IN VITRO* RhE TEST METHODS FOR SKIN IRRITATION

5. The PS comprises the following three elements (9):

I) Essential Test Method Components

II) Minimum List of Reference Chemicals

III) Defined Reliability and Accuracy Values

I) Essential Test Method Components

6. These consist of essential structural, functional, and procedural elements of a validated method that should be included in the protocol of a proposed, mechanistically and functionally similar or modified method. These components include unique characteristics of the method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a similar or modified proposed method is based on the same concepts as the corresponding VRM (9). The essential test method components are described in detail in paragraphs 16 to 21 of the TM and testing should be performed according to the following:

The general conditions (paragraph 16)

The functional conditions, which include:

- viability (paragraph 17);
- barrier function (paragraph 18);
- morphology (paragraph 19);
- reproducibility (paragraph 20); and,
- quality control (paragraph 21)

II) Minimum List of Reference Chemicals

7. Reference Chemicals are used to determine if the reliability and accuracy of a proposed similar or modified method, proven to be structurally and functionally sufficiently similar to the VRM, or representing a minor modification of one of the three validated methods, are comparable or better than those of the VRM (2) (8) (16) (23). The 20 recommended Reference Chemicals listed in Table 1 include chemicals representing different chemical classes (*i.e.* chemical categories based on functional groups), and are representative of the full range of Draize irritancy scores (from non-irritant to strong irritant). The chemicals included in this list comprise 10 UN GHS/EU CLP Category 2 chemicals and 10 non-categorised chemicals, of which 3 are optional UN GHS Category 3 chemicals. Under this Test Method, the optional Category 3 is considered as No Category. The chemicals listed in Table 1 are selected from the chemicals used in the optimisation phase that followed prevalidation and in the validation study of the VRM, with regard to chemical functionality and physical state (14) (18). These Reference Chemicals represent the minimum number of chemicals that should be used to evaluate the accuracy and reliability of a proposed similar or modified method, but should not be used for the development of new methods. In situations where a listed chemical is unavailable, other chemicals for which adequate *in vivo* reference data are available could be used, primarily from the chemicals used in the optimisation phase following prevalidation or the validation study of the VRM. If desired, additional chemicals

representing other chemical classes and for which adequate *in vivo* reference data are available may be added to the minimum list of Reference Chemicals to further evaluate the accuracy of the proposed method.

Table 1: Minimum List of Reference Chemicals for Determination of Accuracy and Reliability Values for Similar or Modified RhE Skin Irritation Methods[1]

Chemical	**CAS Number**	**Physical state**	***In vivo* score**	**VRM *in vitro* Cat.**	**UN GHS/EU CLP *in vivo* Cat.**
1-bromo-4-chlorobutane	6940-78-9	Liquid	0	Cat. 2	No Cat.
diethyl phthalate	84-66-2	Liquid	0	No Cat.	No Cat.
naphthalene acetic acid	86-87-3	Solid	0	No Cat.	No Cat.
allyl phenoxy-acetate	7493-74-5	Liquid	0.3	No Cat.	No Cat.
isopropanol	67-63-0	Liquid	0.3	No Cat.	No Cat.
4-methyl-thio-benzaldehyde	3446-89-7	Liquid	1	Cat. 2	No Cat.
methyl stearate	112-61-8	Solid	1	No Cat.	No Cat.
heptyl butyrate	5870-93-9	Liquid	1.7	No Cat.	No Cat.
hexyl salicylate	6259-76-3	Liquid	2	No Cat.	No Cat.
Cinnamaldehyde	104-55-2	Liquid	2	Cat. 2	No Cat. (*Optional Cat. 3*)[3]
1-decanol[2]	*112-30-1*	*Liquid*	*2.3*	*Cat. 2*	*Cat. 2*
cyclamen aldehyde	103-95-7	Liquid	2.3	Cat. 2	Cat. 2
1-bromohexane	111-25-1	Liquid	2.7	Cat. 2	Cat. 2
2-chloromethyl-3,5-dimethyl-4-methoxypyridine HCl	86604-75-3	Solid	2.7	Cat. 2	Cat. 2
di-n-propyl disulphide[2]	*629-19-6*	*Liquid*	*3*	*No Cat.*	*Cat. 2*
potassium hydroxide (5% aq.)	1310-58-3	Liquid	3	Cat. 2	Cat. 2
benzenethiol, 5-(1,1-dimethylethyl)-2-methyl	7340-90-1	Liquid	3.3	Cat. 2	Cat. 2
1-methyl-3-phenyl-1-piperazine	5271-27-2	Solid	3.3	Cat. 2	Cat. 2
Heptanal	111-71-7	Liquid	3.4	Cat. 2	Cat. 2
Tetrachloroethylene	127-18-4	Liquid	4	Cat. 2	Cat. 2

[1] The chemical selection is based on the following criteria (i), the chemicals are commercially available; (ii), they are representative of the full range of Draize irritancy scores (from non-irritant to strong irritant); (iii), they have a well-defined chemical structure; (iv), they are representative of the chemical functionality used in the validation process; and (v), they are not associated with an extremely toxic profile (*e.g.* carcinogenic or toxic to the reproductive system) and they are not associated with prohibitive disposal costs.

[2] Chemicals that are irritant in the rabbit but for which there is reliable evidence that they are non-irritant in humans (31) (32) (33).

[3] Under the UN GHS, not in the EU CLP

III) Defined Reliability and Accuracy Values

8. For purposes of establishing the reliability and relevance of proposed similar or modified methods to be transferred between laboratories, all 20 Reference Chemicals in Table 1 should be tested in at least three laboratories. However, if the proposed method is to be used in a single laboratory only, multi-laboratory testing will not be required for validation. It is however essential that such validation studies are independently assessed by internationally recognised validation bodies, in agreement with international guidelines (9). In each laboratory, all 20 Reference Chemicals should be tested in three independent runs performed with different tissue batches and at sufficiently spaced time points. Each run should consist of a minimum of three concurrently tested tissue replicates for each included test chemical, NC and PC.

9. The calculation of the reliability and accuracy values of the proposed method should be done considering all four criteria below together, ensuring that the values for reliability and relevance are calculated in a predefined and consistent manner:

1. Only the data of runs from complete run sequences qualify for the calculation of the method within, and between-laboratory variability and predictive capacity (accuracy).

2. The final classification for each Reference Chemicals in each participating laboratory should be obtained by using the mean value of viability over the different runs of a complete run sequence.

3. Only the data obtained for chemicals that have complete run sequences in all participating laboratories qualify for the calculation of the method between-laboratory variability.

4. The calculation of the accuracy values should be done on the basis of the individual laboratory predictions obtained for the 20 Reference Chemicals by the different participating laboratories.

In this context, a **run sequence** consists of three independent runs from one laboratory for one test chemical. A **complete run sequence** is a run sequence from one laboratory for one test chemical where all three runs are valid. This means that any single invalid run invalidates an entire run sequence of three runs.

Within-laboratory reproducibility

10. An assessment of within-laboratory reproducibility should show a concordance of classifications (UN GHS/EU CLP Category 2 and No Category) obtained in different, independent test runs of the 20 Reference Chemicals within one single laboratory equal or higher (≥) than 90%.

Between-laboratory reproducibility

11. An assessment of between-laboratory reproducibility is not essential if the proposed method is to be used in a single laboratory only. For methods to be transferred between laboratories, the concordance of classifications (UN GHS/EU CLP Category 2 and No Category) obtained in different, independent test runs of the 20 Reference Chemicals between

preferentially a minimum of three laboratories should be equal or higher (≥) than 80%.

Predictive capacity (accuracy)

12. The accuracy (sensitivity, specificity and overall accuracy) of the proposed similar or modified method should be comparable or better to that of the VRM, taking into consideration additional information relating to relevance in the species of interest (Table 2). The sensitivity should be equal or higher (≥) than 80% (2) (8) (23). However, a further specific restriction applies to the sensitivity of the proposed *in vitro* method inasmuch as only two *in vivo* Category 2 chemicals, *1-decanol* and *di-n-propyl disulphide,* may be misclassified as No Category by more than one participating laboratory. The specificity should be equal or higher (≥) than 70% (2) (8) (23). There is no further restriction with regard to the specificity of the proposed *in vitro* method, *i.e.* any participating laboratory may misclassify any *in vivo* No Category chemical as long as the final specificity of the test method is within the acceptable range. The overall accuracy should be equal or higher (≥) than 75% (2) (8) (23). Although the sensitivity of the VRM calculated for the 20 Reference Chemicals listed in Table 1 is equal to 90%, the defined minimum sensitivity value required for any similar or modified method to be considered valid is set at 80% since both *1-decanol* (a borderline chemical) and *di-n-propyl disulphide* (a false negative of the VRM) are known to be non-irritant in humans (31) (32) (33), although being identified as irritants in the rabbit test. Since RhE models are based on cells of human origin, they may predict these chemicals as non-irritant (UN GHS/EU CLP No Category).

Table 2. Required predictive values for sensitivity, specificity and overall accuracy for any similar or modified method to be considered valid.

Sensitivity	**Specificity**	**Overall Accuracy**
≥ 80%	≥ 70%	≥ 75%

Study Acceptance Criteria

13. It is possible that one or several tests pertaining to one or more test chemicals does/do not meet the test acceptance criteria for the test and control chemicals or is/are not acceptable for other reasons. To complement missing data, for each test chemical a maximum number of two additional tests is admissible ("retesting"). More precisely, since in case of retesting also PC and NC have to be concurrently tested, a maximum number of two additional runs may be conducted for each test chemical.

14. It is conceivable that even after retesting, the minimum number of three valid runs required for each tested chemical is not obtained for every Reference Chemical in every participating laboratory, leading to an incomplete data matrix. In such cases the following three criteria should all be met in order to consider the datasets acceptable:

1. All 20 Reference Chemicals should have at least one complete run sequence.

2. In each of at least three participating laboratories, a minimum of 85% of the run sequences need to be complete (for 20 chemicals; *i.e.* 3 invalid run sequences are allowed in a single laboratory).

3. A minimum of 90% of all possible run sequences from at least three laboratories need to be complete (for 20 chemicals tested in 3 laboratories; *i.e.* 6 invalid run sequences are allowed in total)."

3. The following Chapters are added

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

INTRODUCTION

1. The *in vitro* micronucleus (MNvit) assay 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. The assay detects the activity of clastogenic and aneugenic chemicals (substances and mixtures) (1) (2) in cells that have undergone cell division during or after exposure to the test substance. This Test Method (TM) allows the use of protocols with and without the actin polymerisation inhibitor cytochalasin B (cytoB). The addition of cytoB prior to the targeted mitosis allows for the identification and selective analysis of micronucleus frequency in cells that have completed one mitosis because such cells are binucleate (3) (4). This TM also allows the use of protocols without cytokinesis block, provided there is evidence that the cell population analysed has undergone mitosis.

2. In addition to using the MNvit assay to identify chemicals (substances and mixtures) that induce micronuclei, the use of a cytokinesis block, immunochemical labelling of kinetochores, or hybridisation with centromeric/telomeric probes (fluorescence *in situ* hybridisation (FISH)), also can provide information on the mechanisms of chromosome damage and micronucleus formation (5) (6) (7) (8) (9) (10) (11) (12) (13) (14) (15) (16). The 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.

3. Micronuclei represent damage that has been transmitted to daughter cells, whereas chromosome aberrations scored in metaphase cells may not be transmitted. Because micronuclei in interphase cells can be assessed relatively objectively, laboratory personnel need only determine whether or not the cells have undergone division and how many cells contain a micronucleus. As a result, the preparations 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 assay. 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.* OECD Test Guideline 473 (chapter B.10 of this Annex) (17). However, the MNvit assay does not allow for the differentiation of chemicals inducing polyploidy from those inducing clastogenicity without special techniques such as FISH described under paragraph 2.

4. The MNvit assay is an *in vitro* method that typically uses cultured human or rodent cells. It provides a comprehensive basis for investigating chromosome damaging potential *in vitro* because both aneugens and clastogens can be detected.

5. The MNvit assay is robust and effective 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 assay using various rodent cell lines (CHO, V79, CHL/IU, and L5178Y) and human lymphocytes (18) (19) (20) (21) (22) (23) (24) (25) (26) (27) (28) (29) (30) (31). These include, in particular, the international validation studies co-ordinated by the Société Française de Toxicologie

Génétique (SFTG) (18) (19) (20) (21) (22) and the reports of the International Workshop on Genotoxicity Testing (4) (16). 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, and the test method has been endorsed as scientifically valid by the ECVAM Scientific Advisory Committee (ESAC) (32) (33) (34). The use of the human TK6 lymphoblastoid cell line (35), HepG2 cells (36) (37) and primary Syrian Hamster Embryo cells (38) has been described, although they have not been used in validation studies.

DEFINITIONS

6. Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS

7. 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 substances being tested. The exogenous metabolic activation system does not entirely mimic *in vivo* conditions. Care should also be taken to avoid conditions that would lead to artifactual positive results which do not reflect intrinsic mutagenicity, and may arise from such factors as marked changes in pH or osmolality, or by high levels of cytotoxicity (39) (40) (41). If the test chemical causes a change in the pH of the medium at the time of addition, the pH should be adjusted, preferably by buffering the stock solution so that all the volumes at all test concentrations, and for all controls, remain the same.

8. 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 substance.

PRINCIPLE OF THE TEST

9. Cell cultures of human or mammalian origin are exposed to the test substance both with and without an exogenous source of metabolic activation unless cells with an adequate metabolizing capability are used. Concurrent solvent/vehicle (VC) and positive control chemicals (PC) are included in all tests.

10. During or after exposure to the test substance, the cells are grown for a period sufficient to allow chromosome or spindle damage to lead to the formation of micronuclei in interphase cells. For induction of aneuploidy, the test substance 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 substance or during the post-exposure period, if one is used. In cultures that have been treated with a cytokinesis blocker, this is 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 during or after exposure to the test substance. 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 substance-induced cytotoxicity or cytostasis should be assessed in the cultures (or in parallel cultures) that are scored for micronuclei.

DESCRIPTION OF THE ASSAY

Preparations

11. Cultured primary human peripheral blood lymphocytes (5) (19) (42) (43) and a number of rodent cell lines such as CHO, V79, CHL/IU, and L5178Y cells may be used (18) (19) (20) (21) (22) (25) (26) (27) (28) (30). The use of other cell lines and types should be justified based on their demonstrated performance in the assay, as described in the Acceptability Criteria section. Because the background frequency of micronuclei will influence the sensitivity of the assay, it is recommended that cell types with a low, stable background frequency of micronucleus formation be used.

12. Human peripheral blood lymphocytes should be obtained from young (approximately 18-35 years of age), healthy, non-smoking individuals with no known recent exposures to genotoxic chemicals or radiation. If cells from more than one donor are pooled for use, the number of donors should be specified. The micronucleus frequency increases with age and this trend is more marked in females than in males (44) and this should be taken into account in the selection of donor cells for pooling.

Media and culture conditions

13. Appropriate culture medium and incubation conditions (culture vessels, CO_2 concentration, temperature, and humidity) should be used for maintaining cultures. Established cell lines and strains should be checked routinely for the stability of the modal chromosome number and the absence of mycoplasma contamination, and should not be used if contaminated or if the modal chromosome number has changed. The normal cell cycle time for the culture conditions used in the testing laboratory should be known. If the cytokinesis-block method is used then the concentration of the cytokinesis inhibitor should be optimised for the particular cell type and should be shown to produce a good yield of binucleate cells for scoring.

Preparation of cultures

14. Established cell lines and strains: cells are propagated from stock cultures, seeded in culture medium at a density such that the cultures will not reach confluency in monolayers, and suspension cultures will not reach excessive density before the time of harvest, and incubated at 37°C.

15. Lymphocytes: whole blood treated with an anti-coagulant (*e.g.* heparin), or separated lymphocytes, are cultured in the presence of a mitogen *e.g.* phytohaemagglutinin (PHA) prior to exposure to the test substance and cytoB.

Metabolic activation

16. Exogenous metabolising systems should be used when using cells with inadequate endogenous metabolic capacity. The most commonly used system is a co-factor-supplemented post-mitochondrial fraction (S9) prepared from the livers of rodents treated with enzyme-inducing agents such as Aroclor 1254 (45) (46) or a combination of phenobarbitone and β-naphthoflavone (46) (47) (48) (49). The latter combination does not conflict with the Stockholm Convention on Persistent Organic Pollutants (50) and Regulation (EC) No 850/2004 on Persistent Organic Pollutants (66) and has been shown to be as effective

as Aroclor 1254 for inducing mixed-function oxidases (46) (47) (48) (49). The S9 fraction typically is used at concentrations ranging from 1-10% (v/v) in the final test medium. The condition of a metabolic activation system may depend upon the class of chemical being tested and in some cases it may be appropriate to utilise more than one S9 concentration.

17. Genetically engineered cell lines expressing specific human or rodent activating enzymes may eliminate the need for an exogenous metabolic activation system, and may be used as the test cells. In such cases the choice of the cell lines used should be scientifically justified, *e.g.* by relevance of the mixed function oxidases for the metabolism of the test substance (51), and their responsiveness to known clastogens and aneugens (see separate section on Acceptability Criteria). It should be recognized that the substance being tested may not be metabolised by the expressed mixed function oxidase(s); in this case, the negative results would not indicate that the test substance cannot induce micronuclei.

Test substance preparation

18. Solid chemicals should be dissolved in appropriate solvents or vehicles and diluted, if appropriate, prior to treatment of the cells. Liquid chemicals may be added directly to the test systems and/or diluted prior to treatment. Gases or volatile chemicals should be tested by appropriate modifications to the standard protocols, such as treatment in sealed vessels (52) (53). Fresh preparations of the test substance should be used unless stability data demonstrate the acceptability of storage.

Test Conditions

Solvents/vehicles

19. The solvent/vehicle should not react with the test substance, or be incompatible with the survival of the cells or with the maintenance of S9 activity at the concentration used. If other than well established solvent/vehicles (*e.g.* water, cell culture medium, dimethyl sulfoxide) are used, their use should be supported by data indicating their compatibility with the test substance and their lack of genetic toxicity. It is recommended that, wherever possible, the use of an aqueous solvent/vehicle should be considered first.

Use of cytoB as a cytokinesis blocker

20. One of the most important considerations in the performance of the MNvit assay is ensuring that the cells being scored have completed mitosis during the treatment or the post-treatment incubation period, if one is used. 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 binucleated cells (5) (54) (55). Micronucleus scoring, therefore, can be limited to cells that have gone through mitosis during or after treatment. The effect of the test substance on cell proliferation kinetics can be measured simultaneously. CytoB should be used as a cytokinesis blocker when human lymphocytes are used because cell cycle times will be variable within cultures and among donors and because not all lymphocytes will respond to PHA. Other methods have been used when testing cell lines to determine if the cells being scored have divided; these are addressed below (see Paragraph 26).

21. The appropriate concentration of cytoB should be determined by the laboratory for each cell type to achieve the optimal frequency of binucleated cells in the solvent/vehicle control cultures. The appropriate concentration of cytoB is usually between 3 and 6 μg/ml.

Measuring cell proliferation and cytotoxicity and choosing exposure concentrations

22. When determining the highest test substance concentration to be tested, concentrations that have the capability of producing artifactual positive responses, such as those producing excessive cytotoxicity, precipitation in the culture medium, and marked changes in pH or osmolality (39) (40) (41), should be avoided.

23. Measurements of cell proliferation are made to ensure that the treated cells have undergone mitosis during the assay and that the treatments are conducted at appropriate levels of cytotoxicity (see Paragraph 29). Cytotoxicity should be determined with and without metabolic activation in cells that require metabolic activation using the relative increase in cell counts (RICC) or relative population doubling (RPD) (see Appendix 2 for formulas) unless cytoB is used. When cytoB is used, cytotoxicity can be determined using the replication index (RI) (see Appendix 2 for formula).

24. 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 (5), and ensures that only cells that divided during or after treatment are scored.

25. In studies with cytoB, cytostasis/cytotoxicity can be quantified from the cytokinesis-block proliferation index (CBPI) (5) (26) (56) or may be derived from the RI from at least 500 cells per culture (see Appendix 2 for formulas). When cytoB is used to assess cell proliferation, a CBPI or RI should be determined from at least 500 cells per culture. These measurements among others can be used to estimate cytotoxicity by comparing values in the treated and control cultures. Assessment of other markers of cytotoxicity (*e.g.* confluency, cell number, apoptosis, necrosis, metaphase counting) can provide useful information.

26. In studies without cytoB, it is necessary to demonstrate that the cells scored in the culture have undergone division during or following treatment with the test substance, otherwise false negative responses may be produced. Methods that have been used for ensuring that divided cells are being scored include incorporation and subsequent detection of bromodeoxyuridine (BrdU) to identify cells that have replicated (57), the formation of clones when cells from permanent cell lines are treated and scored *in situ* on a microscope slide (Proliferation Index (PI)) (25) (26) (27) (28), or the measurement of Relative Population Doubling (RPD) or Relative Increase in Cell Count (RICC) or other proven methods (16) (56) (58) (59) (see Appendix 2 for formulas). Assessment of other markers for cytotoxicity or cytostasis (*e.g.* confluency, cell number, apoptosis, necrosis, metaphase counting) can provide useful information.

27. At least three analysable test concentrations should be evaluated. In order to achieve this, it may be necessary to perform the experiment using a larger number of closely spaced concentrations and analyse micronucleus formation in those concentrations providing the appropriate range of cytotoxicities. An alternative strategy is to perform a preliminary cytotoxicity test to narrow the range for the definitive test.

28. The highest concentration should aim to produce 55 ± 5% cytotoxicity. Higher levels may induce chromosome damage as a secondary effect of cytotoxicity (60). Where cytotoxicity occurs, the test concentrations selected should cover a range from that producing 55 ± 5% cytotoxicity, to little or no cytotoxicity.

29. If no cytotoxicity or precipitate is observed, the highest test concentration should correspond to 0.01 M, 5 mg/mL or 5 µl/mL, whichever is the lowest. The concentrations selected for analysis should, in general, be separated by a spacing of no more than 10. For test substances that exhibit a steep concentration-response curve, it may be necessary to more closely space the test substance concentrations so that cultures in the moderate and low toxicity ranges also will be scored.

30. When solubility is a limiting factor, the maximum concentration, if not limited by cytotoxicity, should be the lowest concentration at which minimal precipitate is visible in cultures, provided there is no interference with scoring. Evaluation of precipitation should be done by methods such as light microscopy, noting precipitate that persists, or appears during culture (by the end of treatment).

Controls

31. Concurrent positive and solvent/vehicle controls both with and without metabolic activation should be included in each experiment.

32. PC are needed to demonstrate the ability of the cells used, and the test protocol, to identify clastogens and aneugens, and to affirm the metabolic capability of the S9 preparation. The PC should employ known inducers of micronucleus formation at concentrations expected to give small, but reproducible increases over background, and demonstrate the sensitivity of the test system. PC concentrations should be chosen so that the effects are clear but do not immediately reveal the identity of the coded slides to the reader.

33. A clastogen that requires metabolic activation (*e.g.* cyclophosphamide; benzo[a]pyrene) should be used to demonstrate both the metabolic competence and the ability of the test system to detect clastogens. Other PC may be used if justified. Because some PC that need metabolic activation may be active without exogenous metabolic activation under certain treatment conditions or in certain cell lines, the need for metabolic activation, and the activity of the S9 preparation, should be tested in the selected cell line and at the selected concentrations.

34. At the present time, no aneugens are known that require metabolic activation for their genotoxic activity (16). Currently accepted PC for aneugenic activity are, for example, colchicine and vinblastine. Other chemicals may be used if they induce micronuclei solely, or primarily, through aneugenic activity. To avoid the need for two PC (for clastogenicity and aneugenicity) without metabolic activation, the aneugenicity control can serve as the PC without S9, and the clastogenicity control can be used to test the adequacy of the metabolic activation system used. PC for both clastogenicity and aneugenicity should be used in cells that do not require S9. Suggested PC are included in Appendix 3.

35. The use of chemical class-related PC may be considered, when suitable chemicals are available. All PC used should be appropriate for the cell type and activation conditions.

36. Solvent/vehicle controls should be included for every harvest time. In addition, untreated NC (lacking solvent/vehicle) should also be used unless there are published or laboratory historical control data demonstrating that no genotoxic or other deleterious effects are induced by the chosen solvent at the concentrations used.

TEST PROCEDURE

Treatment Schedule

37. 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 are treated with the test substance during all stages of their cell cycles. 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 and these are considered in Paragraphs 41-43 (16).

38. Theoretical considerations, together with published data (18) indicate that most aneugens and clastogens will be detected by a short term treatment period of 3 to 6 hrs in the presence and absence of S9, followed by removal of the test substance and a growth period of 1.5 – 2.0 cell cycles (6). Cells are sampled at a time equivalent to about 1.5 – 2.0 times the normal (*i.e.* untreated) cell cycle length either after the beginning or at the end of treatment (See Table 1). Sampling or recovery times may be extended if it is known or suspected that the test substance affects the cell cycling time (*e.g.* when testing nucleoside analogues).

39. Because of the potential cytotoxicity of S9 preparations for cultured mammalian cells, an extended exposure treatment of 1.5 – 2.0 normal cell cycles is used only in the absence of S9. In the extended treatment, options are offered to allow treatment of the cells with the test chemical in the absence or presence of cytoB. These options address situations where there may be concern regarding possible interactions between the test substance and cytoB.

40. The suggested cell treatment schedules are presented in Table 1. These general treatment schedules may be modified depending on the stability or reactivity of the test substance or the particular growth characteristics of the cells being used. All treatments should commence and end while the cells are growing exponentially. These schedules are presented in more details in paragraphs 41-47 following.

Table 1. Cell treatment and harvest times for the MNvit assay

Lymphocytes, primary cells and cell lines treated with cytoB	+ S9	Treat for 3-6 hrs in the presence of S9; remove the S9 and treatment medium; add fresh medium and cytoB; harvest 1.5 – 2.0 normal cell cycles later.
	– S9 Short exposure	Treat for 3-6 hrs; remove the treatment medium; add fresh medium and cytoB;

		harvest 1.5 – 2.0 normal cell cycles later.
	– S9 Extended exposure	*Option A:* Treat for 1.5 – 2 normal cell cycles in the presence of cytoB; harvest at the end of the exposure period. *Option B:* Treat for 1.5 – 2.0 normal cell cycles; remove the test substance; add fresh medium and cytoB; harvest 1.5 – 2.0 normal cell cycles later.
Cell lines treated without cytoB (Identical to the treatment schedules outlined above with the exception that no cytoB is added)		

Lymphocytes, primary cells, and cell lines with cytoB

41. For lymphocytes, the most efficient approach is to start the exposure to the test substance at 44-48 hrs after PHA stimulation, when cycle synchronisation will have disappeared (5). In the initial assay, cells are treated for 3 to 6 hrs with the test substance in the absence and presence of S9. The treatment medium is removed and replaced with fresh medium containing cytoB, and the cells are harvested 1.5 – 2.0 normal cell cycles later.

42. If both initial tests of the short (3-6 hrs) treatment are negative or equivocal, a subsequent, extended exposure treatment without S9 is used. Two treatment options are available and are equally acceptable. However, it might be more appropriate to follow Option A for stimulated lymphocytes where exponential growth may be declining at 96 hrs following stimulation. Also, cultures of cells should not have reached confluence by the final sampling time in Option B.

- Option A: The cells are treated with the test substance for 1.5 – 2.0 normal cell cycles, and harvested at the end of the treatment time.

- Option B: The cells are treated with the test substance for 1.5 – 2.0 normal cell cycles. The treatment medium is removed and replaced with fresh medium, and the cells are harvested after additional 1.5 - 2.0 normal cell cycles.

43. Primary cells and cell lines should be treated in a similar manner to lymphocytes except that it is not necessary to stimulate them with PHA for 44-48 hrs. Cells other than lymphocytes should be exposed such that at the time of study termination, the cells are still in log-phase growth.

Cell lines without cytoB

44. Cells should be treated for 3-6 hrs in the presence and absence of S9. The treatment medium is removed and replaced with fresh medium, and the cells are harvested 1.5 – 2.0 normal cell cycles later.

45. If both initial tests of the short (3-6 hrs) treatment are negative or equivocal, a subsequent, extended exposure treatment (without S9) is used. Two treatment options are available, both of which are equally acceptable:

- Option A: The cells are treated with the test substance for 1.5 – 2.0 normal cell cycles, and harvested at the end of the treatment time.

- Option B: The cells are treated with the test substance for 1.5 – 2.0 normal cell cycles. The treatment medium is removed and replaced with fresh medium, and the cells are harvested after additional 1.5 - 2.0 normal cell cycles.

46. In monolayers, mitotic cells (identifiable as being round and detaching from the surface) may be present at the end of the 3-6 hr treatment. Because these mitotic cells are easily detached, they can be lost when the medium containing the test substance is removed. Care should be taken to collect these when cultures are washed, and to return them to the cultures, to avoid losing cells that are in mitosis, and at risk for micronuclei, at the time of harvest.

Number of cultures

47. Duplicate cultures should be used for each test substance concentration and for the vehicle/solvent and NC cultures. Where minimal variation between duplicate cultures can be demonstrated from historical laboratory data, it may be acceptable for single cultures to be used. If single cultures are used, it is recommended that an increased number of concentrations be analysed.

Cell harvest and slide preparation

48. Each culture is 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. Cell cytoplasm should be retained to allow the detection of micronuclei and (in the cytokinesis-block method) reliable identification of binucleate cells.

49. The slides can be stained using various methods, such as Giemsa or fluorescent DNA specific dyes (59). The use of a DNA specific stain (*e.g.* acridine orange (61) or Hoechst 33258 plus pyronin-Y (62)) 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 (chromosome/chromosomal fragment) of micronuclei if mechanistic information of their formation is of interest (15)(16). Other methods for differentiation between clastogens and aneugens may be used if they have been shown to be effective.

Analysis

50. All slides, including those of the solvent/vehicle and the controls, should be independently coded before the microscopic analysis. Alternatively, coded samples can be analysed using a validated, automated flow cytometric or image analysis system.

51. In cytoB-treated cultures, micronucleus frequencies should be analysed in at least 2000 binucleated cells per concentration (at least 1000 binucleated cells per culture; two cultures per concentration). If single cultures are used, at least 2000 binucleated cells per concentration should be scored from that culture. If substantially fewer than 1000 binucleate cells per culture, or 2000 if a single culture is used, 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 toxic 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; neither should binucleate cells 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 (63) (64) Scoring of mononucleate cells is acceptable if the test substance is shown to interfere with cytoB activity.

52. In cell lines assayed without cytoB treatment, micronuclei should be scored in at least 2000 cells per concentration (at least 1000 cells per culture; two cultures per concentration). Where only one culture per concentration is used, at least 2000 cells should be scored from that culture.

53. 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 being scored have proliferated, as discussed in Paragraphs 24-27.

Acceptability criteria

54. A laboratory proposing to use the MNvit assay described in this TM should demonstrate its ability to reliably and accurately detect chemicals of known aneugenic and clastogenic activity, with and without metabolic activation, as well as known negative chemicals, using the reference chemicals in Appendix 3. As evidence of its ability to perform this TM correctly, the laboratory should provide evidence that the cells being scored for micronucleus formation have completed one nuclear division if the test is performed without the use of cytoB.

55. The chemicals in Appendix 3 are recommended for use as reference chemicals. Substitute or additional chemicals can be included if their activity is known and if they induce micronuclei by the same mechanisms of action, and if they are shown to be relevant to the chemicals that will be tested using the MNvit procedure. Justification could include a validation study employing a broad variety of substances or focused on a narrower spectrum based on the chemical class of the test substance or the mechanism of damage being studied.

56. Solvent/vehicle control and untreated cultures should give reproducibly low and consistent micronuclei frequencies (typically 5-25 micronuclei/1000 cells for the cell types identified in paragraph 11). Other cell types may have different ranges of responses which should be determined when validating them for use in the MNvit assay. Data from negative,

solvent, and PC should be used to establish historical control ranges. These values should be used in deciding the adequacy of the concurrent NC/PC for an experiment.

57. If minor changes to the protocol (*e.g.* use of automated instead of manual scoring techniques; use of a new cell type) are proposed for the assay, then the effectiveness of the change should be demonstrated before the modified protocol can be considered acceptable for use. Demonstration of effectiveness includes demonstration that the major mechanisms of chromosome breakage and gain or loss can be detected, and that appropriate positive and negative results can be achieved for the class of the individual substance, or the broad range of substances, to be tested.

DATA AND REPORTING

Treatment of results

58. 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. Scoring of the numbers of cells with one, two, or more micronuclei could provide useful information, but is not mandatory.

59. Concurrent measures of cytotoxicity and/or cytostasis for all treated and solvent/vehicle control cultures should be determined (58). 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 or PI should be used (see Appendix 2).

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

61. Chemicals that induce micronuclei in the MNvit assay 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.

Evaluation and interpretation of results

62. There is no requirement for verification by additional testing of a clear positive or negative response. Equivocal results may be clarified by analysis of another 1000 cells from all the cultures to avoid loss of blinding. If this approach does not resolve the result, further testing should be performed. Modification of study parameters over an extended or narrowed range of conditions, as appropriate, should be considered in follow-up experiments. Study parameters that might be modified include the test concentration spacing, the timing of treatment and cell harvest, and/or the metabolic activation conditions.

63. There are several criteria for determining a positive result, such as a concentration-related increase or a statistically significant increase in the number of cells containing micronuclei. The biological relevance of the results should be considered first. Consideration of whether the observed values are within or outside of the historical control range can provide guidance when evaluating the biological significance of the response. Appropriate statistical methods may be used as an aid in evaluating the test results (65). However, the

results of statistical testing should be assessed with respect to dose-response relationship. Reproducibility and historical data should also be taken into consideration.

64. Although most experiments will give clearly positive or negative results, in some cases the data set will preclude making a definite judgement about the activity of the test substance. These equivocal or questionable responses may occur regardless of the number of times the experiment is repeated.

65. Positive results from the MNvit assay indicate that the test substance induces chromosome breakage or chromosome loss, in cultured mammalian cells. Negative results indicate that, under the test conditions used, the test substance does not induce chromosome breaks and/or gain or loss in cultured mammalian cells.

Test Report

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

Test chemical:

- identification data and Chemical Abstract Services Registry Number and EC Number;

- physical nature and purity;

- physico-chemical properties relevant to the conduct of the study;

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

Solvent/Vehicle:

- justification for choice of solvent/vehicle;

- solubility and stability of the test substance in solvent/vehicle;

Cells:

- type and source of cells used;

- suitability of the cell type used;

- absence of mycoplasma, if applicable;

- information on cell cycle length, doubling time or proliferation index;

- where lymphocytes are used, sex, age and number of blood donors, if applicable;

- where lymphocytes are used, whether whole blood or separated lymphocytes are exposed;

- number of passages, if applicable;

- methods for maintenance of cell cultures, if applicable;

- modal number of chromosomes;

- normal (negative control) cell cycle time;

Test Conditions:

- identity of cytokinesis blocking substance (e.g. cytoB), if used, and its concentration and duration of cell exposure;

- rationale for selection of concentrations and number of cultures, including cytotoxicity data and solubility limitations, if available;

- composition of media, CO_2 concentration, if applicable;

- concentrations of test substance;

- concentration (and/or volume) of vehicle and test substance added;

- incubation temperature and time;

- duration of treatment;

- harvest time after treatment;

- cell density at seeding, if applicable;

- type and composition of metabolic activation system, including acceptability criteria;

- positive control chemicals and negative controls;

- methods of slide preparation and staining technique used;

- criteria for micronucleus identification;

- numbers of cells analysed;

- methods for the measurements of cytotoxicity;

- any supplementary information relevant to cytotoxicity;

- criteria for considering studies as positive, negative, or equivocal;

- method(s) of statistical analysis used;

- methods, such as use of kinetochore antibody, to characterise whether micronuclei contain whole or fragmented chromosomes, if applicable;

Results:

- measurement of cytotoxicity used, *e.g.* CBPI or RI in the case of cytokinesis-block method; RICC, RPD or PI when cytokinesis-block methods are not used;

other observations when applicable, *e.g.* cell confluency, apoptosis, necrosis, metaphase counting, frequency of binucleated cells;

- signs of precipitation;
- data on pH and osmolality of the treatment medium, if determined;
- definition of acceptable cells for analysis;
- distribution of mono-, bi-, and multi-nucleated 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/vehicle) and positive control chemical data (concentrations and solvents);
- historical negative (solvent/vehicle) and positive control chemical data, with ranges, means and standard deviation and confidence interval (*e.g.* 95%);
- statistical analysis; p-values if any;

Discussion of the results

Conclusions

LITERATURE

(1) Kirsch-Volders, M. (1997), Towards a validation of the micronucleus test. *Mutation Res.,* 392, 1-4.

(2) Parry, J.M. and Sors, A. (1993), The detection and assessment of the aneugenic potential of environmental chemicals: the European Community aneuploidy project, *Mutation Res.*, 287, 3-15.

(3) Fenech, M. and Morley, A.A. (1985), Solutions to the kinetic problem in the micronucleus assay, *Cytobios.,* 43, 233-246.

(4) Kirsch-Volders, M., Sofuni, T., Aardema, M., Albertini, S., Eastmond, D., Fenech, M., Ishidate, M. Jr, Lorge, E., Norppa, H., Surralles, J., von der Hude, W. and Wakata, A. (2000), Report from the *In Vitro* Micronucleus Assay Working Group, *Environ. Mol. Mutagen*., 35, 167-172.

(5) Fenech, M. (2007), Cytokinesis-block micronucleus cytome assay, *Nature Protocols, 2*(5), 1084-1104.

(6) Fenech, M. and Morley, A.A. (1986), Cytokinesis-block micronucleus method in human lymphocytes: effect of *in-vivo* ageing and low dose X-irradiation, *Mutation Res.,* 161, 193-198.

(7) Eastmond, D.A. and Tucker, J.D. (1989), Identification of aneuploidy-inducing agents using cytokinesis-blocked human lymphocytes and an antikinetochore antibody, *Environ. Mol. Mutagen.,* 13, 34-43.

(8) Eastmond, D.A. and Pinkel, D. (1990), Detection of aneuploidy and aneuploidy-inducing agents in human lymphocytes using fluorescence *in-situ* hybridisation with chromosome-specific DNA probes, *Mutation Res.,* 234, 9-20.

(9) Miller, B.M., Zitzelsberger, H.F., Weier, H.U. and Adler, I.D. (1991), Classification of micronuclei in murine erythrocytes: immunofluorescent staining using CREST antibodies compared to *in situ* hybridization with biotinylated gamma satellite DNA, *Mutagenesis,* 6, 297-302.

(10) Farooqi, Z., Darroudi, F. and Natarajan, A.T. (1993), The use of fluorescence *in-situ* hybridisation for the detection of aneugens in cytokinesis-blocked mouse splenocytes, *Mutagenesis,* 8, 329-334.

(11) Migliore, L., Bocciardi, R., Macri, C. and Lo Jacono, F. (1993), Cytogenetic damage induced in human lymphocytes by four vanadium compounds and micronucleus analysis by fluorescence *in situ* hybridization with a centromeric probe, *Mutation Res.,* 319, 205-213.

(12) Norppa, H., Renzi, L. and Lindholm, C. (1993), Detection of whole chromosomes in micronuclei of cytokinesis-blocked human lymphocytes by antikinetochore staining and *in situ* hybridization, *Mutagenesis,* 8, 519-525.

(13) Eastmond, D.A, Rupa, D.S. and Hasegawa, L.S. (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 Res.,* 322, 9-20.

(14) Marshall, R.R., Murphy, M., Kirkland, D.J. and Bentley, K.S. (1996), Fluorescence *in situ* hybridisation (FISH) with chromosome-specific centromeric probes: a sensitive method to detect aneuploidy, *Mutation Res.,* 372, 233-245.

(15) Zijno, P., Leopardi, F., Marcon, R. and Crebelli, R. (1996), Analysis of chromosome segregation by means of fluorescence *in situ* hybridization: application to cytokinesis-blocked human lymphocytes, *Mutation Res.,* 372, 211-219.

(16) Kirsch-Volders, M., Sofuni, T., Aardema, M., Albertini, S., Eastmond, D., Fenech, M., Ishidate Jr., M., Lorge, E., Norppa, H., Surrallés, J., von der Hude, W. and Wakata, A. (2003), Report from the *in vitro* micronucleus assay working group. *Mutation Res.,* 540, 153-163.

(17) OECD (1997), *In Vitro Mammalian Chromosome Aberration Test,* Test Guideline No. 473, OECD Guidelines for Testing of Chemicals, OECD, Paris. Available at: [www.oecd.org/env/testguidelines]

(18)Lorge, E., Thybaud, V., Aardema, M.J., Oliver, J., Wakata, A., Lorenzon G. and Marzin, D. (2006), SFTG International collaborative Study on *in vitro* micronucleus test. I. General conditions and overall conclusions of the study, *Mutation Res.,* 607, 13-36.

(19) Clare, G., Lorenzon, G., Akhurst, L.C., Marzin, D., van Delft, J., Montero, R., Botta, A., Bertens, A., Cinelli, S., Thybaud, V. and Lorge, E. (2006), SFTG International collaborative study on the *in vitro* micronucleus test. II. Using human lymphocytes, *Mutation Res.,* 607, 37-60.

(20) Aardema, M.J., Snyder, R.D., Spicer, C., Divi, K., Morita, T., Mauthe, R.J., Gibson, D.P., Soelter, S., Curry, P.T., Thybaud, V., Lorenzon, G., Marzin, D. and Lorge, E. (2006), SFTG International collaborative study on the *in vitro* micronucleus test, III. Using CHO cells, *Mutation Res.,* 607, 61-87.

(21) Wakata, A., Matsuoka, A., Yamakage, K., Yoshida, J., Kubo, K., Kobayashi, K., Senjyu, N., Itoh, S., Miyajima, H., Hamada, S., Nishida, S., Araki, H., Yamamura, E., Matsui, A., Thybaud, V., Lorenzon, G., Marzin, D. and Lorge, E. (2006), SFTG International collaborative study on the *in vitro* micronucleus test, IV. Using CHO/IU cells, *Mutation Res.,* 607, 88-124.

(22) Oliver, J., Meunier, J.-R., Awogi, T., Elhajouji, A., Ouldelhkim, M.-C., Bichet, N., Thybaud, V., Lorenzon, G., Marzin, D. and Lorge, E. (2006), SFTG International collaborative study on the *in vitro* micronucleus test, V. Using L5178Y cells, *Mutation Res.,* 607, 125-152.

(23) Albertini, S., Miller, B., Chetelat, A.A. and Locher, F. (1997), Detailed data on *in vitro* MNT and *in vitro* CA: industrial experience, *Mutation Res.,* 392, 187-208.

(24) Miller, B., Albertini, S., Locher, F., Thybaud, V. and Lorge, E. (1997), Comparative evaluation of the *in vitro* micronucleus test and the *in vitro* chromosome aberration test: industrial experience, *Mutation Res.,* 392, 45-59.

(25) Miller, B., Potter-Locher, F., Seelbach, A., Stopper, H., Utesch, D. and Madle, S. (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 fur Umwelt-Mutations-forschung, *Mutation Res.,* 410, 81-116.

(26) Kalweit, S., Utesch, U., von der Hude, W. and Madle, S. (1999), Chemically induced micronucleus formation in V79 cells – comparison of three different test approaches, *Mutation Res.* 439, 183-190.

(27) Kersten, B., Zhang, J., Brendler Schwaab, S.Y., Kasper, P. and Müller, L. (1999), The application of the micronucleus test in Chinese hamster V79 cells to detect drug-induced photogenotoxicity, *Mutation Res.* 445, 55-71.

(28) von der Hude, W., Kalweit, S., Engelhardt, G., McKiernan, S., Kasper, P., Slacik-Erben, R., Miltenburger, H.G., Honarvar, N., Fahrig, R., Gorlitz, B., Albertini, S., Kirchner, S., Utesch, D., Potter-Locher, F., Stopper, H. and Madle, S. (2000), *In vitro* micronucleus assay with Chinese hamster V79 cells - results of a collaborative study with *in situ* exposure to 26 chemical substances, *Mutation Res*., 468, 137-163.

(29) Garriott, M.L., Phelps, J.B. and Hoffman, W.P. (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 Res.,* 517, 123-134.

(30) Matsushima, T., Hayashi, M., Matsuoka, A., Ishidate, M. Jr., Miura, K.F., Shimizu, H., Suzuki, Y., Morimoto, K., Ogura, H., Mure, K., Koshi, K. and Sofuni, T. (1999), Validation study of the *in vitro* micronucleus test in a Chinese hamster lung cell line (CHL/IU), *Mutagenesis,* 14, 569-580.

(31) Elhajouji, A., and Lorge, E. (2006), Special Issue: SFTG International collaborative study on *in vitro* micronucleus test, *Mutation Res.,* 607, 1-152.

(32) 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]

(33) 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]

(34) Corvi, R., Albertini, S., Hartung, T., Hoffmann, S., Maurici, D., Pfuhler, S, van Benthem, J., Vanparys P. (2008), ECVAM Retrospective Validation of *in vitro* Micronucleus Test (MNT), *Mutagenesis,* 23, 271-283.

(35) Zhang, L.S., Honma, M., Hayashi, M., Suzuki, T., Matsuoka, A. and Sofuni, T. (1995), A comparative study of TK6 human lymphoblastoid and L5178Y mouse lymphoma cell lines in the *in vitro* micronucleus test, *Mutation Res.,* 347, 105-115.

(36) Ehrlich, V., Darroudi, F., Uhl, M., Steinkellner, S., Zsivkovits, M. and Knasmeuller, S. (2002), Fumonisin B_1 is genotoxic in human derived hepatoma (HepG2) cells, *Mutagenesis,* 17, 257-260.

(37) Knasmüller, S., Mersch-Sundermann, V., Kevekordes, S., Darroudi, F., Huber, W.W., Hoelzl, C., Bichler, J. and Majer, B.J. (2004), Use of human-derived liver cell lines for the detection of environmental and dietary genotoxicants; current state of knowledge, *Toxicol.,* 198, 315-328. .

(38) Gibson, D.P., Brauninger, R., Shaffi, H.S., Kerckaert, G.A., LeBoeuf, R.A., Isfort, R.J. and Aardema, M.J. (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 Res.,* 392, 61-70.

(39) Scott, D., Galloway, S.M., Marshall, R.R., Ishidate, M. Jr., Brusick, D., Ashby, J. and Myhr, B.C. (1991), International Commission for Protection Against Environmental Mutagens and Carcinogens, Genotoxicity under extreme culture conditions. A report from ICPEMC Task Group 9, *Mutation Res.,* 257, 147-205.

(40) Morita, T., Nagaki, T., Fukuda, I. and Okumura, K. (1992), Clastogenicity of low pH to various cultured mammalian cells, *Mutation Res.,* 268, 297-305.

(41) Brusick, D. (1986), Genotoxic effects in cultured mammalian cells produced by low pH treatment conditions and increased ion concentrations, *Environ. Mutagen*., 8, 789-886.

(42) Fenech, M. and Morley, A.A. (1985), Measurement of micronuclei in lymphocytes, *Mutation Res.,* 147, 29-36.

(43) Fenech, M. (1997), The advantages and disadvantages of cytokinesis-blood micronucleus method, *Mutation Res.,* 392, 11-18.

(44) Bonassi, S., Fenech, M., Lando, C., Lin, Y.P., Ceppi, M., Chang, W.P., Holland, N., Kirsch-Volders, M., Zeiger, E., Ban, S., Barale, R., Bigatti, M.P., Bolognesi, C., Jia, C., Di Giorgio, M., Ferguson, L.R., Fucic, A., Lima, O.G., Hrelia, P., Krishnaja, A.P., Lee, T.K., Migliore, L., Mikhalevich, L., Mirkova, E., Mosesso, P., Muller, W.U., Odagiri, Y., Scarffi, M.R., Szabova, E., Vorobtsova, I., Vral, A. and Zijno, A. (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, *Environ. Mol. Mutagen.* 37, 31-45.

(45) Maron, D.M. and Ames, B.N. (1983), Revised methods for the Salmonella mutagenicity test, *Mutation Res.,* 113, 173-215.

(46) Ong, T.-m., Mukhtar, M., Wolf, C.R. and Zeiger, E. (1980), Differential effects of cytochrome P450-inducers on promutagen activation capabilities and enzymatic activities of S-9 from rat liver, *J. Environ. Pathol. Toxicol.,* 4, 55-65.

(47) Elliott, B.M., Combes, R.D., Elcombe, C.R., Gatehouse, D.G., Gibson, G.G., Mackay, J.M. and Wolf, R.C. (1992), Alternatives to Aroclor 1254-induced S9 in *in-vitro* genotoxicity assays. *Mutagenesis,* 7, 175-177.

(48) Matsushima, T., Sawamura, M., Hara, K. and Sugimura, T. (1976), A safe substitute for Polychlorinated Biphenyls as an Inducer of Metabolic Activation Systems, *In:* de Serres, F.J., Fouts, J. R., Bend, J.R. and Philpot, R.M. (eds), *In Vitro* Metabolic Activation in Mutagenesis Testing, Elsevier, North-Holland, pp. 85-88.

(49) Johnson, T.E., Umbenhauer, D.R. and Galloway, S.M. (1996), Human liver S-9 metabolic activation: proficiency in cytogenetic assays and comparison with phenobarbital/beta-naphthoflavone or Aroclor 1254 induced rat S-9, *Environ. Mol. Mutagen.,* 28, 51-59.

(50) UNEP (2001), Stockholm Convention on Persistent Organic Pollutants, United Nations Environment Programme (UNEP). Available at: [http://www.pops.int/]

(51) Doherty, A.T., Ellard, S., Parry, E.M. and Parry, J.M. (1996), An investigation into the activation and deactivation of chlorinated hydrocarbons to genotoxins in metabolically competent human cells, *Mutagenesis,* 11, 247-274.

(52) Krahn, D.F., Barsky, F.C. and McCooey, K.T. (1982), CHO/HGPRT Mutation Assay: Evaluation of Gases and Volatile Liquids, *In:* Tice, R.R., Costa, D.L. and Schaich, K.M. (eds), Genotoxic Effects of Airborne Agents. New York, Plenum, pp. 91-103.

(53) Zamora, P.O., Benson, J.M., Li, A.P. and Brooks, A.L. (1983), Evaluation of an exposure system using cells grown on collagen gels for detecting highly volatile mutagens in the CHO/HGPRT mutation assay, *Environ. Mutagenesis* 5, 795-801.

(54) Fenech, M. (1993), The cytokinesis-block micronucleus technique: a detailed description of the method and its application to genotoxicity studies in human populations, *Mutation Res.,* 285, 35-44.

(55) Phelps, J.B., Garriott, M.L., and Hoffman, W.P. (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 Res.,* 521, 103-112.

(56) Kirsch-Volders, M., Sofuni, T., Aardema, M., Albertini, S., Eastmond, D., Fenech, M., Ishidate, M. Jr., Kirchner, S., Lorge, E., Morita, T., Norppa, H., Surralles, J., Vanhauwaert, A. and Wakata, A. (2004), Corrigendum to "Report from the *in vitro* micronucleus assay working group", *Mutation Res.,* 564, 97-100.

(57) Pincu, M., Bass, D. and Norman, A. (1984), An improved micronuclear assay in lymphocytes, *Mutation Res.,* 139, 61-65.

(58) Lorge, E., Hayashi, M., Albertini, S. and Kirkland, D. (2008), Comparison of different methods for an accurate assessment of cytotoxicity in the *in vitro* micronucleus test. I. Theoretical aspects, *Mutation Res.,* 655, 1-3.

(59) Surralles, J., Xamena, N., Creus, A., Catalan, J., Norppa, H. and Marcos, R. (1995), Induction of micronuclei by five pyrethroid insecticides in whole-blood and isolated human lymphocyte cultures, *Mutation Res.,* 341, 169-184.

(60) 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, Environ. Molec. *Mutagenesis* 35, 191-201.

(61) Hayashi, M., Sofuni, T., and Ishidate, M. Jr. (1983), An Application of Acridine Orange Fluorescent Staining to the Micronucleus Test, *Mutation Res.,* 120, 241-247.

(62) MacGregor, J. T., Wehr, C. M., and Langlois, R. G. (1983), A Simple Fluorescent Staining Procedure for Micronuclei and RNA in Erythrocytes Using Hoechst 33258 and Pyronin Y, *Mutation Res.,* 120, 269-275.

(63) Hayashi, M., Sofuni, T. and Ishidate, M. Jr. (1983), An application of acridine orange fluorescent staining to the micronucleus test, *Mutation Res.,* 120, 241-247.

(64) Fenech, M., Chang, W.P., Kirsch-Volders, M., Holland, N., Bonassi, S. and Zeiger, E. (2003), HUMN project: detailed description of the scoring criteria for the cytokinesis-block micronucleus assay using isolated human lymphocyte cultures, *Mutation Res.,* 534, 65-75.

(65) Hoffman, W.P., Garriott, M.L. and Lee, C. (2003), *In vitro* micronucleus test, *In:* Encyclopedia of Biopharmaceutical Statistics, Second edition. S. Chow (ed.), Marcel Dekker, Inc. New York, NY, pp. 463-467.

(66) Regulation (EC) No 850/2004 of the European Parliament and of the Council of 29 April 2004 on persistent organic pollutants and amending Directive 79/117/EEC, OJ L 229, 30.4.2004, p. 5.

Appendix 1

DEFINITIONS

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

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

Apoptosis: programmed cell death characterized 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.

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

Clastogen: any substance or process which causes structural chromosomal aberrations in populations of cells or 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: harmful effects to cell structure or function ultimately causing cell death.

Genotoxic: a general term encompassing all types of DNA or chromosome damage, including breaks, adducts rearrangements, 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.

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

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).

Test chemical (also referred to as test substance): Any substance or mixture tested using this TM.

Appendix 2

FORMULAS FOR CYTOTOXICITY ASSESSMENT

1. ***When cytoB is used***, evaluation of cytotoxicity should be based on the **Cytokinesis-Block Proliferation Index (CBPI)** or **Replicative Index (RI)** (16) (58). The CBPI indicates the average number of cell cycles per cell during the period of exposure to cytoB, and may be used to calculate cell proliferation. The RI indicates the relative number of nuclei in treated cultures compared to control cultures and can be used to calculate the % cytostasis:

$$\% \text{ Cytostasis} = 100\text{-}100\{(CBPI_T - 1) \div (CBPI_C - 1)\}$$

And:

T = test chemical treatment culture

C = vehicle control culture

Where:

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

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

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

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

T= treated cultures

C= control cultures

2. 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.

3. ***When cytoB is not used***, evaluation of cytotoxicity based on **Relative Increase in Cell Counts (RICC)** or on **Relative Population Doubling (RPD)** is

recommended (58), as both take into account the proportion of the cell population which has divided.

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

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

where:

Population Doubling = [log (Post-treatment cell number ÷ Initial cell number)] ÷ log 2

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

5. 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)

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

6. The PI has been used as a valuable and reliable cytotoxicity parameter also for cell lines cultured *in situ* in the absence of cytoB (25)(26)(27)(28).

Appendix 3

REFERENCE CHEMICALS RECOMMENDED FOR ASSESSING PERFORMANCE[1]

Category	Chemical	CAS No	EC No
1. Clastogens active without metabolic activation			
	Cytosine arabinoside	147-94-4	205-705-9
	Mitomycin C	50-07-7	200-008-6
2. Clastogens requiring metabolic activation			
	Benzo(a)pyrene	50-32-8	200-028-5
	Cyclophosphamide	50-18-0	200-015-4
3. Aneugens			
	Colchicine	64-86-8	200-598-5
	Vinblastine	143-67-9	205-606-0
4. Negative substances			
	Di(2-ethylhexyl)phthalate	117-81-7	204-211-0
	Nalidixic acid	389-08-2	206-864-7
	Pyrene	129-00-0	204-927-3
	Sodium chloride	7647-14-5	231-598-3

[1.] The reference chemicals are the recommended chemicals for use. Substitution or adding of chemicals to the list of reference chemicals can be done if their activity is known and if they induce micronuclei by the same mechanisms of action, and if they are shown to be relevant to the chemicals that will be tested using the MNvit procedure. Depending on the purpose, justification could also include a validation study employing a broad variety of substances or focused on a narrower spectrum based on the chemical class of the test substance or the mechanism of damage being studied.

B.50 Skin Sensitisation: Local Lymph Node Assay: DA

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals and EU Test Methods are periodically reviewed in light of scientific progress, changing regulatory needs, and animal welfare considerations. The first Test Method (TM) (B.42) for the determination of skin sensitisation in the mouse, the Local Lymph Node Assay (LLNA; OECD Test Guideline 429) has been revised (1) The details of the validation of the LLNA and a review of the associated work have been published (2)(3)(4)(5)(6)(7)(8)(9). In the LLNA, radioisotopic thymidine or iodine is used to measure lymphocyte proliferation and therefore the assay has limited use where the acquisition, use, or disposal of radioactivity is problematic. The LLNA: DA (developed by Daicel Chemical Industries, Ltd.) is a non-radioactive modification to the LLNA, which quantifies adenosine triphosphate (ATP) content via bio-luminescence as an indicator of lymphocyte proliferation. The LLNA: DA test method has been validated and reviewed and recommended by an international peer review panel as considered useful for identifying skin sensitising and non-sensitising chemicals, with certain limitations (10) (11) (12) (13). This TM is designed for assessing skin sensitisation potential of chemicals (substances and mixtures) in animals. Chapter B.6 of this Annex and OECD Test Guideline 406 utilise guinea pig tests, notably the guinea pig maximisation test and the Buehler test (14) The LLNA (chapter B.42 of this Annex; OECD Test Guideline 429) and the two non-radioactive modifications, LLNA: DA (chapter B.50 of this Annex; OECD Test Guideline 442 A) and LLNA: BrdU-ELISA (chapter B.51 of this Annex; OECD Test Guideline 442 B), all provide an advantage over the guinea pig tests in B.6 and OECD Test Guideline 406 (14) in terms of reduction and refinement of animal use.

2. Similar to the LLNA, the LLNA: DA studies the induction phase of skin sensitisation and provides quantitative data suitable for dose-response assessment. Furthermore, an ability to detect skin sensitisers without the necessity for using a radiolabel for DNA eliminates the potential for occupational exposure to radioactivity and waste disposal issues. This in turn may allow for the increased use of mice to detect skin sensitisers, which could further reduce the use of guinea pigs to test for skin sensitisation potential (*i.e.* B.6; OECD Test Guideline 406) (14).

DEFINITIONS

3. Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

4. The LLNA: DA is a modified LLNA method for identifying potential skin sensitising chemicals, with specific limitations. This does not necessarily imply that in all instances the LLNA: DA should be used in place of the LLNA or guinea pig tests (i.e. B.6; OECD Test Guideline 406) (14), but rather that the assay is of equal merit and may be employed as an alternative in which positive and negative results generally no longer require further confirmation (10) (11). The testing laboratory should consider all available information on the test substance prior to conducting the study. Such information will include the identity and chemical structure of the test substance; its physicochemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the test substance; and toxicological data on structurally related chemicals. This information should be considered in order to

determine whether the LLNA: DA is appropriate for the test substance (given the incompatibility of limited types of chemicals with the LLNA: DA [see paragraph 5] and to aid in dose selection.

5. The LLNA: DA is an *in vivo* method and, as a consequence, will not eliminate the use of animals in the assessment of allergic contact sensitising activity. It has, however, the potential to reduce animal use for this purpose when compared to the guinea pig tests (B.6; OECD Test Guideline 406) (14). Moreover, the LLNA: DA offers a substantial refinement (less pain and distress) of the way in which animals are used for allergic contact sensitisation testing, since unlike the B.6 and OECD Test Guideline 406, the LLNA: DA does not require that challenge-induced dermal hypersensitivity reactions be elicited. Despite the advantages of the LLNA: DA over B.6 and OECD Test Guideline 406 (14), there are certain limitations that may necessitate the use of B.6 or OECD Test Guideline 406 (e.g. the testing of certain metals, false positive findings with certain skin irritants [such as some surfactant-type substances] (6) (1 and chapter B.42 in this Annex), solubility of the test substance). In addition, chemical classes or substances containing functional groups shown to act as potential confounders (16) may necessitate the use of guinea pig tests (i.e. B.6; OECD Test Guideline 406 (14)). Limitations that have been identified for the LLNA (1, and chapter B.42 in this Annex) have been recommended to apply also to the LLNA: DA (10). Additionally, the use of the LLNA: DA might not be appropriate for testing substances that affect ATP levels (e.g. substances that function as ATP inhibitors) or those that affect the accurate measurement of intracellular ATP (e.g. presence of ATP degrading enzymes, presence of extracellular ATP in the lymph node). Other than such identified limitations, the LLNA: DA should be applicable for testing any substances unless there are properties associated with these substances that may interfere with the accuracy of the LLNA: DA. In addition, consideration should be given to the possibility of borderline positive results when Stimulation Index (SI) values between 1.8 and 2.5 are obtained (see paragraphs 31-32). This is based on the validation database of 44 substances using an SI $\geq$ 1.8 (see paragraph 6) for which the LLNA: DA correctly identified all 32 LLNA sensitisers, but incorrectly identified three of 12 LLNA non-sensitisers with SI values between 1.8 and 2.5 (i.e. borderline positive) (10). However, as the same dataset was used for setting the SI-values and calculating the predictive properties of the test, the stated results may be an over-estimation of the real predictive properties.

PRINCIPLE OF THE TEST METHOD

6. The basic principle underlying the LLNA: DA is that sensitisers induce proliferation of lymphocytes in the lymph nodes draining the site of test substance application. This proliferation is proportional to the dose and to the potency of the applied allergen and provides a simple means of obtaining a quantitative measurement of sensitisation. Proliferation is measured by comparing the mean proliferation in each test group to the mean proliferation in the vehicle treated control (VC) group. The ratio of the mean proliferation in each treated group to that in the concurrent VC group, termed the SI, is determined, and should be $\geq$1.8 before further evaluation of the test substance as a potential skin sensitiser is warranted. The procedures described here are based on the use of measuring ATP content by bioluminescence (known to correlate with living cell number) (17) to indicate an increased number of proliferating cells in the draining auricular lymph nodes (18) (19). The bioluminescent method utilises the luciferase enzyme to catalyse the formation of light from ATP and luciferin according to the following reaction:

$$ATP + Luciferin + O_2 \xrightarrow{Luciferase} Oxyluciferin + AMP + PP_i + CO_2 + Light$$

The emitted light intensity is linearly related to the ATP concentration and is measured using a luminometer. The luciferin-luciferase assay is a sensitive method for ATP quantitation used in a wide variety of applications (20).

DESCRIPTION OF THE ASSAY

Selection of animal species

7. The mouse is the species of choice for this test. Validation studies for the LLNA: DA were conducted exclusively with the CBA/J strain, which is therefore considered the preferred strain (12) (13). Young adult female mice, which are nulliparous and non-pregnant, are used. At the start of the study, animals should be between 8-12 weeks old, and the weight variation of the animals should be minimal and not exceed 20% of the mean weight. Alternatively, other strains and males may be used when sufficient data are generated to demonstrate that significant strain and/or gender-specific differences in the LLNA: DA response do not exist.

Housing and feeding conditions

8. Mice should be group-housed (21), unless adequate scientific rationale for housing mice individually is provided. The temperature of the experimental animal room should be 22 ± 3ºC. Although the relative humidity should be at least 30% and preferably not exceed 70%, other than during room cleaning, the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water.

Preparation of animals

9. The animals are randomly selected, marked to permit individual identification (but not by any form of ear marking), and kept in their cages for at least five days prior to the start of dosing to allow for acclimatisation to the laboratory conditions. Prior to the start of treatment all animals are examined to ensure that they have no observable skin lesions.

Preparation of dosing solutions

10. Solid chemicals should be dissolved or suspended in solvents/vehicles and diluted, if appropriate, prior to application to an ear of the mice. Liquid chemicals may be applied neat or diluted prior to dosing. Insoluble chemicals, such as those generally seen in medical devices, should be subjected to an exaggerated extraction in an appropriate solvent to reveal all extractable constituents for testing prior to application to an ear of the mice. Test substances should be prepared daily unless stability data demonstrate the acceptability of storage.

Reliability check

11. Positive control chemicals (PC) are used to demonstrate appropriate performance of the assay by responding with adequate and reproducible sensitivity to a sensitising test substance for which the magnitude of the response is well characterised. Inclusion of a concurrent PC is recommended because it demonstrates competency of the laboratory to successfully conduct each assay and allows for an assessment of intra- and inter-laboratory reproducibility and comparability. Some regulatory authorities also require a PC for each study and therefore users are encouraged to consult the relevant authorities prior to conducting the LLNA: DA. Accordingly, the routine use of a concurrent PC is encouraged to

avoid the need for additional animal testing to meet such requirements that might arise from the use of a periodic PC (see paragraph 12). The PC should produce a positive LLNA: DA response at an exposure level expected to give an increase in the SI ≥ 1.8 over the negative control (NC) group. The PC dose should be chosen such that it does not cause excessive skin irritation or systemic toxicity and the induction is reproducible but not excessive (e.g. SI > 10 would be considered excessive). Preferred PC are 25% hexyl cinnamic aldehyde (Chemical Abstracts Service [CAS] number 101-86-0) and 25% eugenol (CAS number 97-53-0,) in acetone: olive oil (4:1, v/v). There may be circumstances in which, given adequate justification, other PC, meeting the above criteria, may be used.

12. While inclusion of a concurrent PC group is recommended, there may be situations in which periodic testing (i.e. at intervals ≤6 months) of the PC may be adequate for laboratories that conduct the LLNA: DA regularly (i.e. conduct the LLNA: DA at a frequency of no less than once per month) and have an established historical PC database that demonstrates the laboratory's ability to obtain reproducible and accurate results with PCs. Adequate proficiency with the LLNA: DA can be successfully demonstrated by generating consistent positive results with the PC in at least 10 independent tests conducted within a reasonable period of time (i.e. less than one year).

13. A concurrent PC group should always be included when there is a procedural change to the LLNA: DA (e.g. change in trained personnel, change in test method materials and/or reagents, change in test method equipment, change in source of test animals), and such changes should be documented in laboratory reports. Consideration should be given to the impact of these changes on the adequacy of the previously established historical database in determining the necessity for establishing a new historical database to document consistency in the PC results.

14. Investigators should be aware that the decision to conduct a PC study on a periodic basis instead of concurrently has ramifications on the adequacy and acceptability of negative study results generated without a concurrent PC during the interval between each periodic PC study. For example, if a false negative result is obtained in the periodic PC study, negative test substance results obtained in the interval between the last acceptable periodic PC study and the unacceptable periodic PC study may be questioned. Implications of these outcomes should be carefully considered when determining whether to include concurrent PCs or to only conduct periodic PCs. Consideration should also be given to using fewer animals in the concurrent PC group when this is scientifically justified and if the laboratory demonstrates, based on laboratory-specific historical data, that fewer mice can be used (22).

15. Although the PC should be tested in the vehicle that is known to elicit a consistent response (*e.g.* acetone: olive oil; 4:1, v/v), there may be certain regulatory situations in which testing in a non-standard vehicle (clinically/chemically relevant formulation) will also be necessary (23). If the concurrent PC is tested in a different vehicle than the test substance, then a separate VC for the concurrent PC should be included.

16. In instances where substances of a specific chemical class or range of responses are being evaluated, benchmark substances may also be useful to demonstrate that the test method is functioning properly for detecting the skin sensitisation potential of these types of substances. Appropriate benchmark substances should have the following properties:

- structural and functional similarity to the class of the test substance being tested;
- known physical chemical characteristics;

- supporting data from the LLNA: DA;
- supporting data from other animal models and/or from humans.

TEST PROCEDURE

Number of animals and dose levels

17. A minimum of four animals is used per dose group, with a minimum of three concentrations of the test substance, plus a concurrent NC group treated only with the vehicle for the test substance, and a PC (concurrent or recent, based on laboratory policy in considering paragraphs 11-15). Testing multiple doses of the PC should be considered, especially when testing the PC on an intermittent basis. Except for absence of treatment with the test substance, animals in the control groups should be handled and treated in a manner identical to that of animals in the treatment groups.

18. Dose and vehicle selection should be based on the recommendations given in references (2) and (24). Consecutive doses are normally selected from an appropriate concentration series such as 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, 0.5%, etc. Adequate scientific rationale should accompany the selection of the concentration series used. All existing toxicological information (e.g. acute toxicity and dermal irritation) and structural and physicochemical information on the test substance of interest (and/or structurally related substances) should be considered, where available, in selecting the three consecutive concentrations so that the highest concentration maximises exposure while avoiding systemic toxicity and/or excessive local skin irritation (24) (25). In the absence of such information, an initial pre-screen test may be necessary (see paragraphs 21-24).

19. The vehicle should not interfere with or bias the test result and should be selected on the basis of maximising the solubility in order to obtain the highest concentration achievable while producing a solution/suspension suitable for application of the test substance. Recommended vehicles are acetone: olive oil (4:1 v/v), N,N-dimethylformamide, methyl ethyl ketone, propylene glycol, and dimethyl sulphoxide (6) but others may be used if sufficient scientific rationale is provided. In certain situations it may be necessary to use a clinically relevant solvent or the commercial formulation in which the test substance is marketed as an additional control. Particular care should be taken to ensure that hydrophilic substances are incorporated into a vehicle system, which wets the skin and does not immediately run off, by incorporation of appropriate solubilisers (e.g. 1% Pluronic® L92). Thus, wholly aqueous vehicles are to be avoided.

20. The processing of lymph nodes from individual mice allows for the assessment of inter-animal variability and a statistical comparison of the difference between test substance and VC group measurements (see paragraph 33). In addition, evaluating the possibility of reducing the number of mice in the PC group is only feasible when individual animal data are collected (22). Further, some regulatory authorities require the collection of individual animal data. Regular collection of individual animal data provides an animal welfare advantage by avoiding duplicate testing that would be necessary if the test substance results originally collected in one manner (e.g. via pooled animal data) were to be considered later by regulatory authorities with other requirements (e.g. individual animal data).

Pre-screen test

21. In the absence of information to determine the highest dose to be tested (see paragraph 18), a pre-screen test should be performed in order to define the appropriate dose level to test in the LLNA: DA. The purpose of the pre-screen test is to provide guidance for selecting the maximum dose level to use in the main LLNA: DA study, where information on the concentration that induces systemic toxicity (see paragraph 24) and/or excessive local skin irritation (see paragraph 23) is not available. The maximum dose level tested should be 100% of the test substance for liquids or the maximum possible concentration for solids or suspensions.

22. The pre-screen test is conducted under conditions identical to the main LLNA: DA study, except there is no assessment of lymph node proliferation and fewer animals per dose group can be used. One or two animals per dose group are suggested. All mice will be observed daily for any clinical signs of systemic toxicity or local irritation at the application site. Body weights are recorded pre-test and prior to termination (Day 8). Both ears of each mouse are observed for erythema and scored using Table 1 (25). Ear thickness measurements are taken using a thickness gauge (*e.g.* digital micrometer or Peacock Dial thickness gauge) on Day 1 (pre-dose), Day 3 (approximately 48 hours after the first dose), Day 7 (24 hours prior to termination) and Day 8. Additionally on Day 8, ear thickness could be determined by ear punch weight determinations, which should be performed after the animals are humanely killed. Excessive local irritation is indicated by an erythema score ≥3 and/or ear thickness of ≥25% on any day of measurement (26) (27). The highest dose selected for the main LLNA: DA study will be the next lower dose in the pre-screen concentration series (see paragraph 18) that does not induce systemic toxicity and/or excessive local skin irritation

Table 1: Erythema Scores

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

23. In addition to a 25% increase in ear thickness (26) (27), a statistically significant increase in ear thickness in the treated mice compared to control mice has also been used to identify irritants in the LLNA (28) (29) (30) (31) (32) (33) (34). However, while statistically significant increases can occur when ear thickness is less than 25% they have not been associated specifically with excessive irritation (30) (31) (32) (33) (34).

24. The following clinical observations may indicate systemic toxicity (35) when used as part of an integrated assessment and therefore may indicate the maximum dose level to use in the main LLNA: DA: changes in nervous system function (*e.g.* pilo-erection, ataxia, tremors, and convulsions); changes in behaviour (*e.g.* aggressiveness, change in grooming activity, marked change in activity level); changes in respiratory patterns (*i.e.* changes in frequency

and intensity of breathing such as dyspnea, gasping, and rales), and changes in food and water consumption. In addition, signs of lethargy and/or unresponsiveness and any clinical signs of more than slight or momentary pain and distress, or a >5% reduction in body weight from Day 1 to Day 8 and mortality, should be considered in the evaluation. Moribund animals or animals showing signs of severe pain and distress should be humanely killed (36).

Main study experimental schedule

25. The experimental schedule of the assay is as follows:

- *Day 1:*

  Individually identify and record the weight of each animal and any clinical observation. Apply 1% sodium lauryl sulfate (SLS) aqueous solution to the dorsum of each ear by using a brush dipped in the SLS solution to cover the entire dorsum of each ear with four to five strokes. One hour after the SLS treatment, apply 25 µL of the appropriate dilution of the test substance, the vehicle alone, or the PC (concurrent or recent, based on laboratory policy in considering paragraphs 11-15), to the dorsum of each ear.

- *Days 2, 3 and 7:*

  Repeat the 1% SLS aqueous solution pre-treatment and test substance application procedure carried out on Day 1.

- *Days 4, 5, and 6:*

  No treatment.

- *Day 8:*

  Record the weight of each animal and any clinical observation. Approximately 24 to 30 hours after the start of application on Day 7, humanely kill the animals. Excise the draining auricular lymph nodes from each mouse ear and process separately in phosphate buffered saline (PBS) for each animal. Details and diagrams of the lymph node identification and dissection can be found in reference (22). To further monitor the local skin response in the main study, additional parameters such as scoring of ear erythema or ear thickness measurements (obtained either by using a thickness gauge, or ear punch weight determinations at necropsy) may be included in the study protocol.

Preparation of cell suspensions

26. From each mouse, a single-cell suspension of lymph node cells (LNC) excised bilaterally is prepared by sandwiching the lymph nodes between two glass slides and applying light pressure to crush the nodes. After confirming that the tissue has spread out thinly pull the two slides apart. Suspend the tissue on both slides in PBS by holding each slide at an angle over the Petri dish and rinsing with PBS while concurrently scraping the tissue off of the slide with a cell scraper. Further, the lymph nodes in NC animals are small, so careful operation is important to avoid any artificial effects on SI values. A total volume of 1 mL PBS should be used for rinsing both slides. The LNC suspension in the Petri dish should be homogenised

lightly with the cell scraper. A 20 µL aliquot of the LNC suspension is then collected with a micropipette, taking care not to take up the membrane that is visible to the eye, and subsequently mixed with 1.98 mL of PBS to yield a 2 mL sample. A second 2 mL sample is then prepared using the same procedure so that two samples are prepared for each animal.

Determination of cellular proliferation (measurement of ATP content of lymphocytes)

27. Increases in ATP content in the lymph nodes are measured by the luciferin/luciferase method using an ATP measurement kit, which measures bioluminescence in Relative Luminescence Units (RLU). The assay time from time of animal sacrifice to measurement of ATP content for each individual animal should be kept uniform, within approximately 30 minutes, because the ATP content is considered to gradually decrease with time after animal sacrifice (12) Thus, the series of procedures from excision of auricular lymph nodes to ATP measurement should be completed within 20 minutes by the pre-determined time schedule that is the same for each animal. ATP luminescence should be measured in each 2 mL sample so that a total of two ATP measurements are collected for each animal. The mean ATP luminescence is then determined and used in subsequent calculations (see paragraph 30).

OBSERVATIONS

Clinical observations

28. Each mouse should be carefully observed at least once daily for any clinical signs, either of local irritation at the application site or of systemic toxicity. All observations are systematically recorded with records being maintained for each mouse. Monitoring plans should include criteria to promptly identify those mice exhibiting systemic toxicity, excessive local skin irritation, or corrosion of skin for euthanasia (36).

Body weights

29. As stated in paragraph 25, individual animal body weights should be measured at the start of the test and at the scheduled humane kill.

CALCULATION OF RESULTS

30. Results for each treatment group are expressed as the mean SI. The SI is derived by dividing the mean RLU/mouse within each test substance group and the PC group by the mean RLU/mouse for the solvent/VC group. The average SI for the VCs is then one.

31. The decision process regards a result as positive when $SI \geq 1.8$ (10) However, the strength of the dose-response relationship, the statistical significance and the consistency of the solvent/vehicle and PC responses may also be used when determining whether a borderline result (*i.e.* SI value between 1.8 and 2.5) is declared positive (2) (3) (37).

32. For a borderline positive response between an SI of 1.8 and 2.5, users may want to consider additional information such as dose-response relationship, evidence of systemic toxicity or excessive irritation, and where appropriate, statistical significance together with SI values to confirm that such results are positives (10). Consideration should also be given to various properties of the test substance, including whether it has a structural relationship to known skin sensitisers, whether it causes excessive skin irritation in the mouse, and the nature of the dose-response relationship observed. These and other considerations are discussed in detail elsewhere (4)

33. Collecting data at the level of the individual mouse will enable a statistical analysis for presence and degree of dose-response relationship in the data. Any statistical assessment could include an evaluation of the dose-response relationship as well as suitably adjusted comparisons of test groups (*e.g.* pair-wise dosed group versus concurrent solvent/vehicle control comparisons). Statistical analyses may include, *e.g.* linear regression or William's test to assess dose-response trends, and Dunnett's test for pair-wise comparisons. In choosing an appropriate method of statistical analysis, the investigator should maintain an awareness of possible inequalities of variances and other related problems that may necessitate a data transformation or a non-parametric statistical analysis. In any case, the investigator may need to carry out SI calculations and statistical analyses with and without certain data points (sometimes called "outliers").

DATA AND REPORTING

Data

34. Data should be summarised in tabular form showing the individual animal RLU values, the group mean RLU/animal, its associated error term (*e.g.* SD, SEM), and the mean SI for each dose group compared against the concurrent solvent/vehicle control group.

Test report

35. The test report should contain the following information:

Test and control chemicals:

- identification data (*e.g.* CAS number and EC number, if available; source; purity; known impurities; lot number);
- physical nature and physicochemical properties (*e.g.* volatility, stability, solubility);
- if mixture, composition and relative percentages of components;

Solvent/vehicle:

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

Test animals:

- source of CBA mice;
- microbiological status of the animals, when known;
- number and age of animals;
- source of animals, housing conditions, diet, etc;

Test conditions:

- the source, lot number and manufacturer's quality assurance/quality control data for the ATP kit;
- details of test substance preparation and application;
- justification for dose selection (including results from pre-screen test, if conducted);

- vehicle and test substance concentrations used, and total amount of test substance applied;
- details of food and water quality (including diet type/source, water source);
- details of treatment and sampling schedules;
- methods for measurement of toxicity;
- criteria for considering studies as positive or negative;
- details of any protocol deviations and an explanation on how the deviation affects the study design and results;

Reliability check:

- a summary of results of latest reliability check, including information on test substance, concentration and vehicle used;
- concurrent and/or historical PC and concurrent negative (solvent/vehicle) control data for testing laboratory;
- if a concurrent PC was not included, the date and laboratory report for the most recent periodic PC and a report detailing the historical PC data for the laboratory justifying the basis for not conducting a concurrent PC;

Results:

- individual weights of mice at start of dosing and at scheduled kill; as well as mean and associated error term (*e.g.* SD, SEM) for each treatment group;
- time course of onset and signs of toxicity, including dermal irritation at site of administration, if any, for each animal;
- time of animal termination and time of ATP measurement for each animal;
- a table of individual mouse RLU values and SI values for each dose treatment group;
- mean and associated error term (*e.g.* SD, SEM) for RLU/mouse for each treatment group and the results of outlier analysis for each treatment group;
- calculated SI and an appropriate measure of variability that takes into account the inter-animal variability in both the test substance and control groups;
- dose response relationship;
- statistical analyses, where appropriate;

Discussion of results:

- a brief commentary on the results, the dose-response analysis, and statistical analyses, where appropriate, with a conclusion as to whether the test substance should be considered a skin sensitiser.

LITERATURE

(1) OECD (2010), *Skin Sensitisation: Local Lymph Node Assay*, Test Guideline No. 429, Guidelines for the Testing of Chemicals, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(2) Chamberlain, M. and Basketter, D.A. (1996), The local lymph node assay: status of validation. *Food Chem, Toxicol*., 34, 999-1002.

(3) Basketter, D.A., Gerberick, G.F., Kimber, I. and Loveless, S.E. (1996), The local lymph node assay: A viable alternative to currently accepted skin sensitisation tests. *Food Chem, Toxicol*., 34, 985-997.

(4) Basketter, D.A., Gerberick, G.F. and Kimber, I. (1998), Strategies for identifying false positive responses in predictive sensitisation tests. *Food Chem. Toxicol.,* 36, 327-333.

(5) Van Och, F.M.M., Slob, W., De Jong, W.H., Vandebriel, R.J. and Van Loveren, H. (2000), A quantitative method for assessing the sensitising potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of uncertainty margins. *Toxicol*., 146, 49-59.

(6) ICCVAM (1999), The murine local lymph node Assay: A test method for assessing the allergic contact dermatitis potential of chemicals/compounds: The results of an independent peer review evaluation coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICETAM). NIH Publication No: 99-4494. Research Triangle Park, N.C. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf]

(7) Dean, J.H., Twerdok, L.E., Tice, R.R., Sailstad, D.M., Hattan, D.G., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: II. Conclusions and recommendations of an independent scientific peer review panel. *Reg. Toxicol. Pharmacol.* 34, 258-273.

(8) Haneke, K.E., Tice, R.R., Carson, B.L., Margolin, B.H., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: III. Data analyses completed by the national toxicology program interagency center for the evaluation of alternative toxicological methods. *Reg. Toxicol. Pharmacol. 34*, 274-286.

(9) Sailstad, D.M., Hattan, D., Hill, R.N., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: I. The ICCVAM review process. *Reg. Toxicol. Pharmacol.* 34, 249-257.

(10) ICCVAM (2010), ICCVAM Test Method Evaluation Report. Nonradioactive local lymph node assay: modified by Daicel Chemical Industries, Ltd., based on ATP content test method protocol (LLNA: DA). NIH Publication No. 10-7551A/B. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/immunotox/llna-DA/TMER.htm]

(11) ICCVAM (2009), Independent Scientific Peer Review Panel Report: Updated validation status of new versions and applications of the murine local lymph node assay: a test method for assessing the allergic contact dermatitis potential of chemicals and products.

Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNAPRPRept2009.pdf].

(12) Idehara, K., Yamagishi, G., Yamashita, K. and Ito, M. (2008), Characterization and evaluation of a modified local lymph node assay using ATP content as a non-radio isotopic endpoint. *J. Pharmacol. Toxicol. Meth.,* 58, 1-10.

(13) Omori, T., Idehara, K., Kojima, H., Sozu, T., Arima, K., Goto, H., Hanada, T., Ikarashi, Y., Inoda, T., Kanazawa, Y., Kosaka, T., Maki, E., Morimoto, T., Shinoda, S., Shinoda, N., Takeyoshi, M., Tanaka, M., Uratani, M., Usami, M., Yamanaka, A., Yoneda, T., Yoshimura, I. and Yuasa, A. (2008), Interlaboratory validation of the modified murine local lymph node assay based on adenosine triphosphate measurement. *J. Pharmacol. Toxicol. Meth.,* 58, 11-26.

(14) OECD (1992), *Skin Sensitisation*, Test Guideline No. 406, Guidelines for Testing of Chemicals, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(15) Kreiling, R., Hollnagel, H.M., Hareng, L., Eigler, L., Lee, M.S., Griem, P., Dreessen, B., Kleber, M., Albrecht, A., Garcia, C. and Wendel, A. (2008), Comparison of the skin sensitising potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). *Food Chem. Toxicol*., 46, 1896-1904.

(16) Basketter, D., Ball, N., Cagen, S., Carrillo, J.C., Certa, H., Eigler, D., Garcia, C., Esch, H., Graham, C., Haux, C., Kreiling, R. and Mehling, A. (2009), Application of a weight of evidence approach to assessing discordant sensitisation datasets: implications for REACH. *Reg. Toxicol. Pharmacol*., 55, 90-96.

(17) Crouch, S.P., Kozlowski, R., Slater, K.J. and Fletcher J. (1993), The use of ATP bioluminescence as a measure of cell proliferation and cytotoxicity. *J. Immunol. Meth*., 160, 81-88.

(18) Ishizaka, A., Tono-oka, T. and Matsumoto, S. (1984), Evaluation of the proliferative response of lymphocytes by measurement of intracellular ATP. *J. Immunol. Meth*., 72, 127-132.

(19) Dexter, S.J., Cámara, M., Davies, M. and Shakesheff, K.M. (2003), Development of a bioluminescent ATP assay to quantify mammalian and bacterial cell number from a mixed population. *Biomat.,* 24, 27-34.

(20) Lundin A. (2000), Use of firefly luciferase in ATP-related assays of biomass, enzymes, and metabolites. *Meth. Enzymol*., 305, 346-370.

(21) ILAR (1996), Institute of Laboratory Animal Research (ILAR) Guide for the Care and Use of Laboratory Animals. 7th ed. Washington, DC: National Academies Press.

(22) ICCVAM (2009), Recommended Performance Standards: Murine Local Lymph Node Assay, NIH Publication Number 09-7357, Research Triangle Park, NC: National Institute of Environmental Health Science. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf]

(23) McGarry, H.F. (2007), The murine local lymph node assay: regulatory and potency considerations under REACH. *Toxicol*., 238, 71-89.

(24) Kimber, I., Dearman, R.J., Scholes E.W. and Basketter, D.A. (1994), The local lymph node assay: developments and applications. *Toxicol*., 93, 13-31.

(25) OECD (2002), *Acute Dermal Irritation/Corrosion*, Test Guideline No. 404, Guidelines for Testing of Chemicals, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(26) Reeder, M.K., Broomhead, Y.L., DiDonato, L. and DeGeorge, G.L. (2007), Use of an enhanced local lymph node assay to correctly classify irritants and false positive substances. *Toxicologist*, 96, 235.

(27) ICCVAM (2009), Nonradioactive Murine Local Lymph Node Assay: Flow Cytometry Test Method Protocol (LLNA: BrdU-FC) Revised Draft Background Review Document. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/immunotox/fcLLNA/BRDcomplete.pdf].

(28) Hayes, B.B., Gerber, P.C., Griffey, S.S. and Meade, B.J. (1998), Contact hypersensitivity to dicyclohexylcarbodiimide and diisopropylcarbodiimide in female B6C3F1 mice. *Drug. Chem. Toxicol*., 21, 195-206.

(29) Homey, B., von Schilling, C., Blumel, J., Schuppe, H.C., Ruzicka, T., Ahr, H.J., Lehmann, P. and Vohr, V.W. (1998), An integrated model for the differentiation of chemical-induced allergic and irritant skin reactions. *Toxicol. Appl. Pharmacol.,* 153, 83-94.

(30) Woolhiser, M.R., Hayes, B.B. and Meade, B.J. (1998), A combined murine local lymph node and irritancy assay to predict sensitisation and irritancy potential of chemicals. *Toxicol. Meth.,* 8, 245-256.

(31) Hayes, B.B. and Meade, B.J. (1999), Contact sensitivity to selected acrylate compounds in B6C3F1 mice: relative potency, cross reactivity, and comparison of test methods. *Drug Chem. Toxicol*., 22, 491-506.

(32) Ehling, G., Hecht, M., Heusener, A., Huesler, J., Gamer, A.O., van Loveren, H., Maurer, T., Riecke, K., Ullmann, L., Ulrich, P., Vandebriel, R. and Vohr, H.W. (2005), A European inter-laboratory validation of alternative endpoints of the murine local lymph node assay: first round. *Toxicol*., 212, 60-68.

(33) Vohr, H.W. and Ahr, H.J. (2005), The local lymph node assay being too sensitive? *Arch. Toxicol*., 79, 721-728.

(34) Patterson, R.M., Noga, E. and Germolec, D. (2007), Lack of evidence for contact sensitisation by Pfiesteria extract. *Environ. Health Perspect*., 115, 1023-1028.

(35) ICCVAM (2009), Report on the ICCVAM-NICEATM/ECVAM/JaCVAM Scientific Workshop on Acute Chemical Safety Testing: Advancing *In Vitro* Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/acutetox/Tox_workshop.htm]

(36) OECD (2000), *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation*, Environmental Health and Safety Monograph Series on Testing and Assessment No. 19, ENV/JM/MONO(2000)7, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(37) Kimber, I., Hilton, J., Dearman, R.J., Gerberick, G.F., Ryan, C.A., Basketter, D.A., Lea, L., House, R.V., Ladies, G.S., Loveless, S.E. and Hastings, K.L. (1998), Assessment of the skin sensitisation potential of topical medicaments using the local lymph node assay: An interlaboratory exercise. *J. Toxicol. Environ. Health,* 53 563-79.

(38) OECD (2005), *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment,* Environment, Health and Safety Monograph Series on Testing and Assessment No. 34, ENV/JM/MONO (2005)14, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

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 (38).

Benchmark substance: A sensitising or non-sensitising substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties; (i) a consistent and reliable source(s); (ii) structural and functional similarity to the class of substances being tested; (iii) known physicochemical characteristics; (iv) supporting data on known effects, and (v) known potency in the range of the desired response.

False negative: A substance incorrectly identified as negative or non-active by a test method, when in fact it is positive or active.

False positive: A substance incorrectly identified as positive or active by a test, when in fact it is negative or non-active.

Hazard: The potential for an adverse health or ecological effect. The adverse effect is manifested only if there is an exposure of sufficient level.

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same test substances, can produce qualitatively and quantitatively similar results. Inter-laboratory reproducibility is determined during the pre-validation and validation processes, and indicates the extent to which a test can be successfully transferred between laboratories, also referred to as between-laboratory reproducibility (38).

Intra-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Also referred to as within-laboratory reproducibility (38).

Outlier: An outlier is an observation that is markedly different from other values in a random sample from a population.

Quality assurance: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures, and the accuracy of data transfer, are assessed by individuals who are independent from those performing the testing.

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 (38).

Skin sensitisation: An immunological process that results when a susceptible individual is exposed topically to an inducing chemical allergen, which provokes a cutaneous immune response that can lead to the development of contact sensitisation.

Stimulation Index (SI): A value calculated to assess the skin sensitisation potential of a test substance that is the ratio of the proliferation in treated groups to that in the concurrent vehicle control group.

Test substance (also referred to as test chemical): Any substance or mixture tested using this TM.

B.51 Skin Sensitisation: Local Lymph Node Assay: BrdU-ELISA

INTRODUCTION

1. OECD Guidelines for the Testing of Chemicals and EU Test Methods are periodically reviewed in light of scientific progress, changing regulatory needs, and animal welfare considerations. The first Test Method (TM) (B.42) for the determination of skin sensitisation in the mouse, the Local Lymph Node Assay (LLNA; OECD Test Guideline 429) has been revised (1, and chapter B.42 in this Annex). The details of the validation of the LLNA and a review of the associated work have been published (2) (3) (4) (5) (6) (7) (8) (9). In the LLNA, radioisotopic thymidine or iodine is used to measure lymphocyte proliferation and therefore the assay has limited use where the acquisition, use, or disposal of radioactivity is problematic. The LLNA: BrdU-ELISA [Enzyme-Linked Immunosorbent Assay] is a non-radioactive modification to the LLNA TM, which utilises non-radiolabelled 5-bromo-2-deoxyuridine (BrdU) (Chemical Abstracts Service [CAS] No 59-14-3) in an ELISA-based test system to measure lymphocyte proliferation. The LLNA: BrdU-ELISA has been validated and reviewed and recommended by an international independent scientific peer review panel as considered useful for identifying skin sensitising and non-sensitising chemicals with certain limitations (10) (11) (12). This TM is designed for assessing skin sensitisation potential of chemicals (substances and mixtures) in animals. Chapter B.6 of this Annex and OECD Test Guideline 406 utilise guinea pig tests, notably the guinea pig maximisation test and the Buehler test (13). The LLNA (chapter B.42 of this Annex; OECD Test Guideline 429) and the two non-radioactive modifications, LLNA: BrdU-ELISA (chapter B.51 of this Annex; OECD Test Guideline 442 B) and LLNA: DA (chapter B.50 of this Annex; OECD Test Guideline 442 A), all provide an advantage over the guinea pig tests in B.6 and OECD Test Guideline 406 (13) in terms of reduction and refinement of animal use.

2. Similar to the LLNA, the LLNA: BrdU-ELISA studies the induction phase of skin sensitisation and provides quantitative data suitable for dose-response assessment. Furthermore, an ability to detect skin sensitisers without the necessity for using a radiolabel for DNA eliminates the potential for occupational exposure to radioactivity and waste disposal issues. This in turn may allow for the increased use of mice to detect skin sensitisers, which could further reduce the use of guinea pigs to test for skin sensitisation potential (*i.e.* B.6; OECD Test Guideline 406) (13).

DEFINITIONS

3. Definitions used are provided in Appendix 1.

INITIAL CONSIDERATIONS AND LIMITATIONS

4. The LLNA: BrdU-ELISA is a modified LLNA method for identifying potential skin sensitising chemicals, with specific limitations. This does not necessarily imply that in all instances the LLNA: BrdU-ELISA should be used in place of the LLNA or guinea pig tests (*i.e.* B.6; OECD Test Guideline 406) (13), but rather that the assay is of equal merit and may be employed as an alternative in which positive and negative results generally no longer require further confirmation (10) (11). The testing laboratory should consider all available information on the test substance prior to conducting the study. Such information will include the identity and chemical structure of the test substance; its physicochemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the test substance; and toxicological data on structurally related chemicals. This information should be considered in order to

determine whether the LLNA: BrdU-ELISA is appropriate for the test substance (given the incompatibility of limited types of chemicals with the LLNA: BrdU-ELISA [see paragraph 5]) and to aid in dose selection.

5. The LLNA: BrdU-ELISA is an *in vivo* method and, as a consequence, will not eliminate the use of animals in the assessment of allergic contact sensitising activity. It has, however, the potential to reduce the animal use for this purpose when compared to the guinea pig tests (B.6; OECD Test Guideline 406) (13). Moreover, the LLNA: BrdU-ELISA offers a substantial refinement of the way in which animals are used for allergic contact sensitisation testing, since unlike the B.6 and OECD Test Guideline 406, the LLNA: BrdU-ELISA does not require that challenge-induced dermal hypersensitivity reactions be elicited. Furthermore, the LLNA: BrdU-ELISA does not require the use of an adjuvant, as is the case for the guinea pig maximisation test (Chapter B.6 of this Annex,13). Thus, the LLNA: BrdU-ELISA reduces animal distress. Despite the advantages of the LLNA: BrdU-ELISA over B.6 and OECD Test Guideline 406 (13), there are certain limitations that may necessitate the use of B.6 or OECD Test Guideline 406 (*e.g.* the testing of certain metals, false positive findings with certain skin irritants [such as some surfactant-type substances] (6) (1, and chapter B.42 in this Annex), solubility of the test substance). In addition, chemical classes or substances containing functional groups shown to act as potential confounders (15) may necessitate the use of guinea pig tests (*i.e.* B.6; OECD Test Guideline 406 (13)). Limitations that have been identified for the LLNA (1, and chapter B.42 in this Annex) have been recommended to apply also to the LLNA: BrdU-ELISA (10). Other than such identified limitations, the LLNA: BrdU-ELISA should be applicable for testing any chemicals unless there are properties associated with these chemicals that may interfere with the accuracy of the LLNA: BrdU-ELISA. In addition, consideration should be given to the possibility of borderline positive results when Stimulation Index (SI) values between 1.6 and 1.9 are obtained (see paragraphs 31-32). This is based on the validation database of 43 substances using an SI $\geq$ 1.6 (see paragraph 6) for which the LLNA: BrdU-ELISA correctly identified all 32 LLNA sensitisers, but incorrectly identified two of 11 LLNA non-sensitisers with SI values between 1.6 and 1.9 (*i.e.* borderline positive) (10). However, as the same dataset was used for setting the SI-values and calculating the predictive properties of the test, the stated results may be an over-estimation of the real predictive properties.

PRINCIPLE OF THE TEST METHOD

6. The basic principle underlying the LLNA: BrdU-ELISA is that sensitisers induce proliferation of lymphocytes in the lymph nodes draining the site of test substance application. This proliferation is proportional to the dose and to the potency of the applied allergen and provides a simple means of obtaining a quantitative measurement of sensitisation. Proliferation is measured by comparing the mean proliferation in each test group to the mean proliferation in the vehicle treated control group (VC). The ratio of the mean proliferation in each treated group to that in the concurrent VC group, termed the SI, is determined, and should be $\geq$1.6 before further evaluation of the test substance as a potential skin sensitiser is warranted. The procedures described here are based on the use of measuring BrdU content to indicate an increased number of proliferating cells in the draining auricular lymph nodes. BrdU is an analogue of thymidine and is similarly incorporated into the DNA of proliferating cells. The incorporation of BrdU is measured by ELISA, which utilises an antibody specific for BrdU that is also labelled with peroxidase. When the substrate is added, the peroxidase reacts with the substrate to produce a coloured product that is quantified at a specific absorbance using a microtitre plate reader.

DESCRIPTION OF THE ASSAY

Selection of animal species

7. The mouse is the species of choice for this test. Validation studies for the LLNA: BrdU-ELISA were conducted exclusively with the CBA/JN strain, which is therefore considered the preferred strain (10) (12). Young adult female mice, which are nulliparous and non-pregnant, are used. At the start of the study, animals should be between 8-12 weeks old, and the weight variation of the animals should be minimal and not exceed 20% of the mean weight. Alternatively, other strains and males may be used when sufficient data are generated to demonstrate that significant strain and/or gender-specific differences in the LLNA: BrdU-ELISA response do not exist.

Housing and feeding conditions

8. Mice should be group-housed (16), unless adequate scientific rationale for housing mice individually is provided. The temperature of the experimental animal room should be 22 ± 3ºC. Although the relative humidity should be at least 30% and preferably not exceed 70%, other than during room cleaning, the aim should be 50-60%. Lighting should be artificial, the sequence being 12 hours light, 12 hours dark. For feeding, conventional laboratory diets may be used with an unlimited supply of drinking water.

Preparation of animals

9. The animals are randomly selected, marked to permit individual identification (but not by any form of ear marking), and kept in their cages for at least five days prior to the start of dosing to allow for acclimatisation to the laboratory conditions. Prior to the start of treatment all animals are examined to ensure that they have no observable skin lesions.

Preparation of dosing solutions

10. Solid chemicals should be dissolved or suspended in solvents/vehicles and diluted, if appropriate, prior to application to an ear of the mice. Liquid chemicals may be applied neat or diluted prior to dosing. Insoluble chemicals, such as those generally seen in medical devices, should be subjected to an exaggerated extraction in an appropriate solvent to reveal all extractable constituents for testing prior to application to an ear of the mice. Test substances should be prepared daily unless stability data demonstrate the acceptability of storage.

Reliability check

11. Positive control chemicals (PC) are used to demonstrate appropriate performance of the assay by responding with adequate and reproducible sensitivity as a sensitising test substance for which the magnitude of the response is well characterised. Inclusion of a concurrent PC is recommended because it demonstrates competency of the laboratory to successfully conduct each assay and allows for an assessment of intra-, and inter-laboratory reproducibility and comparability. Some regulatory authorities also require a PC for each study and therefore users are encouraged to consult the relevant authorities prior to conducting the LLNA: BrdU-ELISA. Accordingly, the routine use of a concurrent PC is encouraged to avoid the need for additional animal testing to meet such requirements that might arise from the use of a periodic PC (see paragraph 12). The PC should produce a positive LLNA: BrdU-ELISA response at an exposure level expected to give an increase in

the SI ≥ 1.6 over the negative control (NC) group. The PC dose should be chosen such that it does not cause excessive skin irritation or systemic toxicity and the induction is reproducible but not excessive (*e.g.* SI > 14 would be considered excessive). Preferred PC are 25% hexyl cinnamic aldehyde (CAS No 101-86-0) and 25% eugenol (CAS No 97-53-0) in acetone: olive oil (4:1, v/v). There may be circumstances in which, given adequate justification, other PC, meeting the above criteria, may be used.

12. While inclusion of a concurrent PC group is recommended, there may be situations in which periodic testing *(i.e.* at intervals ≤6 months) of the PC may be adequate for laboratories that conduct the LLNA: BrdU-ELISA regularly (*i.e.* conduct the LLNA: BrdU-ELISA at a frequency of no less than once per month) and have an established historical PC database that demonstrates the laboratory's ability to obtain reproducible and accurate results with PCs. Adequate proficiency with the LLNA: BrdU-ELISA can be successfully demonstrated by generating consistent positive results with the PC in at least 10 independent tests conducted within a reasonable period of time (*i.e.* less than one year).

13. A concurrent PC group should always be included when there is a procedural change to the LLNA: BrdU-ELISA (*e.g.* change in trained personnel, change in test method materials and/or reagents, change in test method equipment, change in source of test animals), and such changes should be documented in laboratory reports. Consideration should be given to the impact of these changes on the adequacy of the previously established historical database in determining the necessity for establishing a new historical database to document consistency in the PC results.

14. Investigators should be aware that the decision to conduct a PC study on a periodic basis instead of concurrently has ramifications on the adequacy and acceptability of negative study results generated without a concurrent PC during the interval between each periodic PC study. For example, if a false negative result is obtained in the periodic PC study, negative test substance results obtained in the interval between the last acceptable periodic PC study and the unacceptable periodic PC study may be questioned. Implications of these outcomes should be carefully considered when determining whether to include concurrent PCs or to only conduct periodic PCs. Consideration should also be given to using fewer animals in the concurrent PC group when this is scientifically justified and if the laboratory demonstrates, based on laboratory-specific historical data, that fewer mice can be used (17).

15. Although the PC should be tested in the vehicle that is known to elicit a consistent response (*e.g.* acetone: olive oil; 4:1, v/v), there may be certain regulatory situations in which testing in a non-standard vehicle (clinically/chemically relevant formulation) will also be necessary (18). If the concurrent PC is tested in a different vehicle than the test substance, then a separate VC for the concurrent PC should be included.

16. In instances where test substances of a specific chemical class or range of responses are being evaluated, benchmark substances may also be useful to demonstrate that the test method is functioning properly for detecting the skin sensitisation potential of these types of test substances. Appropriate benchmark substances should have the following properties:

- structural and functional similarity to the class of the test substance being tested;
- known physical chemical characteristics;
- supporting data from the LLNA: BrdU-ELISA;
- supporting data from other animal models and/or from humans.

TEST PROCEDURE

Number of animals and dose levels

17. A minimum of four animals is used per dose group, with a minimum of three concentrations of the test substance, plus a concurrent NC group treated only with the vehicle for the test substance, and a PC group (concurrent or recent, based on laboratory policy in considering paragraphs 11- 15). Testing multiple doses of the PC should be considered especially when testing the PC on an intermittent basis. Except for absence of treatment with the test substance, animals in the control groups should be handled and treated in a manner identical to that of animals in the treatment groups.

18. Dose and vehicle selection should be based on the recommendations given in the references 2 and 19. Consecutive doses are normally selected from an appropriate concentration series such as 100%, 50%, 25%, 10%, 5%, 2.5%, 1%, 0.5%, etc. Adequate scientific rationale should accompany the selection of the concentration series used. All existing toxicological information (*e.g.* acute toxicity and dermal irritation) and structural and physicochemical information on the test substance of interest (and/or structurally related substances) should be considered, where available, in selecting the three consecutive concentrations so that the highest concentration maximises exposure while avoiding systemic toxicity and/or excessive local skin irritation (19)(20 and chapter B.4 of this Annex). In the absence of such information, an initial pre-screen test may be necessary (see paragraphs 21-24).

19. The vehicle should not interfere with or bias the test result and should be selected on the basis of maximising the solubility in order to obtain the highest concentration achievable while producing a solution/suspension suitable for application of the test substance. Recommended vehicles are acetone: olive oil (4:1 v/v), *N,N*-dimethylformamide, methyl ethyl ketone, propylene glycol, and dimethyl sulphoxide (6) but others may be used if sufficient scientific rationale is provided. In certain situations it may be necessary to use a clinically relevant solvent or the commercial formulation in which the test substance is marketed as an additional control. Particular care should be taken to ensure that hydrophilic test substances are incorporated into a vehicle system, which wets the skin and does not immediately run off, by incorporation of appropriate solubilisers (*e.g.* 1% Pluronic® L92). Thus, wholly aqueous vehicles are to be avoided.

20. The processing of lymph nodes from individual mice allows for the assessment of inter-animal variability and a statistical comparison of the difference between test substance and VC group measurements (see paragraph 33). In addition, evaluating the possibility of reducing the number of mice in the PC group is only feasible when individual animal data are collected (17). Further, some regulatory authorities require the collection of individual animal data. Regular collection of individual animal data provides an animal welfare advantage by avoiding duplicate testing that would be necessary if the test substance results originally collected in one manner (*e.g.* via pooled animal data) were to be considered later by regulatory authorities with other requirements (*e.g.* individual animal data).

Pre-screen test

21. In the absence of information to determine the highest dose to be tested (see paragraph 18), a pre-screen test should be performed in order to define the appropriate dose level to test in the LLNA: BrdU-ELISA. The purpose of the pre-screen test is to provide guidance for

selecting the maximum dose level to use in the main LLNA: BrdU-ELISA study, where information on the concentration that induces systemic toxicity (see paragraph 24) and/or excessive local skin irritation (see paragraph 23) is not available. The maximum dose level tested should be a concentration of 100% of the test substance for liquids or the maximum possible concentration for solids or suspensions.

22. The pre-screen test is conducted under conditions identical to the main LLNA: BrdU-ELISA study, except there is no assessment of lymph node proliferation and fewer animals per dose group can be used. One or two animals per dose group are suggested. All mice will be observed daily for any clinical signs of systemic toxicity or local irritation at the application site. Body weights are recorded pre-test and prior to termination (Day 6). Both ears of each mouse are observed for erythema and scored using Table 1 (20, and chapter B.4 of this Annex). Ear thickness measurements are taken using a thickness gauge (*e.g.* digital micrometer or Peacock Dial thickness gauge) on Day 1 (pre-dose), Day 3 (approximately 48 hours after the first dose), and Day 6. Additionally, on Day 6, ear thickness could be determined by ear punch weight determinations, which should be performed after the animals are humanely killed. Excessive local irritation is indicated by an erythema score ≥3 and/or ear thickness of ≥25% on any day of measurement (21) (22). The highest dose selected for the main LLNA: BrdU-ELISA study will be the next lower dose in the pre-screen concentration series (see paragraph 18) that does not induce systemic toxicity and/or excessive local skin irritation.

Table 1: Erythema Scores

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

23. In addition to a 25% increase in ear thickness (21) (22), a statistically significant increase in ear thickness in the treated mice compared to control mice has also been used to identify irritants in the LLNA (22) (23) (24) (25) (26) (27) (28). However, while statistically significant increases can occur when ear thickness is less than 25% they have not been associated specifically with excessive irritation (25) (26) (27) (28) (29).

24. The following clinical observations may indicate systemic toxicity (30) when used as part of an integrated assessment and therefore may indicate the maximum dose level to use in the main LLNA: BrdU-ELISA: changes in nervous system function (*e.g.* pilo-erection, ataxia, tremors, and convulsions); changes in behaviour (*e.g.* aggressiveness, change in grooming activity, marked change in activity level); changes in respiratory patterns (*i.e.* changes in frequency and intensity of breathing such as dyspnea, gasping, and rales), and changes in food and water consumption. In addition, signs of lethargy and/or unresponsiveness and any clinical signs of more than slight or momentary pain and distress, or a >5% reduction in body weight from Day 1 to Day 6 and mortality should be considered in the evaluation. Moribund animals or animals showing signs of severe pain and distress should be humanely killed (31).

Main study experimental schedule

25. The experimental schedule of the assay is as follows:

- *Day 1:*

  Individually identify and record the weight of each animal and any clinical observation. Apply 25 µL of the appropriate dilution of the test substance, the vehicle alone, or the PC (concurrent or recent, based on laboratory policy in considering paragraphs 11-15), to the dorsum of each ear.

- *Days 2 and 3:*

  Repeat the application procedure carried out on Day 1.

- *Day 4:*

  No treatment.

- *Day 5:*

  Inject 0.5 mL (5 mg/mouse) of BrdU (10 mg/mL) solution intra-peritoneally.

- *Day 6:*

  Record the weight of each animal and any clinical observation. Approximately 24 hours (24 h) after BrdU injection, humanely kill the animals. Excise the draining auricular lymph nodes from each mouse ear and process separately in phosphate buffered saline (PBS) for each animal. Details and diagrams of the lymph node identification and dissection can be found in reference (17). To further monitor the local skin response in the main study, additional parameters such as scoring of ear erythema or ear thickness measurements (obtained either by using a thickness gauge, or ear punch weight determinations at necropsy) may be included into the study protocol.

Preparation of cell suspensions

26. From each mouse, a single-cell suspension of lymph node cells (LNC) excised bilaterally is prepared by gentle mechanical disaggregation through 200 micron-mesh stainless steel gauze or another acceptable technique for generating a single-cell suspension (*e.g.* use of a disposable plastic pestle to crush the lymph nodes followed by passage through a #70 nylon mesh). The procedure for preparing the LNC suspension is critical in this assay and therefore every operator should establish the skill in advance. Further, the lymph nodes in NC animals are small, so careful operation is important to avoid any artificial effects on SI values. In each case, the target volume of the LNC suspension should be adjusted to a determined optimised volume (approximately 15 mL). The optimised volume is based on achieving a mean absorbance of the NC group within 0.1- 0.2.

Determination of cellular proliferation (measurement of BrdU content in DNA of lymphocytes)

27. BrdU is measured by ELISA using a commercial kit (*e.g.* Roche Applied Science, Mannheim, Germany, Catalogue Number 11 647 229 001). Briefly, 100 µL of the LNC suspension is added to the wells of a flat-bottom microplate in triplicate. After fixation and denaturation of the LNC, anti-BrdU antibody is added to each well and allowed to react. Subsequently the anti-BrdU antibody is removed by washing and the substrate solution is then added and allowed to produce chromogen. Absorbance at 370 nm with a reference wavelength of 492 nm is then measured. In all cases, assay test conditions should be optimised (see paragraph 26).

OBSERVATIONS

Clinical observations

28. Each mouse should be carefully observed at least once daily for any clinical signs, either of local irritation at the application site or of systemic toxicity. All observations are systematically recorded with records being maintained for each mouse. Monitoring plans should include criteria to promptly identify those mice exhibiting systemic toxicity, excessive local skin irritation, or corrosion of skin for euthanasia (31).

Body weights

29. As stated in paragraph 25, individual animal body weights should be measured at the start of the test and at the scheduled humane kill.

CALCULATION OF RESULTS

30. Results for each treatment group are expressed as the mean SI. The SI is derived by dividing the mean BrdU labelling index/mouse within each test substance group and the PC group by the mean BrdU labelling index for the solvent/VC group. The average SI for the VCs is then one.

The BrdU labelling index is defined as:

$$\text{BrdU labelling index} = (ABS_{em} - ABS\ blank_{em}) - (ABS_{ref} - ABS\ blank_{ref})$$

Where; em = emission wavelength; and ref = reference wavelength.

31. The decision process regards a result as positive when SI $\geq$ 1.6 (10). However, the strength of the dose-response relationship, the statistical significance and the consistency of the solvent/vehicle and PC responses may also be used when determining whether a borderline result (*i.e.* SI value between 1.6 and 1.9) is declared positive (3) (6) (32).

32. For a borderline positive response between an SI of 1.6 and 1.9, users may want to consider additional information such as dose-response relationship, evidence of systemic toxicity or excessive irritation, and where appropriate, statistical significance together with SI values to confirm that such results are positives (10). Consideration should also be given to various properties of the test substance, including whether it has a structural relationship to known skin sensitisers, whether it causes excessive skin irritation in the mouse, and the nature

of the dose-response observed. These and other considerations are discussed in detail elsewhere (4).

33. Collecting data at the level of the individual mouse will enable a statistical analysis for presence and degree of dose-response relationship in the data. Any statistical assessment could include an evaluation of the dose-response relationship as well as suitably adjusted comparisons of test groups (*e.g.* pair-wise dosed group versus concurrent solvent/vehicle control comparisons). Statistical analyses may include, *e.g.* linear regression or William's test to assess dose-response trends, and Dunnett's test for pair-wise comparisons. In choosing an appropriate method of statistical analysis, the investigator should maintain an awareness of possible inequalities of variances and other related problems that may necessitate a data transformation or a non-parametric statistical analysis. In any case, the investigator may need to carry out SI calculations and statistical analyses with and without certain data points (sometimes called "outliers").

DATA AND REPORTING

Data

34. Data should be summarised in tabular form showing the individual animal BrdU labelling index values, the group mean BrdU labelling index/animal, its associated error term (*e.g.* SD, SEM), and the mean SI for each dose group compared against the concurrent solvent/vehicle control group.

Test report

35. The test report should contain the following information:

Test and control chemicals:

- identification data (*e.g.* CAS number and EC number, if available; source; purity; known impurities; lot number);
- physical nature and physicochemical properties (*e.g.* volatility, stability, solubility);
- if mixture, composition and relative percentages of components;

Solvent/vehicle:

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

Test animals:

- source of CBA mice;
- microbiological status of the animals, when known;
- number and age of animals;
- source of animals, housing conditions, diet, etc.;

Test conditions:

- source, lot number, and manufacturer's quality assurance/quality control data (antibody sensitivity and specificity and the limit of detection) for the ELISA kit;
- details of test substance preparation and application;
- justification for dose selection (including results from pre-screen test, if conducted);
- vehicle and test substance concentrations used, and total amount of test substance applied;
- details of food and water quality (including diet type/source, water source);
- details of treatment and sampling schedules;
- methods for measurement of toxicity;
- criteria for considering studies as positive or negative;
- details of any protocol deviations and an explanation on how the deviation affects the study design and results;

Reliability check:

- a summary of results of latest reliability check, including information on test substance, concentration and vehicle used;
- concurrent and/or historical PC and concurrent negative (solvent/vehicle) control data for testing laboratory;
- if a concurrent PC was not included, the date and laboratory report for the most recent periodic PC and a report detailing the historical PC data for the laboratory justifying the basis for not conducting a concurrent PC;

Results:

- individual weights of mice at start of dosing and at scheduled humane kill; as well as mean and associated error term (*e.g.* SD, SEM) for each treatment group;
- time course of onset and signs of toxicity, including dermal irritation at site of administration, if any, for each animal;
- a table of individual mouse BrdU labelling indices and SI values for each treatment group;
- mean and associated error term (*e.g.* SD, SEM) for BrdU labelling index/mouse for each treatment group and the results of outlier analysis for each treatment group;
- calculated SI and an appropriate measure of variability that takes into account the inter-animal variability in both the test substance and control groups;
- dose-response relationship;
- statistical analyses, where appropriate;

Discussion of results:

- a brief commentary on the results, the dose-response analysis, and statistical analyses, where appropriate, with a conclusion as to whether the test substance should be considered a skin sensitiser.

LITERATURE

(1) OECD (2010), Skin Sensitisation: Local Lymph Node Assay, Test Guideline No. 429, Guidelines for the Testing of Chemicals, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(2) Chamberlain, M. and Basketter, D.A. (1996), The local lymph node assay: status of validation. Food Chem. Toxicol., 34, 999-1002.

(3) Basketter, D.A., Gerberick, G.F., Kimber, I. and Loveless, S.E. (1996), The local lymph node assay: A viable alternative to currently accepted skin sensitisation tests. Food Chem. Toxicol., 34, 985-997.

(4) Basketter, D.A., Gerberick, G.F. and Kimber, I. (1998), Strategies for identifying false positive responses in predictive sensitisation tests. Food Chem. Toxicol., 36, 327-33.

(5) Van Och, F.M.M., Slob, W., De Jong, W.H., Vandebriel, R.J. and Van Loveren, H. (2000), A quantitative method for assessing the sensitising potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of uncertainty margins. Toxicol., 146, 49-59.

(6) ICCVAM (1999), The murine local lymph node Assay: A test method for assessing the allergic contact dermatitis potential of chemicals/compounds: The results of an independent peer review evaluation coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICETAM). NIH Publication No: 99-4494. Research Triangle Park, N.C. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna/llnarep.pdf]

(7) Dean, J.H., Twerdok, L.E., Tice, R.R., Sailstad, D.M., Hattan, D.G., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: II. Conclusions and recommendations of an independent scientific peer review panel. Reg. Toxicol. Pharmacol., 34(3), 258-273.

(8) Haneke, K.E., Tice, R.R., Carson, B.L., Margolin, B.H., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: III. Data analyses completed by the national toxicology program interagency center for the evaluation of alternative toxicological methods. Reg. Toxicol. Pharmacol., 34(3), 274-286.

(9) Sailstad, D.M., Hattan, D., Hill, R.N., Stokes, W.S. (2001), ICCVAM evaluation of the murine local lymph node assay: I. The ICCVAM review process. Reg. Toxicol. Pharmacol., 34(3), 249-257.

(10) ICCVAM (2010), ICCVAM Test Method Evaluation Report. Nonradioactive local lymph node assay: BrdU-ELISA Test Method Protocol (LLNA: BrdU-ELISA). NIH Publication No. 10-7552A/B. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/immunotox/llna-ELISA/TMER.htm]

(11) ICCVAM (2009), Independent Scientific Peer Review Panel Report: Updated validation status of new versions and applications of the murine local lymph node assay: a test method for assessing the allergic contact dermatitis potential of chemicals and products. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNAPRPRept2009.pdf]

(12) Takeyoshi, M., Iida, K., Shiraishi, K. and Hoshuyama, S. (2005), Novel approach for classifying chemicals according to skin sensitising potency by non-radioisotopic modification of the local lymph node assay. J. Appl. Toxicol., 25, 129-134.

(13) OECD (1992), Skin Sensitisation, Test Guideline No. 406, Guidelines for Testing of Chemicals, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(14) Kreiling, R., Hollnagel, H.M., Hareng, L., Eigler, L., Lee, M.S., Griem, P., Dreessen, B., Kleber, M., Albrecht, A., Garcia, C. and Wendel, A. (2008), Comparison of the skin sensitising potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). Food Chem. Toxicol., 46, 1896-1904.

(15) Basketter, D., Ball, N., Cagen, S., Carrilo, J.C., Certa, H., Eigler, D., Garcia, C., Esch, H., Graham, C., Haux, C., Kreiling, R. and Mehling, A. (2009), Application of a weight of evidence approach to assessing discordant sensitisation datasets: implications for REACH. Reg. Toxicol. Pharmacol., 55, 90-96.

(16) ILAR (1996), Institute of Laboratory Animal Research (ILAR) Guide for the Care and Use of Laboratory Animals. 7th ed. Washington, DC: National Academies Press.

(17) ICCVAM (2009), Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/docs/immunotox_docs/llna-ps/LLNAPerfStds.pdf]

(18) McGarry, H.F. (2007), The murine local lymph node assay: regulatory and potency considerations under REACH. Toxicol., 238, 71-89.

(19) Kimber, I., Dearman, R.J., Scholes E.W. and Basketter, D.A. (1994), The local lymph node assay: developments and applications. Toxicol., 93, 13-31.

(20) OECD (2002), Acute Dermal Irritation/Corrosion, Test Guideline No. 404, Guidelines for Testing of Chemicals, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(21) Reeder, M.K., Broomhead, Y.L., DiDonato, L. and DeGeorge, G.L. (2007), Use of an enhanced local lymph node assay to correctly classify irritants and false positive substances. Toxicologist, 96, 235.

(22) ICCVAM (2009), Nonradioactive Murine Local Lymph Node Assay: Flow Cytometry Test Method Protocol (LLNA: BrdU-FC) Revised Draft Background Review Document. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/immunotox/fcLLNA/BRDcomplete.pdf].

(23) Hayes, B.B., Gerber, P.C., Griffey, S.S. and Meade, B.J. (1998), Contact hypersensitivity to dicyclohexylcarbodiimide and diisopropylcarbodiimide in female B6C3F1 mice. Drug Chem. Toxicol., 21, 195-206.

(24) Homey, B., von Schilling, C., Blumel, J., Schuppe, H.C., Ruzicka, T., Ahr, H.J., Lehmann, P. and Vohr, V.W. (1998), An integrated model for the differentiation of chemical-induced allergic and irritant skin reactions. Toxicol. Appl. Pharmacol., 153, 83-94.

(25) Woolhiser, M.R., Hayes, B.B. and Meade, B.J. (1998), A combined murine local lymph node and irritancy assay to predict sensitisation and irritancy potential of chemicals. Toxicol. Meth., 8, 245-256.

(26) Hayes, B.B. and Meade, B.J. (1999), Contact sensitivity to selected acrylate compounds in B6C3F1 mice: relative potency, cross reactivity, and comparison of test methods. Drug. Chem. Toxicol., 22, 491-506.

(27) Ehling, G., Hecht, M., Heusener, A., Huesler, J., Gamer, A.O., van Loveren, H., Maurer, T., Riecke, K., Ullmann, L., Ulrich, P., Vandebriel, R. and Vohr, H.W. (2005), A European inter- laboratory validation of alternative endpoints of the murine local lymph node assay: first round. Toxicol., 212, 60-68.

(28) Vohr, H.W. and Ahr, H.J. (2005), The local lymph node assay being too sensitive? Arch. Toxicol., 79, 721-728.

(29) Patterson, R.M., Noga, E. and Germolec D. (2007), Lack of evidence for contact sensitisation by Pfiesteria extract. Environ. Health Perspect., 115, 1023-1028.

(30) ICCVAM (2009), Report on the ICCVAM-NICEATM/ECVAM/JaCVAM Scientific Workshop on Acute Chemical Safety Testing: Advancing *In Vitro* Approaches and Humane Endpoints for Systemic Toxicity Evaluations. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: [http://iccvam.niehs.nih.gov/methods/acutetox/Tox_workshop.htm].

(31) OECD (2000), Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation, Environmental Health and Safety Monograph Series on Testing and Assessment No. 19, ENV/JM/MONO(2000)7, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

(32) Kimber, I., Hilton, J., Dearman, R.J., Gerberick, G.F., Ryan, C.A., Basketter, D.A., Lea, L., House, R.V., Ladies, G.S., Loveless, S.E. and Hastings, K.L. (1998), Assessment of the skin sensitisation potential of topical medicaments using the local lymph node assay: An interlaboratory exercise. J. Toxicol. Environ.l Health, 53, 563-79.

(33) OECD (2005), Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, Environment, Health and Safety Monograph Series on Testing and Assessment No. 34, ENV/JM/MONO(2005)14, OECD, Paris. Available at: [http://www.oecd.org/env/testguidelines]

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 (33).

Benchmark substance: A sensitising or non-sensitising substance used as a standard for comparison to a test substance. A benchmark substance should have the following properties: (i) a consistent and reliable source(s); (ii) structural and functional similarity to the class of substances being tested; (iii) known physical/chemical characteristics; (iv)supporting data on known effects; and (v) known potency in the range of the desired response.

False negative: A test substance incorrectly identified as negative or non-active by a test method, when in fact it is positive or active (33).

False positive: A test substance incorrectly identified as positive or active by a test, when in fact it is negative or non-active (33).

Hazard: The potential for an adverse health or ecological effect. The adverse effect is manifested only if there is an exposure of sufficient level.

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same test substance, can produce qualitatively and quantitatively similar results. Inter-laboratory reproducibility is determined during the pre-validation and validation processes, and indicates the extent to which a test can be successfully transferred between laboratories, also referred to as between-laboratory reproducibility (33).

Intra-laboratory reproducibility: A determination of the extent that qualified people within the same laboratory can successfully replicate results using a specific protocol at different times. Also referred to as within-laboratory reproducibility (33).

Outlier: An outlier is an observation that is markedly different from other values in a random sample from a population.

Quality assurance: A management process by which adherence to laboratory testing standards, requirements, and record keeping procedures, and the accuracy of data transfer, are assessed by individuals who are independent from those performing the testing.

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 (33).

Skin sensitisation: An immunological process that results when a susceptible individual is exposed topically to an inducing chemical allergen, which provokes a cutaneous immune response that can lead to the development of contact sensitisation.

Stimulation Index (SI): A value calculated to assess the skin sensitisation potential of a test substance that is the ratio of the proliferation in treated groups to that in the concurrent vehicle control group.

Test substance (also referred to as test chemical): Any substance or mixture tested using this TM."