

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

DRAFT PROPOSAL FOR AN UPDATE OF TEST GUIDELINE 431

***In Vitro* Skin Corrosion: Reconstructed Human Epidermis (RhE) Test Method**

INTRODUCTION

1. Skin corrosion refers to the production of irreversible damage to the skin manifested as visible necrosis through the *epidermis* and into the *dermis*, following the application of a test material [as defined by the United Nations (UN) Globally Harmonised System of Classification and Labelling of Chemicals (GHS)] (1). This Test Guideline provides an *in vitro* procedure allowing the identification of corrosive chemical substances and mixtures.

2. The assessment of skin corrosivity has typically involved the use of laboratory animals (OECD Test Guideline 404 (TG 404); adopted in 1981 and revised in 1992 and 2002)(2). In relation to animal welfare TG 404 was revised in 2002, allowing for the determination of skin corrosion by applying a tiered testing strategy, using validated *in vitro* or *ex vivo* test methods, thus avoiding pain and suffering of animals. In addition to TG 431 (originally adopted in 2004)(3), two other *in vitro* test methods for testing of corrosivity have been validated and adopted as OECD Test Guidelines 430 (4) and 435 (5).

3. This Test Guideline is based on a Reconstructed Human *Epidermis* (RhE) model, which in its overall design (the use of human derived epidermal keratinocytes as cell source and use of representative tissue and cyto-architecture) closely mimics the biochemical and physiological properties of the upper parts of the human skin, *i.e.* the *epidermis*. This updated Test Guideline also includes a set of Performance Standards (PS)(Annex 1) for the assessment of similar and modified RhE-based test methods (6), in accordance with the principles of Guidance Document No. 34 (7).

4. There are two validated test methods that adhere to this Test Guideline using a RhE model, commercially available as EpiSkin™ and EpiDerm™ (EPI-200) RhE Test methods (designated the Validated Reference Methods-VRMs). Prevalidation studies (8), followed by a formal validation study of these two *in vitro* test methods for assessing skin corrosion (9)(10) have been conducted (11)(12). The outcome of these studies and other published literature (13) led to the recommendation that the following two VRMs could be used for regulatory purposes for the assessment of *in vivo* skin corrosivity (14)(15)(16).

5. Before a proposed similar or modified *in vitro* RhE test method other than the VRMs can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined to ensure its similarity to the VRMs, in accordance with the requirements of the PS set out in this Test Guideline (Annex 1). Any proposed new or updated test method following the PS of this Test Guideline should be submitted to the OECD for adoption and inclusion into the Test Guideline before being used for regulatory purposes.

DEFINITIONS

6. Definitions used are provided in Annex 2.

INITIAL CONSIDERATIONS

7. The test method described in this Test Guideline allows the identification of corrosive chemical substances and mixtures in accordance with the UN GHS. It further enables the identification of non-corrosive substances and mixtures when supported by a weight of evidence determination using other existing information (*e.g.* pH, structure-activity relationships, human and/or animal data) (1)(2)(17). It does not normally provide adequate information on skin irritation, nor does it allow the sub-categorisation of corrosive substances as permitted in the UN GHS (1). The Test Guideline is expected to be generally applicable across chemical classes (9)(10)(13) and to solids, liquids, and semi-solids. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other prior treatment of the sample is required. Gases and aerosols have not been assessed yet in validation studies (10)(13)(18). While it is conceivable that these can be tested using RhE technology, the current Test Guideline does not allow testing of gases and aerosols.”

8. This Test Guideline also includes a set of Performance Standards (PS) (Annex 1) for determining the performance (reliability and relevance) of similar and modified skin corrosion test methods that are structurally and mechanistically similar to the VRMs (6), in accordance with the principles of Guidance Document No. 34 (7). These PS include a list of 24 reference chemicals by which to evaluate assay performance, the essential test method components that should be included in the protocol for the test method to be considered structurally and mechanistically similar, and the minimum reliability and accuracy values necessary for the test method to be considered comparable to the VRMs. Within the reference chemical list, a subset of 12 proficiency chemicals (Annex 1, Table 1) is provided that can be used by laboratories to demonstrate proficiency in using *in vitro* human skin models.

9. A limitation of the Test Guideline, as demonstrated by the validation study (9)(10)(11)(12), is that it does not allow the sub-categorisation of corrosives in accordance with the UN GHS (1). Thus, the regulatory framework in member countries will decide how this Test Guideline will be used. For a full evaluation of local skin effects after a single dermal exposure, it is recommended to follow the sequential testing strategy as appended to TG 404 (2). This testing strategy includes the conduct of *in vitro* tests for skin corrosion (as described in this Test Guideline) and skin irritation before considering testing in live animals. It is recognized that the use of human skin is subject to national and international ethical considerations and conditions.

PRINCIPLE OF THE TEST

10. The test material 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. The RhE test method is based on the premise that corrosive substances are able to penetrate the *stratum corneum* by diffusion or erosion, and are cytotoxic to the cells in the underlying layers. Cell viability is measured by dehydrogenase conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1], into a blue

formazan salt that is quantitatively measured after extraction from tissues (19). Corrosive substances are identified by their ability to decrease cell viability below defined threshold levels (*i.e.* ≤ xx%).

DEMONSTRATION OF PROFICIENCY

12. Prior to routine use of any of the two validated RhE test method that adheres to this Test Guideline, laboratories should demonstrate technical proficiency, using the twelve Proficiency Chemicals recommended in Annex 1 (Table 1). For similar or modified test methods developed under this Test Guideline that are structurally and mechanistically similar to the RhE test method, the PS requirements described in Annex 1 of this Test Guideline should be used to demonstrate similar or better performance of the test method prior to its use for regulatory purposes.

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

PROCEDURE

14. The following is a generic description of the components and procedures of a RhE test method for skin corrosion assessment similar to the VRMs. A RhE model should be reconstructed, and can be in-house prepared or obtained commercially, *e.g.* the EpiDerm™ and EPISKIN™ models (19) (20) (21) (22) (23) (24). Standard Operating procedures for the EpiSkin™ and EpiDerm™ test methods are available (25) (26). Any new or modified model should meet the performance criteria in the Performance Standards. Human skin models used for this test method should comply with the following:

RHE TEST METHOD COMPONENTS

General Conditions

15. Non-transformed human keratinocytes should be used to reconstruct the epithelium. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum*, *stratum granulosum*) should be present under a functional *stratum corneum*. The *stratum corneum* should be 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% (IC₅₀) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET₅₀) upon application of the 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

16. The assay used for determining the magnitude of viability is the MTT-assay (27). 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 < xx**. An acceptability range (upper and lower limit) for the negative control OD values are established by the RhE model developer/supplier, and the acceptability ranges for the VRMs are given in Table 1. 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 1. Acceptability ranges for negative control OD values?

	Lower acceptance limit	Higher acceptance limit
EpiSkin™	≥xx	≤xx
EpiDerm™	≥xx	≤xx

Barrier function

17. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of certain cytotoxic marker chemicals (*e.g.* SDS or Triton X-100), as estimated by IC₅₀ or ET₅₀ (*e.g.* for the EpiDerm™ and EPISKIN™ models this is xx hours). The tissue should demonstrate reproducibility over time and between laboratories. Moreover it should be capable of predicting the corrosive potential of the proficiency chemicals (Table 1) when used in the testing protocol selected.

Morphology

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

Reproducibility

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

Quality control (QC)

20. 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 16), *barrier function* (paragraph 17) and *morphology* (paragraph 18) are the most relevant. These data should be provided to the test method users, so that they are able to include this information in the test report. An acceptability range (upper and lower limit) for the IC₅₀ or the ET₅₀ should be established by the RhE model developer/supplier (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 test methods are given in Table 2.

Table 2. Examples of QC batch release criteria

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM) (x hours treatment with SDS)(REF)	IC ₅₀ = xx mg/ml	IC ₅₀ = xx mg/ml

EpiDerm™ SIT (EPI-200) (x% Triton X-100)(REF)	ET ₅₀ = xx hr	ET ₅₀ = xx hr
--	--------------------------	--------------------------

Application of the Test and Control Substances

21. At least two tissue replicates should be used for each test substance and exposure the control in each run. For liquid as well as solid chemicals, sufficient amount of test substance should be applied to uniformly cover the epidermis surface while avoiding an infinite dose, *i.e.* a minimum of 25 µL/cm² or 25 mg/cm² should be used. For solid substances, the epidermis surface should be moistened with deionised or distilled water before application, to improve contact between the test substance and the epidermis surface. Whenever possible, solids should be tested as a fine powder. The application method should be appropriate for the test substance (see *e.g.* reference 10). At the end of the exposure period, the test material should be carefully washed from the *epidermis* with an aqueous buffer, or 0.9% NaCl.

22. Concurrent NC and positive controls (PC) should be used in each run to demonstrate that viability, barrier function and resulting tissue sensitivity of the tissues are within a defined historical acceptance range. The suggested PC chemicals are glacial acetic acid or 8N KOH. The suggested NCs are 0.9% NaCl or water.

Cell Viability Measurements

23. The MTT assay is a validated quantitative assay which should be used to measure cell viability under this Test Guideline (19)(27). It should be documented that the tissue treated with NC is stable in culture (provide similar viability measurements) for the duration of the test exposure period.

24. Optical properties of the test substance or its chemical action on the MTT may interfere with the assay leading to a false estimate of viability (because the test substance may prevent or reverse the colour generation as well as cause it). This may occur when a specific test substance is not completely removed from the tissue by rinsing or when it penetrates the *epidermis*. If the test substance acts directly on the MTT, is naturally coloured, or becomes coloured during tissue treatment, additional controls should be used to detect and correct for test substance 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 VRMs [ref to be added]. Non specific colour (NSC) due to these interferences should not exceed 30% of NC (for corrections). If NSC > 30%, the test substance is considered as incompatible with the test method.

Interpretation of Results and Prediction Model

25. The OD values obtained for each test sample can be used to calculate percent viability relative to the NC, which is arbitrarily set at 100%. The cut-off percentage cell viability value distinguishing corrosive from non-corrosive test materials (or discriminating between different corrosive classes), or the statistical procedure(s) used to evaluate the results and identify corrosive materials, should be clearly defined and documented, and be shown to be appropriate. In general, these cut-off values are established during test method optimisation, tested during a prevalidation phase, and confirmed in a validation study. The prediction model for corrosivity associated with the EpiDerm™ model is (13):

The test substance is considered to be corrosive to skin:

- i) if the viability after 3 minutes exposure is less than 50%, or
- ii) if the viability after 3 minutes exposure is greater than or equal to 50 % and the viability after 60 minutes exposure is less than 15%.

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

- i) if the viability after 3 minutes exposure is greater than or equal to 50% and the viability after 60 minutes exposure is greater than or equal to 15%.

The prediction model for corrosivity associated with the EpiSkin™ model is (XX):

The test substance is considered to be corrosive to skin:

- i) if the viability after 3 minutes exposure is less than 35%, or
- ii) if the viability after 3 minutes exposure is greater than or equal to 35 % and the viability after 60 minutes exposure is less than 35%.

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

- i) if the viability after 240 minutes exposure is greater than or equal to 35%.

DATA AND REPORTING

Data

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

Test report

27. The test report should include the following information:

Test and Control Substances:

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

Justification of the RhE model and protocol used:

Test Conditions:

- Cell system used;
- Calibration information for measuring device, and band pass used for measuring cell viability

(e.g. spectrophotometer);

- 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

Results:

- Tabulation of data from individual test substances;
- Description of other effects observed

Discussion of the results

Quality assurance statement for Good Laboratory Practice compliant studies:

statement should indicate all inspections made during the study and the dates any results were reported to the Study Director. The statement should also confirm that the final report reflects the raw data.

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) 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>]
- (3) OECD (2004), *In vitro* skin model. OECD Guideline for the Testing of Chemicals No. 431, OECD, Paris. Available at: [<http://www.oecd.org/env/testguidelines>]
- (4) 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>]
- (5) OECD (2006), *In vitro* membrane barrier test method. OECD Guideline for the Testing of Chemicals No. 435, OECD, Paris. Available at: [<http://www.oecd.org/env/testguidelines>]
- (6) ICCVAM (2004), Recommended Performance Standards for *In Vitro* Test Methods for Skin Corrosion. NIH Publication Number 04-4510. Research Triangle Park, NC: National Institute of Environmental Health Sciences. Available at: http://iccvam.niehs.nih.gov/docs/dermal_docs/ps/ps044510.pdf
- (7) 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>]
- (8) Botham, P.A., Chamberlain, M., Barratt, M.D., Curren, R.D., Esdaile, D.J., Gardner, J.R., Gordon, V.C., Hildebrand, B., Lewis, R.W., Liebsch, M., Logemann, P., Osborne, R., Ponc, M., Regnier, J.F., Steiling, W., Walker, A.P., and Balls, M. (1995), A prevalidation study on *in vitro* skin corrosivity testing. The report and recommendations of ECVAM Workshop 6. *ATLA* 23, 219-255.
- (9) Barratt, M.D., Brantom, P.G., Fentem, J.H., Gerner, I., Walker, A.P., and Worth, A.P. (1998), The ECVAM international validation study on *in vitro* tests for skin corrosivity. 1. Selection and distribution of the test chemicals. *Toxicol. in Vitro* 12, 471-482.
- (10) Fentem, J.H., Archer, G.E.B., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., Holzhutter, H.-G., and Liebsch, M. (1998), The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the Management Team. *Toxicol. in Vitro* 12, 483-524.
- (11) Balls, M., Blaauboer, B.J., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H., and Zucco, F. (1995), Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM workshops. *ATLA* 23, 129-147.
- (12) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (1997), Validation and Regulatory Acceptance of Toxicological Test Methods. NIH Publication No. 97-

3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA. <http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf>.
- (13) Liebsch, M., Traue, D., Barrabas, C., Spielmann, H., Uphill, P., Wilkins, S., McPherson, J.P., Wiemann, C., Kaufmann, T., Remmele, M. and Holzhütter, H.-G. (2000), The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. *ATLA* 28, pp. 371-401.
- (14) ECVAM (1998), ECVAM News & Views. *ATLA* 26, 275-280.
- (15) ECVAM (2000), ECVAM News & Views. *ATLA* 28, 365-67.
- (16) ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (2002), ICCVAM evaluation of EpiDerm™, EPISKIN™ (EPI-200), and the Rat Skin Transcutaneous Electrical Resistance (TER) assay: In Vitro test methods for assessing dermal corrosivity potential of chemicals. NIH Publication No. 02-4502. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA. Available: [http://iccvam.niehs.nih.gov/methods/epiddocs/epis_brd.pdf]
- (17) Worth, A.P., Fentem, J.H., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., and Liebsch, M. (1998), An Evaluation of the Proposed OECD Testing Strategy for Skin Corrosion. *ATLA* 26: 709-720.
- (18) Kandarova, H. (2006),
- (19) Tinois, E., Gaetani, Q., Gayraud, B., Dupont, D., Rougier, A., and Pouradier, D.X. (1994), The Episkin model: Successful reconstruction of human *epidermis in vitro*. In: *In vitro* Skin Toxicology. Edited by A. Rougier, A.M. Goldberg and H.I. Maibach: 133-140.
- (20) Cannon, C. L., Neal, P. J., Southee, J. A., Kubilus, J., and Klausner, M., (1994), New epidermal model for dermal irritancy testing. *Toxicol. in Vitro* 8, 889 - 891.
- (21) Ponec, M., Boelsma, E., Weerheim, A., Mulder, A., Bouwstra, J., and Mommaas, M., (2000), Lipid and ultrastructural characterization of reconstructed skin models. *Inter. J. Pharmaceu.* 203, 211 - 225.
- (22) Tinois, E., Tiollier, J., Gaucherand, M., Dumas, H., Tardy, M., and Thivolet, J. (1991), *In vitro* and post –transplantation differentiation of human keratinocytes grown on the human type IV collagen film of a bilayered dermal substitute. *Exp. Cell Res.* 193: 310-319.
- (23) Parenteau, N.L., Bilbo, P., Molte, C.J., Mason, V.S., and Rosenberg, H. (1992), The organotypic culture of human skin keratinocytes and fibroblasts to achieve form and function. *Cytotech.* 9, 163-171.
- (24) Wilkins, L.M., Watson, S.R., Prosky, S.J., Meunier, S.F., Parenteau, N.L. (1994), Development of a bilayered living skin construct for clinical applications. *Biotech. Bioeng.* 43/8, 747-756.
- (25) SOP EpiSkin
- (26) SOP EpiDerm

- (27) Mosmann, T. (1983), Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods* 65, 55-63.
- (28) Marshall, N.J., Goodwin, C.J., Holt, S.J. (1995), A critical assessment of the use of microculture tetrazolium assays to measure cell growth and function. *Growth Reg.* 5, 69-84.

ANNEX 1PERFORMANCE STANDARDS FOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED IN VITRO RHE TEST METHODS¹

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 reliability and relevance (accuracy) for specific testing purposes. These PS, based on validated and accepted test methods, can be used to evaluate the reliability and relevance (accuracy) of other analogous test methods (colloquially referred to as “me-too” test methods) that are based on similar scientific principles and measure or predict the same biological or toxic effect (7).

2. Prior to adoption of modified test 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 methods 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 method, or, if appropriate, to a limited assessment of reliability and accuracy using established PS (7).

3. Similar (me-too) or modified test methods of any of the two VRMs (EPISKINTM and EpiDermTM (EPI-200)) proposed for use under this Test Guideline should be evaluated to determine their reliability and accuracy using reference chemicals (Table 1) representing the full range of the TG 404 *in vivo* corrosivity scores.

4. These PS are based on the US-ICCVAM PS (6) for evaluating the validity of new or modified RhE test methods. The PS consists of: (i) essential test method components; (ii) recommended reference chemicals (including a set of proposed proficiency chemicals), and; (iii) defined reliability and relevance (accuracy) values that the proposed test method should meet or exceed.

1) Essential Test Method Components

5. These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed, mechanistically and functionally similar or modified test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a similar or modified proposed test method is based on the same concepts as the corresponding VRMs. The essential test method components are described in detail in paragraphs 16 to 24 of the Test Guideline and testing should be performed according to the following:

The general conditions (paragraph 15)

Functional conditions, which include:

Viability (paragraph 16)

Barrier function (paragraph 17)

Morphology (paragraph 18)

Reproducibility (paragraph 19)

Quality control (paragraph 20)

¹ Please note that any proposed new or updated test method following the PS of this Test Guideline should be submitted to the OECD for adoption and inclusion into the Test Guideline before being used for regulatory purposes.

If any of these criteria are not met, then these performance standards cannot be used for validation of the new or modified test method.

II) Minimum List of Reference Chemicals

6. Reference Chemicals are used to determine if the reliability and accuracy of a proposed similar or modified test method, proven to be structurally and functionally sufficiently similar to the VRMs, or representing a minor modification of one of the VRMs, are comparable or better than those of the VRMs. The 24 recommended Reference Chemicals listed in Table 1 include substances representing different chemical classes (*i.e.* chemical categories based on functional groups), and are representative of the full range of TG 404 *in vivo* scores. The substances included in this list comprise 24 UN GHS Category 1A/1b/1C respectively. The substances listed in Table 1 are selected from the substances used in the validation study of the VRMs, with regard to chemical functionality and physical state. These Reference Chemicals represent the minimum number of chemicals that should be used to evaluate the reliability and relevance (accuracy) of a proposed similar or modified test method. In situations where a listed substance is unavailable, other substances for which adequate *in vivo* reference data are available could be used, primarily from the substances used in the validation study of the VRMs. If desired, additional substances 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 test method.

Table 1. Minimum List of Reference and Proficiency Chemicals for Determination of Reliability and Accuracy Values for Similar of Modified *In Vitro* HrE Model Corrosivity Test Methods. (Recommended Proficiency Chemicals are Indicated in Bold)

Chemical ¹	CASRN	Chemical Class ²	UN GHS <i>In Vivo</i>	Solid/Liquid	pH ³
<i>In Vivo</i> Corrosives					
Phosphorus tribromide	7789-60-8	inorganic acid	1b	L	1.0
Sulfuric acid (10%)	7664-93-9	inorganic acid	1b/1c	L	1.2
Boron trifluoride dihydrate	13319-75-0	inorganic acid	1b	L	1.5
Octanoic (Caprylic) acid	124-07-2	organic acid	1b/1c	L	3.6
2-tert-Butylphenol	88-18-6	phenol	1b/1c	L	3.9
Hexanoic acid	142-62-1	organic acid	1b/1c	L	3.9
Dimethyldipropylenetriamine	10563-29-8	organic base	1a	L	8.3
Dimethylisopropylamine	996-35-0	organic base	1b/1c	L	8.3
1,2-Diaminopropane	78-90-0	organic base	1a	L	8.3
n-Heptylamine	111-68-2	organic base	1b/1c	L	8.4
Potassium hydroxide (10% aq.)	1310-58-3	inorganic base	1b	L	13.1
Phosphorus pentachloride	10026-13-8	inorganic acid	1a	S	ND
<i>In Vivo</i> Non-corrosives					
Sulfamic acid	5329-14-6	inorganic acid	NC	S	1.5
Isostearic acid	30399-84-9	organic acid	NC	L	3.6
Phenethyl bromide	103-63-9	electrophile	NC	L	3.6
Eugenol	97-53-0	phenol	NC	L	3.7
1,9-Decadiene	1647-16-1	neutral organic	NC	L	3.9
<i>o</i> -Methoxyphenol	90-05-1	phenol	NC	L	3.9
Sodium lauryl sulfate (20% aq.)	151-21-3	surfactant	NC	L	3.9
Tetrachloroethylene	127-18-4	neutral organic	NC	L	4.5
4-Amino-1,2,4-triazole	584-13-4	organic base	NC	S	5.5
4-(methylthio)-Benzaldehyde	3446-89-7	electrophile	NC	L	6.8
Sodium carbonate (50% aq.)	7664-93-9	inorganic base	NC	L	11.7
Dodecanoic acid (lauric acid)	143-07-7	organic acid	NC	S	ND

Abbreviations: aq = aqueous; CASRN = Chemical Abstracts Service Registry Number; UN GHS= United Nations Globally Harmonised System (1).

¹These substances, sorted first by corrosives versus non-corrosives and then by pH, were selected from among the 60 substances used by EU-ECVAM to validate EPISKIN™ (9) (10). Unless otherwise indicated, the substances were tested at the purity level obtained when purchased from a commercial source (9). The goal of the selection process is to include, to the extent possible, substances that: are representative of the range of corrosivity responses (*e.g.* non-corrosives; weak to strong corrosives) that the validated reference test method is capable of measuring or predicting; are representative of the chemical classes used in the validation process; reflect the performance characteristics of the validated reference test

method; have chemical structures that are well-defined; induce reproducible results in the validated reference test method; induce definitive results in the *in vivo* reference test method; are commercially available; and are not associated with prohibitive disposal costs.

²Chemical class assigned by Barratt et al. (1998) (9).

³The pH values were obtained from Fentem et al. (1998) (10) and Barratt et al. (1998)(9).

III) Defined Reliability and Accuracy Values

7. For purposes establishing the reliability and relevance (accuracy) of proposed similar or modified test methods to be transferred between laboratories, all 24 Reference Chemicals should be tested in at least three laboratories. However, if the proposed test 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.

8. The calculation of the reliability and accuracy values of the proposed test 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 test 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 test 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 24 Reference Chemicals by the different participating laboratories.

In this context, a **run sequence** consists of two 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 both 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 variability should show a concordance of classifications obtained in different, independent test runs of the 24 Reference Chemicals within one single laboratory equal or higher (\geq) than 90%.

Between-laboratory reproducibility

11. An assessment of between-laboratory reproducibility is not essential if the proposed test method is to be used in a single laboratory only. For methods to be transferred between laboratories, the concordance of classifications obtained in different, independent test runs of the 24 Reference Chemicals between preferentially a minimum of three laboratories should be equal or higher (\geq) than 80%.

Predictive capacity (accuracy)

12. The accuracy (sensitivity, specificity and overall accuracy) of the proposed similar or modified test method should be comparable or better to that of the VRMs. The sensitivity and specificity should be equal or higher (\geq) than xx%, respectively. The overall accuracy should be equal or higher (\geq) than xx%.

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

Sensitivity	Specificity	Overall Accuracy
xx%	xx%	xx%

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 valid runs required for each tested chemical is not obtained for every Reference Chemicals 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 for purposes of similar validation studies:

1. All 24 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 24 chemicals: x invalid run sequences 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 24 chemicals tested in 3 laboratories: x invalid run sequences allowed in total).

In this context, a **run sequence** consists of x 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 runs are valid. This means that any single invalid run invalidates an entire run sequence. In terms of cell viability measurements, the median CV should not exceed 35% for studies conducted in different laboratories (10) (16). The median CV for replicate studies conducted in the same laboratory should be less than the median CV for studies conducted in different laboratories.

ANNEX 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.

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 sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemicals being examined (9).

ET₅₀: 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 IC₅₀.

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

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

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).

Mixture: Used in the context of the UN GHS (1) as a mixture or solution composed of two or more chemicals in which they do not react.

NC: Negative Control

OD: Optical Density

PC: Positive Control

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

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.

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.

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

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

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test. 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.

Tiered testing strategy: Testing which uses test methods in a sequential manner; the test methods selected in each succeeding level are determined by the results in the previous level of testing.

UN GHS (United Nations Globally Harmonized System of Classification and Labelling): 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).