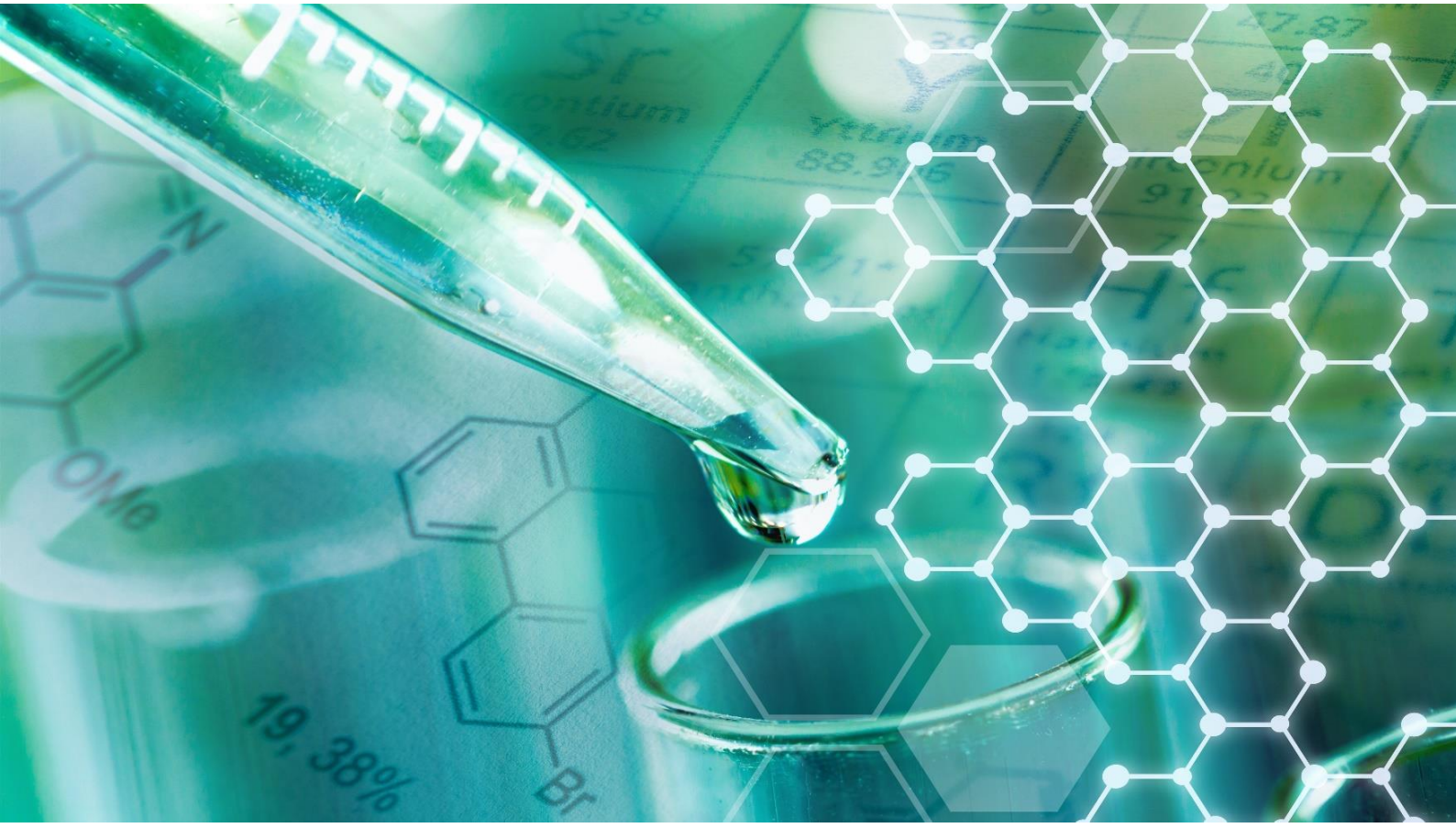
Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Epidermal Sensitisation Assay (EpiSensA) Test Methods

Series on Testing and Assessment No. 396



OECD

Series on Testing and Assessment
No. 396

Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Epidermal Sensitisation Assay (EpiSensA) Test Methods

2024

Please cite this publication as:

OECD (2024), *Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Epidermal Sensitisation Assay (EpiSensA) Test Methods*, OECD Series on Testing and Assessment, No. 396, OECD Publishing, Paris.

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 countries in North and South America, Europe and the Asia and Pacific region, as well as the European Union, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several Partner countries and from interested international organisations attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (https://www.oecd.org/en/topics/chemical-safety-and-biosafety.html).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank, Basel, Rotterdam and Stockholm Conventions and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

PERFORMANCE STANDARDS FOR THE ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED IN VITRO EPIDERMAL SENSITISATION ASSAY (EpiSensA) TEST METHODS

(Intended for the developers of new or modified similar test methods)

Foreword

This document contains the Performance Standards (PS) for the assessment of proposed in vitro assays, similar to the in vitro Epidermal Sensitisation Assay (EpiSensA), or modified. The EpiSensA is an in vitro method for identifying the skin sensitisation potential of chemicals. It is included as Appendix IC of Test Guideline 442D, which addresses Key Event 2 i.e. keratinocytes activation, on the Adverse Outcome Pathway for skin sensitisation. The PS were developed by Japan in parallel to the development of the EpiSensA.

In 2023, the WNT discussed how to deal with the limited availability outside Japan of the Labcyte 3D skin model used in the EpiSensA. The development of PS was agreed as the solution to overcome this problem, in order to facilitate the development of me-too assays, in particular with other types of reconstructed skin models.

This document was developed with the support of the Expert group on alternative methods for skin sensitisation. The PS were approved by the Working Party of the National Coordinators of the Test Guidelines Programme (WNT) at their 36th meeting in April 2024, following two WNT written commenting rounds in 2023/2024.

This document is published under the responsibility of the Chemicals and Biotechnology Committee.

INTRODUCTION

1. Performance standards (PS) have been developed to facilitate the validation of new or modified Epidermal Sensitisation Assay (EpiSensA) test method for in vitro skin sensitisation and allow for timely amendment of the Test Guideline (1) for their inclusion. A New or modified EpiSensA test method will however only be added to the Test Guideline after review and agreement that all criteria described in the PS are met, including similarity to the EpiSensA test method (the validated reference method, VRM) according to the essential test method components and achievement of the target values for reproducibility and predictive capacity for the proposed reference substances. Mutual Acceptance of Data (MAD) will only be guaranteed for test methods validated according to the PS, if these test methods have been reviewed and included in this Test Guideline by the OECD.

2. The purpose of PS is to provide the basis by which new similar or modified test methods, both proprietary (i.e. copyrighted, trademarked, registered) and non-proprietary can demonstrate to have sufficient reliability and relevance for specific testing purposes. The PS, based on a scientifically valid and accepted test method, can be used to evaluate the reliability and relevance 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 (2). In addition, modified test methods which propose potential improvements to an approved test method, should be evaluated to determine the effect of the proposed changes on the test method’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, as well as the data and documentation available in relation to these changes, modified test methods may: i) either be found unsuitable for a PS-based validation (e.g. if the changes are so substantial that the method is not any longer deemed sufficiently similar with regard to the PS), in which cases they should be subjected to the same validation process as described for a new test method (1), or ii) be suitable for a limited assessment of reliability and relevance using established PS (2).

3. Similar (me-too) or modified test methods proposed for use under the Test Guideline on in vitro skin sensitisation: the EpiSensA test method (1) should be evaluated to determine their reliability and relevance using reference substances representing the full

range of in vivo skin sensitisation effects. The proposed similar or modified test methods should have reliability, accuracy, sensitivity and specificity values which are comparable or better than those derived from the VRM EpiSensA test methods as described in paragraphs 41 to 45. The reliability of the new or modified test method, as well as its ability to correctly identify skin sensitiser test chemicals should be determined prior to its use for testing chemicals.

4. These PS comprise the following three elements:

I) Essential test method components

II) Minimum list of reference substances

III) Defined reliability and accuracy values

ESSENTIAL TEST METHOD COMPONENTS

5. The EpiSensA test method is an *in vitro* assay that quantifies changes in the expression of four marker genes associated with keratinocyte activation (i.e. activating transcription factor 3 (*ATF3*), glutamate-cysteine ligase modifier subunit (*GCLM*), DnaJ (Hsp40) homolog subfamily B (*DNAJB4*), and interleukin-8 (*IL-8*)) in the Reconstructed human Epidermis (RhE) model following an exposure to the test chemical of interest. A 6 hour-exposure time was found optimal for the VRM. Relative changes in marker gene expression are quantified using Reverse Transcription quantitative Polymerase Chain Reaction (RT-qPCR). Cytotoxicity is also assessed concurrently to determine whether upregulated expression of the marker genes occurs at sub-cytotoxic concentrations (cell viability ≥ 80%). The relative induction of marker genes is calculated in comparison to vehicle-treated controls. Test chemicals are considered positive in the EpiSensA test method if the expression of at least one marker gene exceeds the respective cut-off value (*ATF3*, 15-fold; *GCLM*, 2-fold; *DNAJB4*, 2-fold; *IL- 8*, 4-fold) with cell viability remaining ≥ 80%. For this purpose, the mean maximum fold-induction (Imax) value is determined using data from concentrations at which mean cell viability remains ≥ 80%.

6. The Essential Test Method Components consist of essential structural, functional, and procedural elements of the VRM 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 to be considered for similar or modified test methods related to EpiSensA test method are described in detail in the following paragraphs.

7. For specific parameters (e.g. cut-off values), or modified procedures (e.g. RhE models used), adequate values or procedures should be provided for the proposed similar or modified test method. These specific values or procedures may vary depending on the specific test method and/or its modification.

General conditions for RhE model (3)

8. 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 multi layered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic benchmark chemicals, e.g. sodium lauryl sulphate (SLS). The barrier function should be demonstrated and should be assessed by determination of the concentration at which a benchmark chemical reduces the viability of the tissues by 50% (IC50) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET50) upon application of the benchmark chemical at a specified, fixed concentration. The containment properties of the RhE model should prevent the passage of material around the stratum corneum to the viable tissue, which would lead to poor modelling of skin exposure. The acceptability range for the test method is included in Table 1. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, or fungi.

Table 1. QC batch release criteria of RhE model using the EpiSensA test method

RhE model	Lower acceptance limit	Upper acceptance limit
LabCyte EPI-MODEL24 (18 hours treatment with SLS)	IC_{50}=1.4 mg/mL	IC_{50}=4.0 mg/mL

SLS: sodium lauryl sulphate

Functional conditions

Barrier function

9. The stratum corneum and its lipid composition should be sufficient to resist the rapid penetration of cytotoxic benchmark chemicals, e.g. SLS, as estimated by IC50 or ET50.

Morphology

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

Quality control (QC)

11. The RhE model should only be used if the developer/supplier demonstrates that each batch of the RhE model used meets defined production release criteria, among which those for barrier function (paragraph 9) and morphology (paragraph 10) 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 ET50 should be established by the RhE model developer/supplier. Only results produced with qualified tissues can be accepted for reliable prediction.

Procedure conditions

12. The following is a description of the procedures for the EpiSensA test method, which comprises two steps: concentration-finding study and main study (Gene expression analysis). The EpiSensA standard operating procedure is available in the Tracking System for Alternative methods towards Regulatory acceptance (TSAR) (4).

Preparation of RhE model

13. The EpiSensA test method should be conducted utilising an RhE model. The VRM used the LabCyte EPI-MODEL24 kit (#401124), which can be obtained from Japan Tissue Engineering Co., Ltd. (J-TEC). Other RhE models can be used after a validation study based on the present Performance Standards is conducted.

14. RhE models are cultured at 37°C with 5% CO_2 in a humidified atmosphere.

Vehicle selection and assessment of test chemical solubility

15. Assessment of solubility is conducted prior to testing. The solubility of each chemical is evaluated and confirmed visually. For this purpose, test chemicals are dissolved or stably dispersed at a concentration of 50% in acetone: olive oil; 4:1 (AOO, 20% v/v of olive oil in acetone) as a first vehicle option (e.g. 0.1 g of test chemical is measured, and 0.1 mL of AOO is added), distilled water (DW) as a second vehicle option, or 50 % v/v ethanol in DW (50% EtOH) as a third vehicle option. If the test chemical is not soluble or does not stably disperse (i.e. a colloid or suspension in which the test chemical does not settle or separate from the vehicle into different phases within 10 minutes of preparation at room temperature) at a concentration of 50% in any of the vehicles, the highest soluble concentration should be determined by 2-fold serial dilutions beginning with 50% down to 0.0122%. If the test chemical is not soluble or does not form a stable dispersion at 0.0122%, the chemical is not applicable for testing using EpiSensA. The appropriate vehicle is defined as the vehicle that dissolves the test chemical or forms a stable dispersion at the highest concentration tested. It should be verified whether the highest concentration determined can be prepared at weight per volume in a volumetric flask. If the highest soluble or stably dispersed concentration is determined to be 0.0488%, 0.0244%, or 0.0122%, the subsequent concentration-finding study (paragraph 16-22) should be skipped, and main study should be performed (see paragraph 23). In cases in which a vehicle other than AOO, DW, or 50% EtOH is used, appropriate scientific rationale for use of that vehicle should be provided.

Concentration-finding study

16. A concentration-finding assay is performed to determine the concentrations of test chemical to be used for the main study (see *Main study* (*Gene expression analysis)* paragraphs). In the main study, test chemical concentrations that show ≥ 80% mean cell viability should be used, but at least one mean cell viability should be <80% for negative judgement. Therefore, the lowest test chemical concentration that induces a < 80% cell viability is determined in concentration- finding study.

Preparation of test chemicals and control substances for concentration-finding study

17. Test chemicals are prepared on the day of testing and dissolved or stably dispersed in an appropriate vehicle. Starting from the highest concentration determined as specified in paragraph 15, 4-fold serial dilutions are prepared to 0.0122 or 0.0244% (w/v) in the corresponding vehicle. The corresponding vehicles utilised for the preparation of the test chemicals are used as the vehicle controls. Both non-treated control and killed control are used for calculation of cell viability. Non-treated control is used to define 100% cell viability,

and killed control is used to define 0% cell viability (see paragraph 21). Triton X-100 is used as the control substance for killed control in the EpiSensA test method. Triton X-100 should be prepared as a 10% (w/v) solution in DW.

Application of test chemicals and control substances for concentration-finding study

18. For each test chemical, one run is needed to determine the concentration to be used in *the Main study (gene expression analysis)*. One tissue unit for each test chemical concentration and non- treated control and two tissue units for the killed control are used for the cell viability assay. Test chemicals prepared as a working solution (5 µL) and Triton X-100 solution (10 µL) are applied to the centre of each epidermis surface using a positive-displacement pipette and tips. The treated tissue units are then incubated for 6 hours at 37°C with 5% CO_2 in a humidified atmosphere.

Cytotoxicity assessment

19. Cell viability is measured by a lactate dehydrogenase (LDH) assay utilising formazan as the dye. LDH is a stable cytoplasmic enzyme present in all cell types, and it is released into the cell culture medium as a result of damage to the plasma membrane. The LDH assay measures the amount of formazan dye produced by released LDH. The criteria for interference of test chemical in LDH assay (i.e. inhibition of LDH reaction) is described in the TSAR (4).

20. After a 6-hour exposure, 50 µL of the medium for each sample is placed into the wells of a 96-well plate, and an equal volume (i.e. 50 µL) of substrate solution containing lactate and tetrazolium salt is added to each well. The plate is incubated for 30 minutes at room temperature with protection from light, and the reaction is stopped by adding 25 µL/well of 1 mol/L hydrochloric acid (HCl). The absorbance of each well is then measured at 490 or 492 nm along with the reference wavelength (≥ 600 nm) using a 96-well plate absorbance reader. Δabs. is calculated by subtracting the absorbance at reference wavelength from the absorbance at 490 or 492 nm. The absorbance should be measured immediately (no longer than 1 hour) after the addition of HCl.

21. Cell viability can be calculated using the following equation:

$$\textbf{Cell viability}\ (\%) = \mathbf{100} - \frac{\boldsymbol{\Delta}\textbf{abs. of test chemical treatment} - \boldsymbol{\Delta}\textbf{abs. of non-treated control}}{\textbf{mean}\ \boldsymbol{\Delta}\textbf{abs. of killed control} - \boldsymbol{\Delta}\textbf{abs. of non-treated control}} \times \mathbf{100}$$

22. If the LDH assay is not applicable to a test chemical of interest, another cytotoxicity assay (e.g. MTT assay or ATP assay) may be used. In the MTT assay, the activation of the

metabolism in a mitochondria in a cell 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. ATP assay is a homogeneous method, in which tissues are lysed and the number of viable cells is determined based on quantitation of the ATP present in tissues. Therefore, total RNA cannot be isolated from the tissues used for the MTT assay or ATP assay. For this reason, tissues used for the cytotoxicity assessment are required besides tissues used for the gene expression analysis when the MTT assay or ATP assay is used. Other methods (e.g. XTT assay) can be used if sufficient scientific rationale is provided based on the PS.

Main study (Gene expression analysis)

Preparation of test chemicals and control substances for main study

23. An appropriate vehicle (AOO, DW, or 50% EtOH; see paragraph 15) should be used to dissolve or stably disperse the test chemical. The lowest concentration that resulted in < 80% cell viability in the concentration-finding study should serve as the highest concentration (i.e. the starting concentration) in the main study for each chemical and is used in the negative judgement (see paragraph 35). If the cell viability was ≥ 80% at any of the tested concentrations in the *concentration- finding study*, the highest soluble or stably dispersed concentration of the test chemical should be used as the starting concentration. Based on the starting concentration, 2-fold serial dilutions are prepared using the corresponding vehicle to obtain working solutions (at least 3 concentrations are used, including the lowest concentration that resulted in < 80% cell viability in the concentration-finding study or the highest soluble or stably dispersed concentration). If the highest soluble or stably dispersed concentration determined in the *vehicle selection and assessment of test chemical solubility* (paragraph 15) is 0.0488% (w/v), only 3 concentrations (0.0488, 0.0244, and 0.0122% w/v) are used. If the concentration determined in the solubility check is 0.0244 or 0.0122% (w/v), only 2 concentrations (0.0244, and 0.0122% w/v) or only 1 concentration (0.0122% w/v) are used, respectively. Likewise, if the lowest test chemical concentration that induces a < 80% cell viability is 0.0244 or 0.0122% (w/v) in the *concentration-finding study*, only 2 concentrations (0.0244, and 0.0122% w/v) or only 1 concentration (0.0122% w/v) are used, respectively. The vehicle control is prepared as described in paragraph 17. Clotrimazole (CAS no. 23593-75-1, ≥ 98% purity) and 4-nitrobenzyl bromide (4-NBB) (CAS no. 100-11-8, ≥ 98% purity) are used as positive controls in the EpiSensA test method, and 0.78% (w/v) clotrimazole and 0.10% (w/v) 4-NBB solutions are prepared in

AOO (working solution). To calculate cell viability, non-treated and killed controls are prepared as described in paragraph 17.

Application of test chemicals and control substances for main study

24. For each test chemical, one run is required to obtain a prediction. Three tissue units for each test chemical concentration, positive control substance and vehicle controls, two tissue units for the killed control, and one tissue unit for the non-treated control are used for the gene expression analysis. Other than the number of tissue units, application is conducted in the same condition as described in paragraph 18.

Cytotoxicity assessment

25. After a 6-hour exposure, cell viability is determined as described in paragraphs 19-22.

RNA isolation

26. For gene expression analysis, the tissue surface is washed three times with 0.5 mL of phosphate-buffered saline, and the tissue is collected and lysed using one of two lysis methods that were used during test method development and validation (TRIzol reagent and a vortex mixer, or a shredder column and centrifuge).

27. Total RNA, including mRNA, is isolated from lysed RhE tissue samples using a commercially available kit and reagents (e.g. RNeasy Mini kit which was used during test method development and validation).

28. The RNA concentration is quantified, and the RNA quality is analysed from each sample using an RNA analysis equipment, e.g. NanoDrop™ (Thermo Fisher Scientific), with following the protocols provided by the instrument supplier. More than 500 ng of RNA is required for complementary DNA (cDNA) synthesis. RNA concentration and quality should correspond to the recommendations described by the supplier of the reagents which are used in subsequent RT-qPCR (e.g. ≥ 100 ng/µl RNA concentration and A260/A280 in range 1.8-2.0).

RT-qPCR

29. cDNA is synthesised using the commercially available reagents (e.g. Superscript III First-Strand Synthesis System which was used during test method development and validation).

30. After cDNA synthesis, the expression levels of marker genes (i.e. *ATF3*, *GCLM*, *DNAJB4*, and *IL-8*) and the endogenous control gene (i.e. glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*)) are analysed using RT-qPCR. The method described in the *EpiSensA standard operating procedure* (4) should be used (i.e. TaqMan Gene Expression Assay and TaqMan Universal PCR Master Mix). In cases in which another reagent for gene expression is used, appropriate scientific rationale for use of that reagent should be provided.

Data evaluation

31. Relative gene expression is analysed using RT-qPCR. Based on the threshold cycle (Ct) value, the ΔCt and ΔΔCt values as well as fold-induction are calculated according to the following equation:

$$\Delta\textbf{Ct value of marker gene} = \textbf{Ct value of marker gene} - \textbf{Ct value of GAPDH}$$

$$\Delta\Delta\textbf{Ct value of marker gene} = \Delta\textbf{Ct value of marker gene} - \Delta\textbf{Ct value of marker gene (vehicle control)}$$

$$\textbf{fold induction} = 2^{-\Delta\Delta\textbf{Ct value of marker gene}}$$

Cell viability is also calculated according to the equation provided in paragraph 21.

Acceptance criteria

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

- The cell viability of at least two tissue units of the vehicle control should be ≥ 95%. If the cell viability of only one vehicle control is < 95%, the Ct values obtained from the remaining two tissue units should be used.
- The mean cell viability of both positive controls (i.e. 0.78% [w/v] clotrimazole and 0.10% [w/v] 4-nitrobenzyl bromide (4-NBB)) should be ≥ 80%.
- In the 0.78% (w/v) clotrimazole positive control, the mean fold-induction values for *ATF3* and *IL-8* should exceed the cut-off value (i.e. the *ATF3* fold-induction value should be > 15, and the *IL-8* fold-induction value should be > 4).
- In the 0.10% (w/v) 4NBB positive control, the mean fold-induction values for *GCLM* and *DNAJB4* should exceed the cut-off value (i.e. the *GCLM* fold-induction value should be > 2, and the *DNAJB4* fold-induction value should be > 2).

33. The following acceptance criteria should be met in order to consider a tested concentration's result valid:

- The result of at least one tested concentration that shows ≥ 80% mean cell viability should be used. If the mean cell viability is < 80% for a given tested concentration, the result for that tested concentration should be excluded for a positive prediction but might be used for a negative prediction.
- When the mean *GAPDH* Ct value for a given test chemical concentration is within ±1 of the mean *GAPDH* Ct value of the corresponding vehicle control, the result obtained at that concentration is acceptable.

Prediction model

34. Each test chemical is evaluated in one run to derive a prediction (positive or negative). An EpiSensA prediction is considered positive if at least one of the following conditions is met:

- The Imax for *ATF3* is > 15 for at least one tested concentration.
- The Imax for *GCLM* is > 2 for at least one tested concentration.
- The Imax for *DNAJB4* is > 2 for at least one tested concentration.
- The Imax for *IL-8* is > 4 for at least one tested concentration.

35. The EpiSensA prediction is considered negative if:

- The mean fold-induction value of the marker genes does not exceed the respective cut-off values for any of the four genes at any of the tested concentrations and
- At least one mean cell viability at the tested concentrations is < 80%.

36. Cases can occur in which the mean fold-induction value of all four marker genes does not exceed the respective cut-off values at the tested concentration but the mean cell viability at all tested concentrations is ≥ 80%. In such cases, an additional main study should be performed using 2- fold serial dilutions beginning with the concentration greater than the highest concentration used in the first main study. However, if the test chemical does not produce a mean cell viability of < 80% at either the highest soluble or stably dispersed concentration (for solid substances) or 100% (for liquids), the test chemical result is judged as negative.

37. If all mean cell viabilities are < 80% at the tested concentrations greater than or equal to 0.0122% (w/v), the prediction is considered inconclusive.

38. Other cases can occur in which the fold-induction value of a marker gene exceeds the cut-off value only at the lowest concentration showing < 80% mean cell viability. In such a case, the test chemical should be retested using a narrower concentration-response analysis and lower dilution factor (e.g. $\sqrt{2}$ [=1.41]-fold dilution) in order to determine whether induction has occurred at a cytotoxic level (80 to 95% mean cell viability).

MINIMUM LIST OF REFERENCE SUBSTANCES

39. Reference substances are used to determine if the reliability and relevance of a proposed similar or modified test method, proven to be structurally and functionally sufficiently similar to the VRM, or representing a minor modification of the VRM, are comparable or better than those of the VRM. The recommended reference substances listed in Table 2 include substances representing the full range of in vivo skin sensitisation effects, which act via various mechanisms, and are representative of different chemical categories based on their functional groups. The substances included in this list comprise non-sensitisers and skin sensitisers including the various potency categories as established by the Local Lymph Node Assay (LLNA) EC3 value (weak, moderate, strong and extreme). These substances were selected from the substances used in the validation study of the VRM and evaluated during its independent peer-review conducted by JaCVAM.

40. The 20 reference substances listed in Table 2 represent the minimum number of substances that should be used to evaluate the reliability and relevance of a proposed similar or modified test method to discriminate skin sensitisers from non-sensitisers. All 20 reference substances from Table 2 should be used to assess the predictive capacity (PC) and between-laboratory reproducibility (BLR) of the similar or modified test method to discriminate skin sensitisers from non-sensitisers (representing 14 skin sensitisers having various potencies and 6 non-sensitisers). The within- laboratory reproducibility (WLR) on the other hand should be assessed on the basis of a subset of 12 of the 20 reference substances listed in Table 2 (in bold italics, comprising 8 skin sensitisers having various potencies and 4 non-sensitisers). The use of these reference substances for the development/optimisation of new similar test methods should be avoided. In situations where a listed substance is unavailable, other substances for which adequate in vivo reference data (e.g. reliable LLNA data, such as LLNA data for which tested concentrations and corresponding stimulation index (SI) values are available, and sensitisation potency can be determined with high confidence) are available

could be used, primarily from the substances used in the validation study of the VRM. If desired, additional substances representing other chemical classes and for which adequate in vivo reference data (e.g. reliable LLNA data, such as LLNA data for which tested concentrations and corresponding SI values are available, and sensitisation potency can be determined with high confidence) are available may be added to the list of reference substances to further evaluate the relevance of the proposed test method.

Table 2: Minimum List of Reference Substances for Determination of Reproducibility and Predictive Capacity of similar or modified EpiSensA test method

No.	Proficiency substances	CAS No.	Physical state	*in vivo* prediction[1]	Vehicle[2]	EpiSensA results for each marker gene[2, 3]				VRM in vitro prediction
						ATF3	*GCLM*	*DNAJB4*	*IL-8*	
1	***2,4-Dinitrochlorobenzene***	97-00-7	Solid	Sensitiser (GHS Cat. 1A)	AOO	p	p	p	p/n	Positive
2	***p-Phenylenediamine***	106-50-3	Solid	Sensitiser (GHS Cat. 1A)	AOO	p/n	p	p	p/n	Positive
3	Metol	55-55-0	Solid	Sensitiser (GHS Cat. 1A)	DW	p	p	p	p/n	Positive
4	Tetrachloro-salicylanilide	1154-59-2	Solid	Sensitiser (GHS Cat. 1A)	AOO	p	n	p	p	Positive
5	***Lauryl gallate***	1166-52-5	Solid	Sensitiser (GHS Cat. 1A)	AOO	p/n	n	p/n	p/n	Negative
6	***Methyl heptine carbonate***	111-12-6	Liquid	Sensitiser (GHS Cat. 1A)	AOO	p	p	p	P	Positive
7	Isoeugenol	97-54-1	Liquid	Sensitiser (GHS Cat. 1A)	AOO	p/n	p	p	p/n	Positive
8	***Glyoxal 40% solution in water***	107-22-2	Liquid	Sensitiser (GHS Cat. 1A)	DW	p/n	p	p/n	p/n	Positive
9	***Abietic acid***	514-10-3	Solid	Sensitiser (GHS Cat. 1B)	AOO	p	p	p	p	Positive
10	Dibutyl aniline	613-29-6	Liquid	Sensitiser (GHS Cat. 1B)	AOO	p/n	n	n	p	Positive
11	Amyl cinnamic aldehyde	122-40-7	Liquid	Sensitiser (GHS Cat. 1B)	AOO	p	n	p	p	Positive
12	***Benzisothiazolinone***	2634-33-5	Solid	Sensitiser (GHS Cat. 1B)	AOO	p/n	p/n	p/n	p/n	Positive[5]
13	Imidazolidinyl urea	39236-46-9	Solid	Sensitiser (GHS Cat. 1B)	DW	p/n	n	p/n	p	Positive
14	***Farnesol***	4602-84-0	Liquid	Sensitiser (GHS Cat. 1B)	AOO	p	p/n	p	p	Positive
15	***Cetrimide***	57-09-0	Solid	Non-sensitiser (Not classified)	50%EtOH	n	n	n	n	Negative
16	***Lactic acid*** [4]	50-21-5	Liquid	Non-sensitiser (Not classified)	DW	n	n	n	n	Negative
17	Benzyl butyl phthalate	85-68-7	Liquid	Non-sensitiser (Not classified)	AOO	n	n	n	n	Negative
18	***Diethyl phthalate***	84-66-2	Liquid	Non-sensitiser (Not classified)	AOO	p/n	n	p/n	p	Positive
19	***Hexane***	110-54-3	Liquid	Non-sensitiser (Not classified)	AOO	n	n	n	n	Negative
20	1-Iodehexane	638-45-9	Liquid	Non-sensitiser (Not classified)	AOO	n	p	p	p	Positive

[1]: The *in vivo* hazard and potency prediction is based on LLNA data (TG497, SD Annex3) (Urbisch, 2015). The *in vivo* potency is derived using the criteria based on UN GHS Sub-categorisation.
[2]: Based on historical results (Mizumachi et al, 2018) (EpiSensA validation report).
[3]: "p" indicates that the fold-induction of marker gene exceeds the cut-off value with ≥ 80% viability. "n" indicates that the fold-induction of marker gene doesn't exceed the cut-off value with ≥ 80% viability. "p/n" means both "p" and "n" are acceptable because both "p" and "n" results were obtained at the validation study.
[4]: MTT assay should be performed instead of LDH assay.
[5]: Reference substance which was not 100% concordant between laboratories.
Bold italics: A subset of 12 out of the 20 reference substances that should be used for an assessment of within-laboratory reproducibility (WLR).

DEFINED RELIABILITY AND ACCURACY VALUES

41. For purposes of establishing the reliability and relevance of proposed similar or modified test methods falling within the Test Guideline for in vitro skin sensitisation: EpiSensA test method (1), all of the reference substances listed in Table 2 should be tested. It is however essential that all PS based validation studies are independently assessed by internationally recognised validation bodies, in agreement with international guidelines (2). The 20 reference substances should be tested in each of at least three laboratories. For the purpose of evaluating the within-laboratory reproducibility the subset of 12 reference substances identified in bold italics in Table 2 should be tested by each participating laboratory to produce three qualified runs to derive three predictions in each laboratory. The remaining 8 reference substances should be tested by each laboratory in one single qualified run to derive one prediction per laboratory for a total of three predictions for the three laboratories. Finally, all 20 reference substances should be used to assess predictive capacity.

42. The calculation of the within-laboratory reproducibility, between-laboratory reproducibility, accuracy, sensitivity and specificity values of the proposed test method should be done according to the rules described below to ensure that a predefined and consistent approach is used:

1. WLR should be calculated based on concordance of classifications obtained by each participating laboratory for the subset of 12 reference substances identified in bold italics in Table 2, using three qualified runs.

2. BLR should be calculated based on concordance of classifications obtained by at least three participating laboratories for the 20 reference substances listed in Table 2. BLR should be calculated based on concordance of classifications using only qualified runs. For the 12 substances for which each laboratory should generate three classifications (for WLR assessment), one single final classification should be derived per laboratory based on the mode of the three predictions obtained. These single final classifications should then be used for BLR assessment.

3. The calculation of the accuracy values should be done using all qualified runs generated by at least three laboratories with the 20 reference substances. The calculations should be based on the predictions obtained with each qualified run. With respect to the three replicate runs performed for WLR assessment, the prediction of skin sensitisation potential should be derived per laboratory based on the mode of the three predictions obtained. In this context, a qualified run means a

run that meets the acceptance criteria for the negative (vehicle) and positive control, as defined in the SOP of a new or modified EpiSensA test method and paragraphs 37 to 38 of the Test Guidelines on an in vitro skin sensitisation: EpiSensA test method (1). Otherwise, the run is considered as non-qualified.

Within-laboratory reproducibility

43. An assessment of within-laboratory reproducibility should show a concordance of predictions (positive versus negative) obtained in three different, independent qualified runs of the 12 recommended reference substances (shown in bold italics in Table 2) within each participating laboratory equal or higher (≥) than 80.0% (actual for EpiSensA average 88.8 % based on the validation dataset).

Between-laboratory reproducibility

44. For similar or modified test methods, the concordance of predictions (positive versus negative) between a minimum of three laboratories, obtained for the 20 recommended reference substances (shown in Table 2), should be equal or higher (≥) than 80.0% (actual for EpiSensA test method: 94.1 % based on 20 reference substances indicated in Table 2 with the exclusion of metol and dibutyl aniline for which no EpiSensA data on BLR is included in the validation dataset (5)).

Predictive capacity

45. The accuracy, sensitivity and specificity of the proposed similar or modified test method should be comparable or better to that of the VRM. The accuracy and sensitivity obtained with the 20 reference substances listed in Table 2 should be equal or higher (≥) than 85.0%, and the specificity should be equal or higher (≥) than 65.0% (actual for EpiSensA based on the 20 reference substances: 85.0 % accuracy, 92.9 % sensitivity and 66.7% specificity). The predictive capacity of EpiSensA calculated on the basis of the full validation dataset is reported in paragraph 4 of the Test Guideline (1). Furthermore, no strong sensitisers (i.e. UN GHS Cat. 1A) should be under-predicted as non- sensitiser, unless a clear rationale can be given.

References

1) OECD (2024), Test guideline for EpiSensA test method (in preparation)

2) OECD (2005), OECD Series on Testing and Assessment No. 34. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment.

3) OECD, Guidance document, No. 220, Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Reconstructed Human Epidermis (RhE) Test Methods for Skin Irritation Testing as described in TG 439

4) EURL ECVAM. (2023). EpiSensA standard operating procedure. Available at: https://tsar.jrc.ec.europa.eu/test-method/tm2018-01-0.

5) OECD (2023). Series on Testing & Assessment No. 384: Epidermal Sensitisation Assay (EpiSensA) Validation Study Report; Organisation for Economic Cooperation and Development, Paris. Available at: https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications- number.htm

Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Epidermal Sensitisation Assay (EpiSensA) Test Methods

Series on Testing and Assessment No. 396

This document contains the Performance Standards (PS) for the assessment of proposed in vitro assays, similar to the in vitro Epidermal Sensitisation Assay (EpiSensA), or modified. The EpiSensA is an in vitro method for identifying the skin sensitisation potential of chemicals. The EpiSensA method has been included in Test Guideline 442D, upon a proposal from Japan, reviewed, consolidated and finally approved by the Working Party of the National Coordinators of the Test Guidelines Programme. This was a project on the work plan of the Test Guidelines Programme until 2024.