PERFORMANCE STANDARDS FOR THE ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED IN VITRO SKIN SENSITISATION DPRA AND ADRA 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 proposed similar or modified test methods based on the Direct Peptide Reactivity Assay (DPRA) and the Amino acid Derivative Assay (ADRA) and to allow for their timely inclusion in the Test Guidelines. (1) (2) Proposed similar or modified test methods based on in chemico covalent binding to proteins will only be added to the Test Guideline, however, after a review process to confirm that all criteria stipulated in the PS for similarity to the validated reference methods (VRM)—namely, DPRA and ADRA—have been met, that the proposed similar or modified test method includes all essential test method components, and that test performance achieves the target values for reproducibility and predictive capacity of the proposed reference chemicals. 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 these Performance Standards (PS) is to provide a basis by which proposed similar or modified test methods, both proprietary (i.e., copyrighted, trademarked, registered) and non-proprietary, can demonstrate sufficient reliability and relevance for 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. (3) In addition, modified test methods which propose potential improvements to an approved test method should be evaluated to determine the effect of the proposed modifications on the test method’s performance and the extent to which such modifications affect the information available for the other components of the validation process. Depending on the number and nature of the proposed modifications as well as the data and documentation available to supports the modifications, proposed similar or modified test methods should either be subjected to the same validation process as any new test method or, where appropriate, to a limited assessment of reliability and relevance using established PS. (3)

3. Similar (me-too) or modified test methods proposed for use under Test Guidelines for a test method based on in chemico covalent binding to proteins (1) (2) should be evaluated to determine their reliability and relevance using a set of reference chemicals (Table 1) that represent the full range of in vivo skin sensitisation effects. The proposed similar or modified test methods should demonstrate reliability, accuracy, sensitivity, and specificity values that are at least as good as those derived from the VRM—DPRA and ADRA—and as described below in paragraphs 8 to 12. The reliability of the proposed similar or modified test method as well as its ability to correctly predict the skin sensitization potential of test chemicals should be validated prior to its use in testing chemicals.

4. These PS comprise the following three elements:

I) Essential test method components

II) Minimum list of reference chemicals

III) Defined reliability and accuracy values

ESSENTIAL TEST METHOD COMPONENTS

5. The Essential Test Method Components comprise the essential structural, functional, and procedural elements of a 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 VRM. (3) The essential test method components are described in detail in the following paragraphs.

● Proteins, peptides, amino acids, and their derivatives that are relevant to covalent binding to proteins in the skin sensitization process should be used as nucleophilic reagents in the assay based on covalent binding to proteins.

● This test method is based on the principle that, since skin sensitizers undergo in vivo covalent binding to proteins, skin sensitization potential can be predicted by assessing whether or not a test chemical undergoes in chemico covalent binding with a nucleophilic reagent containing thiol groups like cysteine or amino groups like lysine.

● Nucleophilic reagents containing thiol groups are susceptible to the formation of oxidative dimers, which can significantly compromise the quality of test results.

MINIMUM LIST OF REFERENCE CHEMICALS

6. Reference chemicals are used to determine if the reproducibility and predictive capacity of a proposed similar or modified test method that has been shown to be sufficiently similar, both structurally and functionally, to the VRM or represents only a minor modification of the VRM are at least as good as that of the VRM. (4) (5) The recommended reference chemicals listed in Table 1 represent the full range of in vivo skin sensitization effects that act via a variety of mechanisms and are representative of different chemical categories based on their functional groups. The chemicals included in this list include skin sensitizers of various potencies based on LLNA EC3 values—e.g., weak, moderate, strong, and extreme—as well as non-sensitizers. These chemicals were selected from those used in the validation studies of the VRM and evaluated during the independent peer reviews conducted by EURL ECVAM and JaCVAM. (4) (5)

7. The 20 reference chemicals listed in Table 1 represent the minimum number of chemicals that should be used to evaluate the reproducibility and predictive capacity of a proposed similar or modified test method to distinguish

skin sensitizers from non-sensitizers. These 20 reference chemicals were selected from the 40 test chemicals used in the ADRA validation study, 13 of which were also used in the DPRA validation study. The figures for reproducibility and predictive capacity given in paragraphs 10, 11, and 12, however, are based only on ADRA results. All 20 reference chemicals listed in Table 1 should be used to assess the predictive capacity and between-laboratory reproducibility (BLR) of the proposed similar or modified test method to distinguish skin sensitizers from non-sensitizers, including 13 sensitizers of various potencies and 7 non-sensitizers. In contrast to this, the within-laboratory reproducibility (WLR) should be assessed on the basis of a subset of 12 of the 20 reference chemicals, which are listed in Table 1 in bold italics and include 8 sensitizers of various potencies and 4 non-sensitizers. The use of these reference chemicals for the development and optimization of proposed similar test methods should be avoided. In situations where a listed chemical is unavailable, it should be substituted with another chemical for which adequate in vivo reference data is available, preferably from the chemicals used in the validation of the VRM. To further evaluate the accuracy of the proposed test method, additional chemicals representing other chemical classes and for which adequate in vivo reference data are available may be added to the list of reference chemicals.

Although benzyl salicylate (No. 6) is known to be a moderate sensitizer and benzyl cinnanmate (No. 16) a weak one, these two chemicals were both predicted to be non-sensitizers in both DPRA and ADRA. Yet their chemical structures are such that it is hard to conceive of either reacting strongly with thiol or amino groups.

One possible explanation of their sensitization potential is that, since both these chemicals have an ester structure in common, in vivo hydrolysis of these esters gives chemicals that become sensitizers after undergoing oxidative metabolism. Thus, although correctly predicted to be sensitizers in LLNA testing, both these chemicals gave false negative results when tested using DPRA and ADRA.

Table 1: List of reference chemicals for determination of reproducibility (12 chemicals for WLR, 20 chemicals for BLR) and predictive capacity (20 chemicals) in a proposed similar or modified protein reactivity assay

No.	Test chemicals	CAS No.	Physical state	Molecular weight	Mechanism	*in vivo* prediction[1]	DPRA prediction	ADRA prediction[2]
12 Test chemicals for Within-Laboratory Reprodusibility and Between-Laboratory Reproducibility								
1	4-(Methylamino) phenol hemisulfate salt	55-55-0	Solid	221.23	pre-hapten, Michael acceptor	Sensitizer (strong)	Pos[3]	Pos
2	Lauryl gallate	1166-52-5	Solid	338.44	pre-hapten, Michael acceptor	Sensitizer (strong)	Pos[3]	Pos
3	Chloramine T	7080-50-4	Solid	281.69	Acylation	Sensitizer (strong)	Pos[4]	Pos
4	Cinnamaldehyde	14371-10-9	Liquid	132.16	Michael acceptor	Sensitizer (moderate)*	Pos[3,*] (Positive control)	Pos
5	2-Mercaptobenzothiazole	149-30-4	Solid	167.25	S_N2, acylation	Sensitizer (moderate)	Pos[4]	Pos
6	Benzyl salicylate	118-58-1	Liquid	228.25	S_N2, acylation	Sensitizer (moderate)	Pos/Neg[4]	Neg
7	Ethyl acrylate	140-88-5	Liquid	100.12	Michael acceptor	Sensitizer (weak)	Pos[3]	Pos
8	Imidazolidinyl urea	39236-46-9	Solid	388.29	Acylation	Sensitizer (weak)	Pos[4]	Pos
9	Salicylic acid	69-72-7	Solid	138.12	Non-reactive	Non-sensitizer	Pos/Neg[3]	Neg
10	Benzyl alcohol	100-51-6	Liquid	108.14	Non-reactive	Non-sensitizer	Pos/Neg[4]	Neg
11	Glycerol	56-81-5	Liquid	92.09	Non-reactive	Non-sensitizer	Neg[4]	Neg
12	Isopropanol	67-63-0	Liquid	60.1	Non-reactive	Non-sensitizer	Neg[4]	Neg
8 Test chemicals for Between-Laboratory Reproducibility								
13	*p* -Benzoquinone	106-51-4	Solid	108.09	Michael acceptor	Sensitizer (extreme)	Pos[4]	Pos
14	Dihydroeugenol	2785-87-7	Liquid	166.22	pro-hapten, S_N2, Michael acceptor	Sensitizer (moderate)	Pos/Neg[4]	Pos/Neg
15	Palmitoyl Chloride	112-67-4	Liquid	274.87	Acylation	Sensitizer (moderate)	Pos[3]	Pos
16	Benzyl cinnamate	103-41-3	Solid	238.29	Michael acceptor, S_N2	Sensitizer (weak)	Neg[4]	Neg
17	Farnesol	4602-84-0	Liquid	222.37	Schiff base (oxidation form)	Sensitizer (weak)	Neg[3]	Pos
18	Dimethyl isophthalate	1459-93-4	Solid	194.19	Non-reactive	Non-sensitizer	Neg[4]	Neg
19	Methyl salicylate	119-36-8	Liquid	152.15	Non-reactive	Non-sensitizer	Pos/Neg[4]	Neg
20	4-Aminobenzoic acid	150-13-0	Solid	137.14	Non-reactive	Non-sensitizer	Neg[4]	Neg

Chemicals highlighted in pink were predicted to be sensitizers, those highlighted in blue were predicted to be non-sensitizers, and those highlighted in yellow had non-concordant results.

[1]Predictions of in vivo hazard (potency) are based on LLNA data. (4) (9) (11) (12) In vivo potency is derived using criteria proposed by ECTOC. (10)

[2]Result of ADRA validation study. (5)

[3]Predictions based on published data. (9) (11) (12) (13)
[4]Result of DPRA validation study. (4)
[*]Predictions based on test results for cinnamic aldehyde (CAS No. 104-55-2).

Chemicals were selected from the test chemicals used in validation of ADRA that underwent a peer review organized by JaCVAM. (5) They were first sorted into non-sensitizers and skin sensitizers, then ranked on the basis of their testing purpose and skin sensitisation potency. The selection includes chemicals that
(i) are representative of the range of skin sensitization potency tested with the VRM (e.g., weak, moderate, strong, and extreme sensitizers as well as non-sensitizers),
(ii) reflect the performance characteristics of the VRM for BLR and predictive capacity,
(iii) have chemical structures that are well-defined,
(iv) include a variety of mechanisms of action, (6) (7) (8)
(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 prohibitively expensive to dispose of.

The in vivo categories are based on EC3 values from the LLNA test methods (weak: EC3 > 10%, moderate: EC3 ≥ 1%, strong: EC3 ≥ 0.1%, and extreme: EC3 < 0.1%).

DEFINED RELIABILITY AND ACCURACY VALUES

8. In order to assess the reliability and relevance of proposed similar or modified test methods based on in chemico covalent binding to proteins, (1) (2) all reference chemicals listed in Table 1 should be tested. Validation studies based on performance standards should be assessed independently by internationally recognized validation bodies in agreement with international guidelines. (3) The 20 reference chemicals should each be tested by at least three laboratories. Within-laboratory reproducibility should be evaluated using the subset of 12 reference chemicals listed in Table 1 to conduct three qualified tests resulting in three predictions at each laboratory. The remaining 8 reference chemicals should be used to conduct a single qualified test resulting in one prediction per laboratory at each laboratory. Finally, results from all 20 reference chemicals should be used to assess predictive capacity. Each qualified test must comprise at least two qualified independent repetitions. If the first two repetitions are concordant, a third repetition is unnecessary. If the first two repetitions are non-concordant, a third repetition is needed to determine the outcome. Each repetition comprises three replicates of the test chemical solution, tested concurrently with three replicates of the negative and positive control reagents.

9. The calculation of values for within-laboratory reproducibility, between-laboratory reproducibility, accuracy, sensitivity, and specificity should be done according to the rules described below to ensure the use of a predefined and consistent approach.

1. WLR should be calculated based on concordance of predictions made using only qualified test results obtained from the subset of 12 reference chemicals listed in Table 1 for which at least three qualified tests are available.

2. BLR should be calculated based on concordance of predictions made using only qualified test results obtained from the 20 reference chemicals listed in Table 1 for which at least one qualified test per laboratory is

available. For the subset of 12 chemicals that were tested three times each for assessing WLR, a single prediction should be derived based on the mode of the three predictions and used to assess BLR.

3. Values for accuracy should be calculated using all qualified test results obtained from the 20 reference chemicals at each laboratory. The calculations should be based on the individual predictions made for each qualified test result of each reference chemical in each laboratory. Accuracy is given as a percentage, calculated by dividing the sum of all sensitizers that were correctly predicted to be sensitizers and all non-sensitizers that were correctly predicted to be non-sensitizers by the total number (20) of chemicals tested.

   The calculations should take into account the fact that the 12 chemicals used to assess both BLR and WLR were each tested nine times, whereas the 8 chemicals used to assess only BLR were tested three times each.

   Test results are considered qualified test results if they were obtained from tests that comprise the number of repetitions that satisfy the acceptance criteria for the negative and positive control, as defined in the SOP and the Test Guidelines for test methods based on in chemico covalent binding to proteins. (1) (2) Otherwise, the test results are not qualified.

Within-laboratory reproducibility

10. Assessments of the WLR of the proposed similar or modified test method should demonstrate that at least 80.0% of the predictions obtained from three independent qualified test results for each chemical in the recommended subset of 12 reference chemicals listed in Table 1 in bold italics are concordant. (97.3% for ADRA per the validation dataset.) (5)

Between-laboratory reproducibility

11. Assessments of the BLR of the proposed similar or modified test method should demonstrate that at least 80.0% of the predictions obtained for the 20 reference chemicals shown in Table 1 at a minimum of three laboratories are concordant. (95.0% for ADRA per the validation dataset.) (5)

Predictive capacity

12. Assessments of the predictive capacity of the proposed similar or modified test method should be comparable to that of the VRM, and calculations should demonstrate an accuracy, sensitivity and specificity of at least 80.0% for the 20 reference chemicals listed in Table 1. (Accuracy of 87.5%, sensitivity of 80.8%, and specificity of 100% for ADRA per the validation dataset.) Predictive capacities for both DPRA and ADRA were calculated on the basis of the full validation dataset and are reported in the DPRA and ADRA validation study reports. (4) (5) Also, a clear rationale should be given for any under-predictions (false negatives) of strong or extreme sensitizers.

LITERATURE

1. OECD (2015), Test Guideline on an In Chemico Skin Sensitization: Direct Peptide Reactivity Assay (DPRA), OECD Guidelines for the Testing of Chemicals No 442C, OECD, Paris.

2. OECD (TBD), Test Guideline on an In Chemico Skin Sensitization: Amino acid Derivative Reactivity Assay (ADRA), OECD Guidelines for the Testing of Chemicals No 442C, OECD, Paris.

3. 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.

4. EURL ECVAM (2012), Direct Peptide Reactivity Assay (DPRA) ECVAM Validation Study Report. Available at: https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/files-dpra/DPRA%20Validation%20Study%20Report.pdf

5. Fujita, *et al*., (TBD), ADRA validation study (under preparation)

6. 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.

7. 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.

8. 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.

9. Natsch *et al*. (2013), A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *Journal of Applied Toxicology*, **33**(11):1337-52, DOI:10.1002/jat.2868.

10. ECETOC (2003), Contact sensitization: Classification according to potency. European Centre for Ecotoxicology & Toxicology of Chemicals (Technical Report No. 87).

11. G. F. Gerberick *et al*. (2007), Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. *Toxicological Sciences*, 97(2):417-427.

12. Urbisch D *et al*. (2015), Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regulatory Toxicology and Pharmacology*, 71(2):337-51

13. Bauch *et al*. (2012), Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. *Regulatory Toxicology and Pharmacology*, 63(3):489-504.