Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

22-May-2015

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

PERFORMANCE STANDARDSFOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED IN VITRO SKIN SENSITISATION ARE-NRF2 LUCIFERASE TEST METHODS

Series on Testing and Assessment

No. 213

JT03376931

ENV/JM/MONO(2015)6
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 213

PERFORMANCE STANDARDSFOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED IN VITRO SKIN SENSITISATION ARE-NRF2 LUCIFERASE TEST METHODS

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
Paris 2015

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, 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 countries with special status at the OECD, 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 eleven 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; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (http://www.oecd.org/chemicalsafety/).

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

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 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.

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment link

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This document includes Performance Standards (PS) for the assessment of proposed similar or modified *in vitro* skin sensitisation ARE-Nrf2 luciferase test methods. The PS were developed by the European Union Reference Laboratory (EURL)-ECVAM in parallel to the development of the Test Guideline 442D (adopted 5th February 2015), which is currently based on the Keratinosens™ test method. In 2012 the WNT discussed how to deal with the issue of monopoly when a single test method including proprietary elements is available for a hazard/endpoint. The development of PS was agreed as the solution to overcome this problem.

The Test Guideline 442D makes reference to the PS document and explain that PS are available to facilitate the development and validation of similar test methods, and that the Mutual Acceptance of Data will only be guaranteed for test methods, developed according to the Performance Standards, if they have been reviewed and adopted by OECD. This means that PS are not "acceptability criteria" for data generated using a Test Guideline, and that the Test Guidelines and the Mutual Acceptance of Data are intended for the users of Test Guidelines, while the Performance Standards are intended for developers of "me-too" tests.

The PS were approved by the Working Group of the National Coordinators of the Test Guidelines Programme by written procedure on 30 June 2014 and declassified by the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biotechnology on 21 April 2015. This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biotechnology.

PERFORMANCE STANDARDSFOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED *IN VITRO* SKIN SENSITISATION ARE-NRF2 LUCIFERASE TEST METHODS

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

INTRODUCTION

1. Performance standards (PS) have been developed to facilitate the validation of new or modified *in vitro* ARE-Nrf2 luciferase test methods similar to the KeratinoSens™ and allow for timely amendment of the Test Guideline (1) for their inclusion. New or modified *in vitro* ARE-Nrf2 luciferase test methods 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 KeratinoSens™ (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 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, 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 relevance using established PS (2).

3. Similar (me-too) or modified test methods proposed for use under the Test Guideline on an *in vitro* skin sensitisation: ARE-Nrf2 luciferase test method (1) should be evaluated to determine their reliability and relevance using reference substances (Table 1) 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 validated reference method (VRM) KeratinoSens™ as described in paragraphs 8 to 12. 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. 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 VRM (2). The essential test method components in the present PS comprise the elements described below:

- Transgenic human cell lines relevant to the skin sensitisation process (e.g. keratinocytes, dendritic cells and other relevant cells) and that have a stable insertion of luciferase reporter gene should be used.

- The reporter gene must be under the control of the ARE-element of the human AKR1C2 gene or alternative ARE elements.

- If cell lines containing alternative ARE elements are used, the specific dependence of the chosen ARE element on Nrf2 and the Nrf2 dependence of the reporter gene activity should be demonstrated (e.g. by transiently or permanently inactivating Nrf2 activity in the reporter cell line and by verifying that sensitiser-induced luciferase activity is reduced due to reduced or lost expression of Nrf2).

- It should be demonstrated that the cell line has a stable insertion of the luciferase reporter gene.

- Clones should be selected on the basis of performance e.g. based on the best signal to noise ratio of the light output of luciferase induction, and on the highest dynamic range of luciferase induction when cells are treated with weak sensitisers.

- The optimal cell seeding number as well as media composition (e.g. DMSO, FCS) should be defined to ensure obtaining significant luciferase induction by skin sensitisers, including weak sensitisers.

- The measurement of luciferase reporter gene activity and the appropriate luciferase substrate used should have sufficient light output to ensure sufficient sensitivity and low variability.

- Finally, cell cytotoxicity should be assessed.

MINIMUM LIST OF REFERENCE SUBSTANCES

6. 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 (5). The recommended reference substances listed in Table 1 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 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 EURL ECVAM (3) (4) (5) (6).

7. The 20 reference substances listed in Table 1 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 1 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 12 skin sensitisers having various potencies and 8 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 1 (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 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 are available may be added to the list of reference substances to further evaluate the relevance of the proposed test method.

Table 1: List of reference substances for determination of reproducibility (n=12 substances, in bold italics for WLR; n=20 substances for BLR) and predictive capacity (n=20 substances) of similar or modified *in vitro* skin sensitisation ARE-Nrf2 luciferase test methods

Reference substances [a)]	**CASRN**	**Physical state**	***In vivo* categories** [b)]	**VRM *In vitro* prediction**
NON-SENSITISERS based on *in vivo* results				
4-methoxy-acetophenone [c)]	***100-06-1***	***Solid***	***Non-sensitiser***	***Positive***
Glycerol	***56-81-5***	***Liquid***	***Non-sensitiser***	***Negative***
Isopropanol	***67-63-0***	***Liquid***	***Non-sensitiser***	***Negative***
Salicylic acid	***69-72-7***	***Solid***	***Non-sensitiser***	***Negative***
Chlorobenzene	108-90-7	Liquid	Non-sensitiser	Negative
Lactic acid	50-21-5	Liquid	Non-sensitiser	Negative
Methyl salicylate	119-36-8	Liquid	Non-sensitiser	Negative

Reference substances [a)]	**CASRN**	**Physical state**	***In vivo* categories** [b)]	**VRM *In vitro* prediction**
Sulfanilamide	63-74-1	Solid	Non-sensitiser	Negative
SKIN SENSITISERS based on *in vivo* results				
Ethylene glycol dimethacrylate	***97-90-5***	***Liquid***	***Skin sensitiser (weak)***	***Positive***
Phenyl benzoate	***93-99-2***	***Solid***	***Skin sensitiser (weak)***	***Negative***
Citral	***5392-40-5***	***Liquid***	***Skin sensitiser (moderate)***	***Positive***
Isoeugenol	***97-54-1***	***Liquid***	***Skin sensitiser (moderate)***	***Positive***
Methyldibromo glutaronitrile	***35691-65-7***	***Solid***	***Skin sensitiser (strong)***	***Positive***
para-phenylenediamine	***106-50-3***	***Solid***	***Skin sensitiser (strong/extreme)***	***Positive***
2,4-Dinitrochlorobenzene	***97-00-7***	***Solid***	***Skin sensitiser (extreme)***	***Positive***
4-Nitrobenzylbromide	***100-11-8***	***Solid***	***Skin sensitiser (extreme)***	***Positive***
Eugenol	97-53-0	Liquid	Skin sensitiser (weak)	Negative [d)]
2-Mercaptobenzothiazole	149-30-4	Solid	Skin sensitiser (moderate)	Positive
4-Methylaminophenol sulphate	55-55-0	Solid	Skin sensitiser (strong)	Positive
Oxazolone	15646-46-5	Solid	Skin sensitiser (extreme)	Positive

[a)] The substances, sorted first by non-sensitisers followed by skin sensitisers and then ranked on the basis of their testing purpose and skin sensitisation potency were selected from the test chemicals used in validation and in-house studies of KeratinoSens™ which were peer-reviewed by EURL ECVAM (3) (4) (5) (6). The selection includes, to the extent possible, substances that: (i) are representative of the range of skin sensitisation potency (e.g. non-sensitisers, weak, moderate, strong and extreme skin sensitisers) tested in the VRM; (ii) reflect to the extent possible the performance characteristics of the VRM for BLR and PC; (iii) have chemical structures that are well-defined; (iv) include a variety of mechanisms of action (including pro-haptens, oxidising chemicals, adduct forming chemicals, Michael acceptors, Schiff base formation, acyl transfer chemicals, aryl electrophile, electrophile-H-polar chemical) (7) (8) (9); (v) include a variety of chemical categories based on their organic functional groups; (vi) induce to the extent possible definitive results in the *in vivo* reference test method; (vii) are commercially available; and (viii) are not associated with prohibitive disposal costs.

[b)] The *in vivo* categories are based on EC3 values from the LLNA test methods (10 < weak; 1< moderate <10; 0.1< strong < 1; extreme < 0.1).

[c)] Reference substance not having KeratinoSens™ data on between-laboratory variability.

[d)] Reference substance which was not 100% concordant between laboratories.

Abbreviations: BLR = Between laboratory reproducibility; CASRN = Chemical Abstracts Service Registry Number; PC = Predictive capacity; WLR = Within-laboratory reproducibility.

DEFINED RELIABILITY AND ACCURACY VALUES

8. 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: ARE-Nrf2 luciferase test method (1), all of the reference substances listed in Table 1 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 1 should be tested by each participating laboratory to produce three qualified experiments to derive three predictions in each laboratory. The remaining 8 reference substances should be tested by each laboratory in one single qualified experiment 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. One qualified experiment consists of at least two qualified independent repetitions in case concordant results are obtained, or of three qualified independent repetitions in case of discordant results in the first two repetitions. Each repetition consists of three replicates for each concentration tested. Negative (solvent) control and positive control should be tested in three replicates concurrently with the test chemical in each repetition.

9. 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. Within-laboratory reproducibility (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 1, using three qualified experiments.

2. Between-laboratory reproducibility (BLR) should be calculated based on concordance of classifications obtained by at least three participating laboratories for the 20 reference substances listed in Table 1. BLR should be calculated based on concordance of classifications using only qualified experiments. For the 12 substances for which each laboratory will 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 experiments generated by at least three laboratories with the 20 reference substances. The calculations should be based on the predictions obtained with each qualified experiment. A weighted calculation should be used to take into account the fact that the 20 reference substances will have a different number of experiments, i.e. the 12 substances used for both BLR and WLR assessment will have nine experiments each, whereas the 8 substances used for BLR assessment only will have three experiments each. In summary, the final outcome of each individual qualified experiment obtained for each substance (from all participating laboratories) should be captured as an independent classification in the calculations and correction factors should be applied so that all substances have an equal weight in the calculations. The positive and negative predictions for each substance should thus be divided by the total number of predictions for that substance so that each substance contributes with a final weight of 1 in the calculations.

In this context, a qualified experiment consists of an experiment containing the necessary number of repetitions meeting the the acceptance criteria for the negative and positive control, as defined in the SOP

and paragraphs 36 to 38 of the Test Guidelines on an *in vitro* skin sensitisation: ARE-Nrf2 luciferase test method (1). Otherwise, the experiment is considered as non-qualified.

Within-laboratory reproducibility

10. An assessment of within-laboratory reproducibility should show a concordance of predictions (positive versus negative) obtained in three different, independent qualified experiments of the 12 recommended reference substances (shown in bold italics in Table 1) within each participating laboratory equal or higher (≥) than 80.0% (actual for KeratinoSens™ 85.7% based on the validation dataset).

Between-laboratory reproducibility

11. 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 1), should be equal or higher (≥) than 80.0% (actual for KeratinoSens™: 94.7% based on 19 reference substances indicated in Table 1 with the exclusion of 4-methoxy-acetophenone for which no KeratinoSens™ data on BLR is available, and 85.7% based on the validation dataset).

Predictive capacity

12. 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, sensitivity and specificity obtained with the 20 reference substances listed in Table 1 should all be equal or higher (≥) than 80.0% (actual for KeratinoSens™ based on the 20 reference substances and using a weighted calculation: 87.0% accuracy, 86.7% sensitivity and 87.5% specificity. The predictive capacity of KeratinoSens™ calculated on the basis of the full validation dataset is reported in paragraph 10 of the Test Guideline (1). Furthermore no strong or extreme sensitisers should be under-predicted as non-sensitiser, unless a clear rationale can be given.

LITERATURE

1. OECD (2014), Test Guideline on an In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method, OECD Guidelines for the Testing of Chemicals N°442D, Section 4, OECD, Paris.

2. OECD (2005). Guidance Document No.34 on "*the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment"*. OECD Series on Testing and Assessment. Organisation for Economic Cooperation and Development, Paris, France, 2005, 96 pp.

3. EURL-ECVAM (2014). Recommendation on the KeratinoSens™ assay for skin sensitisation testing, 42 pp. Available at: http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations/recommendation-keratinosens-skin-sensitisation.

4. Emter R., Ellis G., Natsch A.(2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers *in vitro*. Toxicology and Applied Pharmacology 245, 281-290.

5. Natsch A., Bauch C., Foertsch L., Gerberick F., Normann K., Hilberer A., Inglis H., Landsiedel R., Onken S., Reuter H., Schepky A., Emter R. (2011). The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers *in vitro*: results of a ring-study in five laboratories. Toxicol. *In Vitro* 25, 733-744.

6. Natsch A., Ryan C.A., Foertsch L., Emter R., Jaworska J., Gerberick G.F., Kern P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. Journal of Applied Toxicology, 33, 1337-1352.

7. Aptula A.O., Patlewicz G., Roberts D.W., Schultz T.W. (2006). Non-enzymatic glutathione reactivity and *in vitro* toxicity: a non-animal approach to skin sensitization. Toxicology *In Vitro* 20, 239-247.

8. Casati S., Aeby P., Kimber I., Maxwell G., Ovigne J.M., Roggen E., Rovida C., Tosti L., Basketter D. (2009). Selection of chemicals for the development and evaluation of in vitro methods for skin sensitisation testing. Altern Lab Anim. 37, 305-312.

9. Rovida C., Martin S.F., Vivier M., Weltzien H.U., Roggen E. (2013). Advanced tests for skin and respiratory sensitization assessment. ALTEX 30, 231-352.