OECD/OCDE

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

DRAFT PROPOSAL FOR A NEW TEST GUIDELINE

In Vitro Skin Sensitization: Keratinocyte-based ARE-Nrf2 Luciferase Reporter Gene Test Method

INTRODUCTION

1. A skin sensitizer refers to a substance that will lead to an allergic response following skin contact (1). There is general agreement regarding the key biological events underlying skin sensitization summarized as an Adverse Outcome Pathway (AOP) which includes four events recognized as key ones, going from the molecular initiating event to the adverse health effect, i.e., allergic contact dermatitis in humans or contact hypersensitivity in rodents (2)(3). The first key event is the molecular interaction with skin proteins, and in particular the covalent binding to cysteine and/or lysine residues. The second key event takes place in the keratinocyte and includes inflammatory responses as well as gene expression associated with specific cell signaling pathways such as the antioxidant/electrophile response element (ARE)-dependent pathways. The third key event is the activation of dendritic cells, typically assessed by expression of specific cell surface markers, chemokines and cytokines. The fourth key event is T-cell proliferation, which is indirectly measured in the murine Local Lymph Node Assay (4).

2. The present Test Guideline (TG) provides with an *in vitro* procedure (the keratinocyte-based ARE- Nrf2 luciferase reporter gene assay) to be used for supporting the identification of skin sensitization hazard of chemicals. So far, the assessment of skin sensitization has typically involved the use of laboratory animals. The classical methods based on guinea-pig, the Magnusson Kligman Guinea Pig Maximisation Test (GMPT) and the Buehler Test TG 406 (5), study both the induction and elicitation phases of skin sensitization and are in use since decades. More recently, a murine test, the Local Lymph Node Assay (LLNA) (TG 429)(4) and the two non-radioactive modifications, LLNA: DA (TG 442A)(6) and LLNA: BrdU-ELISA (TG 442B)(7) all measuring the induction response, have gained acceptance since they provide advantages over the guinea pig tests in terms of both animal welfare and scientific benefits. More recently *in chemico* and *in vitro* test methods have been considered scientifically valid to evaluate the skin sensitization hazards of chemicals. Given the complexity of the biological mechanisms underlying skin sensitization effects and the limitations of the currently available non-animal test methods (*in vitro, in silico in chemico*), it is likely that combinations of mechanistically-based test methods within Integrated Approaches to Testing and Assessment (IATA) are needed to be able to substitute the regulatory animal tests currently in use (2)(3).

3. The keratinocyte-based ARE-Nrf2 luciferase reporter gene test method is proposed to address the second key event as explained in paragraph 1. A majority of skin sensitizers

were reported to induce genes that are regulated by the antioxidant response element (ARE) (8)(9). Small electrophilic chemicals such as skin sensitizers can act on the sensor protein Keap1 (Kelch-like ECH-associated protein 1), by i.e., covalent modification of its Cysteine residue, resulting in its dissociation from the transcription factor Nrf2 (nuclear factor-erythroid 2-related factor 2). The dissociated Nrf2 can then activate ARE-dependent genes coding for i.e., phase II detoxifying enzymes (8)(10)(11).

4. The ARE-Nrf2 luciferase reporter gene test method KeratinoSens™ has undergone a validation study (9)(12)(13) followed by an independent peer review conducted by the European Reference Laboratory for Alternative Methods (EURL ECVAM), and was considered scientifically valid to be used as part of an IATA approach, to support the identification of the sensitization potential of test chemicals for hazard classification and labeling purposes (14).

5. Definitions are provided in Annex 1.

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

6. Since activation of the Keap1-Nrf2-ARE pathway addresses only the second key event of the skin sensitization AOP, information from test methods based on the activation of this pathway are unlikely to be sufficient to conclude on the skin sensitization potential of chemicals. Therefore data generated with the present Test Guideline should be considered in the context of integrated approaches, such as the IATA, combining them with complementary information derived from *in vitro* assays addressing other key events of skin sensitization (e.g. in *chemico* reactivity assays such as the Direct Peptide Reactivity Assay) as well as non-testing methods including read-across from chemical analogues. Examples on how to use the keratinocyte-based ARE-Nrf2 luciferase reporter gene test method in combination with other information are reported in literature (13)(15)(16)(17)(18).

7. Based on the dataset from the validation study and in-house testing used for the independent peer-review of the test method, the KeratinoSens™ test method proved to be transferable to laboratories experienced in cell culture and reproducible within- and between-laboratories (14). Furthermore, the accuracy of the KeratinoSens™ to discriminate skin sensitizers from non-sensitizers as judged vs. evidence from the LLNA was found to be of 75% (sensitivity 75%, n=77; specificity 75% n=104) by EURL ECVAM (14). These figures are similar to those recently published based on in-house testing of about 145 chemicals (77% accuracy, 79% sensitivity, 72% specificity) (13). Taken together, this information indicates the usefulness of the KeratinoSens™ assay to contribute to the identification of skin sensitization hazard.

8. On the basis of the data used for independent peer-review, the KeratinoSens™ test method was shown to be applicable to a broad variety of chemicals covering relevant ranges of organic functional groups (Annex 2), reaction mechanisms, skin sensitization potency (as determined with *in vivo* studies) and physico-chemical properties (9)(12)(13)(14). Mainly substances were tested, although data also exist on the testing of mixtures from natural

sources (19). The test method is applicable to chemicals soluble either in water or DMSO. In general test chemicals with a cLogP of up to 5 have been successfully tested whereas extremely hydrophobic molecules with a cLogP above 7 are outside the applicability of the test method (14).

9. Negative results should be interpreted with caution as chemicals with an exclusive reactivity towards lysine-residues can be detected as negative by the test method. Furthermore, because of the limited metabolic capability of the cell line used and because of the experimental conditions, pro-haptens (i.e. chemicals requiring enzymatic activation) requiring biotransformation by P450 enzymes, and pre-haptens (i.e. chemicals requiring biochemical activation) with a slow oxidation rate may also not be systematically detected. Finally, highly cytotoxic chemicals cannot be reliably tested (EURL ECVAM, in preparation). In cases where evidence can be demonstrated on the non-applicability of the Test Guideline to other specific categories of chemicals, the test method should not be used for those specific categories.

10. In addition to supporting discrimination between skin sensitizers and non-sensitizers, the KeratinoSens^{TM} assay also provides concentration-response information that may contribute to the assessment of sensitizing potency (18). However, further work based on, preferentially human data, is required to determine to which extent KeratinoSens^{TM} results relate to potency categories.

PRINCIPLE OF THE TEST

11. The keratinocyte-based ARE-Nrf2 luciferase reporter gene test method makes use of immortalized adherent cell line derived from HaCaT human keratinocytes transfected with a selectable plasmid. The plasmid contains the luciferase gene under the transcriptional control of the SV40 promoter fused with the ARE from the AKR1C2 gene which was identified as one of the genes up-regulated by contact sensitizers in dendritic cells (20)(21). This allows to quantitatively measure (by luminescence detection) luciferase gene induction, using well established light producing luciferase substrates, as an indicator of the activity of the Nrf2 transcription factor in cells following exposure to electrophilic chemicals. The luciferase signal reflects the activation by sensitizers of endogenous Nrf2 dependent genes, and the strict dependence of the luciferase signal in the recombinant cell line on Nrf2 has been demonstrated (22).

12. The keratinocyte-based ARE-Nrf2 luciferase reporter gene test method can be used, within an IATA approach, to support the identification of the skin sensitization hazards of chemicals. Chemicals are identified as skin sensitizers if they induce a statistically significant induction of the luciferase gene above a given threshold (i.e., > 1.5 fold or 50% increase in the case of KeratinoSens^{TM}), below a defined concentration which does not significantly affect cell viability (i.e., below 1000 μM and at a concentration at which the cellular viability is above 70% in the case of KeratinoSens^{TM}). For this purpose, the maximal fold induction of the luciferase gene over solvent control (I_{max}) is determined. Furthermore, since cells are exposed to series of concentrations of the test chemicals, the concentration needed for a statistically significant luciferase gene induction above the threshold (i.e., EC1.5 value in the

case of KeratinoSens™) should be extrapolated from the dose response curve. Finally, parallel cytotoxicity measurements should be conducted to assess whether gene induction levels occur at sub-cytotoxic concentrations.

13. Prior to routine use of the keratinocyte-based ARE-Nrf2 luciferase reporter gene test method that adheres to this Test Guideline, laboratories should demonstrate technical proficiency, using the Proficiency Chemicals listed in Annex 2. A laboratory can use these chemicals to demonstrate their technical competence in performing the keratinocyte-based ARE-Nrf2 luciferase reporter gene test method prior to submitting results for regulatory purposes.

14. This Test Guideline also includes a set of Performance Standards (PS) (Annex 5) that should be used to determine the reliability, relevance and limitations of similar or modified *in vitro* keratinocyte-based ARE-Nrf2 luciferase reporter gene test methods to be used for regulatory purposes. The similarity of the new similar or modified test method to the Validated Reference Method (VRM) KeratinoSens™ should be demonstrated in accordance with the requirements of the PS set out in this Test Guideline (Annex 5). Mutual Acceptance of Data (MAD) will only be guaranteed for test methods, validated according to the PS of this Test Guideline, if these test methods have been reviewed and included in this Test Guideline by the OECD.

PROCEDURE

15. The test method considered scientific valid to be used within this Test Guideline is the KeratinoSens™ test method (9)(12)(13)(14)). The Standard Operating Procedures (SOP) for the KeratinoSens™ is available (23). It is recommended that this SOP is consulted when implementing and using the test method in the laboratory. The following is a description of the main components and procedures of the keratinocyte-based ARE-Nrf2 luciferase reporter gene test method.

Preparation of the Keratinocyte Cultures

16. Transgenic cell lines having a stable insertion of the luciferase reporter gene under the control of the ARE-element should be used (e.g., the KeratinoSens cell line-). Upon receipt, cells are propagated to a defined passage number (e.g., 2 to 4 in the case of KeratinoSens™) and multiple vials of the resulting cell population stored in liquid nitrogen as a homogeneous stock. The cells propagated from this original stock can be kept in culture only for a defined maximum passage number (e.g., 25 in the case of KeratinoSens™). Cells from this stock are then used for routine testing. Frozen cells are thawed, re-suspended in maintenance medium, centrifuged (e.g., to eliminate the Dimethyl sulfoxide (DMSO) used for freezing), and cultured in the appropriate maintenance medium (in the case of KeratinoSens™ this represents DMEM containing 9.1% Fetal Calf Serum (FCS), to which Geneticin is added only at the next passage).

17. When cells reach 80-90% confluence, cells are washed (twice in the KeratinoSens™ assay using a buffer solution containing EDTA, Ethylenediamin-tetra-acetic acid trinatrium salt), detached (using e.g., Trypsin-EDTA), re-suspended in maintenance medium and split at a ratio varying from 1:3 to 1:12 in fresh medium to be cultured again up to 80-90% confluence. The higher the split ratio used, the longer time cells will need to reach 80-90% confluence again. In the case of KeratinoSens™, antibiotics against microbial contaminations are not used in the standard cultivation of the cells, nor are they used when cells are seeded for testing.

18. For testing, cells should be 80-90% confluent, and care should be taken to ensure that cells are never grown to full confluence. One day prior to testing cells are harvested as described above, re-suspended (at e.g., a density of 80'000 cells / ml in media containing 9.1% FCS without Geneticin for KeratinoSens™), and distributed into 96-well plates (e.g., 10'000 cells / well in the case of KeratinoSens™). Attention should be paid to avoid sedimentation of the cells during seeding to ensure homogeneous cell number distribution across all wells. If this is not the case, this step may give raise to high well-to-well variability. in the case of KeratinoSens™, four parallel plates are prepared for each batch where up to seven chemicals can be tested. Three micro-well plates are used for luciferase activity (white plates), and one micro-well plate for cell viability assay (transparent plate).

Preparation of the Test Chemicals and Control Substances

19. The test chemicals and control substances are prepared on the day of testing. For the KeratinoSens™ test method, test chemicals are dissolved in DMSO to a final concentration of 200 mM. The DMSO solutions can be considered self-sterilizing, so that no sterile filtration is needed. Chemicals not soluble in DMSO are dissolved and diluted in sterile water and the solutions sterilized by filtration through a 0.2 µM filter. Chemicals which have no defined molecular weight (such as small polymers) are tested considering a pro forma molecular weight of 200, or, in other words, the stock solution is prepared to a concentration of 40 mg / ml or 4 % (w/v).

20. Based on the 200mM stock DMSO solutions of the test chemicals, serial dilutions are made using a two-fold dilution factor, to obtain 12 master concentrations of the tested chemical ranging from 0.098 to 200 mM. For test chemicals not soluble in DMSO, all the dilutions are made using sterile water.

21. The negative (solvent) control used in the case of KeratinoSens™ is DMSO, for which six wells per plate are prepared

22. The positive control used in the case of KeratinoSens™ is cinnamic aldehyde (CAS n. 14371-10-9), for which a series of 5 master concentrations ranging from 0.4 to 6.4 mM are prepared based on a 6.4 mM stock solution and using also the two-fold dilution factor.

23. The obtained concentrations of the test chemicals and control substances are further diluted 25-fold into culture medium containing 1% FCS, and then used for treatment with a further 4-fold dilution factor (see paragraph 25). As a consequence, the final concentrations

of the tested chemicals range from 0.98 to 2000 μM, for the positive control they range from 4 to 64 μM, and the final negative (solvent) control concentration is at 1% (which corresponds to the same concentration of DMSO found in all tested chemicals and positive controls).

Application of the Test Chemicals and Control Substances

24. For each test chemical and control substance, three experiments consisting each of testing the full dilution series (i.e., 12 in the case of in the case of KeratinoSens™) in three replicate plates are needed to derive a prediction (n=9).

25. After seeding as described in paragraph 17, cells are grown for 24 h in the 96-wells microtiter plates. The medium is then removed and replaced with fresh culture medium containing the test chemicals and control substances. For example, in the case of in the case of KeratinoSens™, 150 µl fresh culture medium containing 1% FCS and without Geneticin is added to which 50 µl of the 25-fold diluted test chemicals and control substances are added. At least one well per plate should be left empty (no cells and no treatment) to assess the background values.

26. The treated plates are then incubated for an established exposure time (i.e., 48 hours at 37°C in the presence of 5% CO_2 in the case of KeratinoSens™). Care should be taken to avoid evaporation of volatile compounds and cross-contamination between wells by volatile compounds by e.g., covering the plates with a foil prior to the incubation with the test chemicals).

Luciferase Activity Measurements

27. Three factors are critical to ensure appropriate luminescence readings:
- the choice of a sensitive luminometer,
- the use of a plate format with sufficient height to avoid light-cross-contamination; and
- the use of a luciferase substrate with sufficient light output to ensure sufficient sensitivity and low variability.

Prior to testing, a control experiment setup as described in Annex 4 should be carried out to ensure that these three points are met.

28. After the exposure time with the tested and control substances chemicals (i.e., 48h incubation in the case of KeratinoSens™), cells in the triplicate plates for luminescence readings are washed with a phosphate buffered saline, and the relevant lysis buffer added to each well for e.g., 20 min at room temperature. Care should be taken to avoid formation of foam. Furthermore, between processing of successive plates, the time should be equal or greater than the cycle time for the luminometer to read one plate (typical cycle time of 10 min for one plate) in order to ensure constant lysis time for each plate.

29. Plates with the cell lysate are then placed in the luminometer for reading, and in the case of KeratinoSens™ the luminometer is programmed to: (i) add the luciferase substrate to each

well (i.e., 50 μl), (ii) wait for 1 second, and (iii) integrate the luciferase activity for 2 seconds. In case alternative settings are used, e.g., depending on the model of luminometer used, these should be justified.

Cytotoxicity Measurements Assessment

30. For the cell viability assay, after the exposure time (i.e., 48h incubation time in the case of KeratinoSens™), the medium in the parallel plate for cytotoxicity reading is replaced with fresh medium containing MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1) (i.e., 0.135 μg/well in the case of KeratinoSens™). Plates are incubated for a fixed period of time (i.e., 4 hours at 37°C in the presence of 5% CO_2 in the case of KeratinoSens™). The MTT medium is removed and cells are lysed (e.g., adding 10% SDS solution to each well to dissolve cells overnight, protected from light, at 37°C and in the presence of 5% CO_2 in the case of KeratinoSens™). After shaking, the absorption is measured with the help of a photometer at i.e., 600nm in the case of KeratinoSens™.

DATA AND REPORTING

Data Evaluation

31. An automated excel sheet is used to perform the data evaluation. The plate readout of the triplicate serial dilutions obtained for luciferase activity measurements and the cytotoxicity measurements are introduced in the dedicated areas. The file automatically calculates the gene induction and the wells having statistically significant induction over a given threshold, and based on this outcome the evaluation sheet derives:

- the maximal gene induction (I_{max}) value observed at any concentration of the tested chemical and positive control; and
- the EC1.5 value representing the concentration for which a gene induction above the 1.5 threshold (i.e., 50% enhanced gene activity) was obtained.
- The IC_{50} and IC_{30} concentration values for 50% and 30% reduction of cellular viability.
- It indicates whether significant 1.5-fold gene induction occurs below the IC_{30} concentration.

32. The results from the different repetitions are summarized in the 'Summary sheet'. This sheet also generates for each chemical a plot summarizing the gene induction and cytotoxicity dose-response in all repetitions. The automatically calculated I_{max} and the EC 1.5 values should visually be checked with the help of the generated graphs, as uneven dose-response curves or large variation may lead to wrong extrapolations which may need to be corrected manually.

33. Special care should be taken in the very rare cases where a statistically non-significant induction above 1.5-fold which is observed followed by a higher concentration with a statistically significant induction. In these cases, if a clear dose-response for induction is

apparent from the graphs, the four parameters needed for the extrapolation of EC1.5 values (concentration and fold-induction below the threshold of 1.5 as well as concentration and fold-induction above the threshold) may then be used to manually calculate the EC1.5 value according to the formula described below (same formula as used in extrapolation of LLNA EC3 values). However, these runs are only considered as valid and positive if the fold induction at any higher concentration is statistically significant and above the threshold of 1.5.

$$EC1.5 = (a - c) \times \left[\frac{(1.5 - d)}{(b - d)}\right] + c$$

Where
a = Concentration in µM of the highest concentration with less than 1.5-fold induction
b = fold-induction at the highest concentration with less than 1.5-fold induction
c = Concentration in µM of the lowest concentration with higher than 1.5-fold induction
d = fold-induction in µM at the lowest concentration with higher than 1.5-fold induction

34. Furthermore, in the very rare cases of biphasic dose-response curves which do cross the threshold of 1.5 twice, the EC1.5 value is also not correctly calculated. These cases are easily spotted by inspection of the dose-response-plot, and the appropriate measures should be taken to ensure correct data reporting. Finally, for the rare chemicals, which generate a 1.5-fold or higher induction already at the lowest test dose of 0.98 µM, the EC1.5 value cannot be calculated automatically, for these chemicals the EC1.5 value of <0.98 is manually set based on visual inspection of the dose-response curve.

Acceptance Criteria

35. The following acceptance criteria should be met when using the KeratinoSens™ test method. First, the gene induction obtained with the positive control, cinnamic aldehyde, must be statistically significant above the threshold of 1.5 in at least one dose.

36. Second, the average induction in the three replicates for cinnamic aldehyde at 64 µM should be between 2 and 8, and the EC 1.5 value should be between 7 µM and 30 µM. At least one of these criteria must be met, otherwise the run is discarded. If only one criteria is fulfilled, it is recommended to carefully check the dose-response of cinnamic aldehyde in order to decide on acceptability.

37. Finally, the average coefficient of variability of the luminescence reading for the negative (solvent) control DMSO should be below 20% in each experiment (3 × 6 wells). If the variability is higher, results should be discarded.

Decision Criteria

38. Test chemicals are identified as potential skin sensitizers in the KeratinoSens™ test method if the following conditions are met:

- the I_{max} is statistically significantly higher than 1.5-fold as compared to the basal luciferase activity and the EC1.5 value is below 1000 µM in at least two out of the three

repetitions (if the I_{max} is exactly equal to 1.5, the chemical is still rated negative and no EC1.5 value is calculated by the evaluation sheet). If an EC1.5 value is calculated automatically in the summary sheet, this indicates that the gene induction is statistically significant at the corresponding concentration according to a T-test.
- At the lowest concentration with a gene induction above 1.5 fold (i.e. at the EC 1.5 determining concentration), the cellular viability should be above 70%. If this is not the case, a warning ('cytotox') appears in the summary sheet.
- There should be an apparent overall dose-response for luciferase induction, which is similar between the repetitions.

Compounds that only induce the gene activity at cytotoxic levels are not rated positive (e.g., non-sensitizing skin irritants).

39. Since the above parameters are automatically calculated by the automated excel sheet, a careful inspection of the dose-response curves for both endpoints (luciferase activity and cytotoxicity), both in the individual repetitions and in the summary file should be conducted as a quality control. Indeed, uneven dose response curves can lead to wrong extrapolations which can be detected by visual inspection.

40. In rare cases, test chemicals, which induce the gene activity very close to the cytotoxic levels, can be positive in some repetitions at non-cytotoxic levels, and in other repetitions only at cytotoxic levels. Such test chemicals shall be retested with more narrow dose-response analysis using e.g., 1.414-fold dilutions between wells instead of two-fold dilutions, to determine if induction has occurred at cytotoxic levels or not (9).

Test Report

41. The test report should include the following information:

Test Chemicals and Control Substances

- Chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known;
- The CAS Registry Number (RN), if known;
- Purity and composition of the test chemical and control substances (in percentage(s) by weight), to the extent this information is available;
- Physico-chemical properties relevant to the conduct of the study (e.g., DMSO solubility, water solubility, physical state, volatility, pH, stability, chemical class);
- Treatment of the test chemical and control substances prior to testing, if applicable (*e.g.,* warming, grinding);
- Storage conditions and stability, if known;

Test Method Conditions

- Name and address of the sponsor, test facility and study director;

- Description of test method used;
- Cell line used, its storage conditions and source (*e.g.,* the facility from which they were obtained);
- Passage number and confluence of cells used for testing;
- Cell counting method used for seeding prior to testing and measures taken to ensure homogeneous cell number distribution (cf. paragraph 18);
- Luminometer used (*e.g.,* model), including instrument settings, luciferase substrate used, and demonstration of appropriate luminescence measurements based on the control test described in Annex 4;
- Reference to historical negative and positive control results and, if applicable, historical data demonstrating acceptable concurrent benchmark control ranges;
- The procedure used to ensure the integrity (*i.e.,* accuracy and reliability) of the test method over time (*e.g.,* periodic testing of proficiency chemicals);

Test Procedure

- Number of replicates used;
- Identity of the negative (solvent) and positive controls used (if applicable, also the benchmark control);
- Test chemical dose, application and exposure time used (if different than the one recommended)
- Duration of exposure (if different than the one recommended);
- Description of evaluation and decision criteria used;
- Description of study acceptance criteria used;
- Description of any modifications of the test procedure;

Results

- Tabulation of I_{max}, EC1.5 and viability values obtained for the test chemical and for the positive control for each experiment as well as the mean values and SD obtained for all experiments and an indication of the rating of the test chemical according to the prediction model;

- Coefficient of variation obtained with the luminescence readings for the negative control for each experiment;
- Description of inspection of the automated excel sheet calculations;
- Description of any other effects observed;

Discussion of the Results

Conclusion

OECD/OCDE

LITERATURE

1. United nations (UN) (2011). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Fourth revised edition, UN New York and Geneva, 2011. Available at: [http://www.unece.org/trans/danger/publi/ghs/ghs_rev04/04files_e.html]

2. OECD (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. Series on Testing and Assessment No. 168. Accessible at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2012)10/PART1&docLanguage=En

3. Adler S, Basketter D, Creton S, Pelkonen O, van Benthem J, Zuang V, Andersen KE, Angers-Loustau A, Aptula A, Bal-Price A, Benfenati E, Bernauer U, Bessems J, Bois FY, Boobis A, Brandon E, Bremer S, Broschard T, Casati S, Coecke S, Corvi R, Cronin M, Daston G, Dekant W, Felter S, Grignard E, Gundert-Remy U, Heinonen T, Kimber I, Kleinjans J, Komulainen H, Kreiling R, Kreysa J, Leite SB, Loizou G, Maxwell G, Mazzatorta P, Munn S, Pfuhler S, Phrakonkham P, Piersma A, Poth A, Prieto P, Repetto G, Rogiers V, Schoeters G, Schwarz M, Serafimova R, Tähti H, Testai E, van Delft J, van Loveren H, Vinken M, Worth A, Zaldivar JM. (2011). Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Archives of Toxicology 85, 367-485.

4. OECD (2010).OECD Guidelines for Chemical Testing No. 429. Skin sensitization: Local Lymph Node assay. Paris, France: Organisation for Economic Cooperation and Development; 2010.

5. OECD (1992). OECD Guidelines for the Testing of Chemicals No. 406: Skin Sensitisation. Paris, France: Organisation for Economic Cooperation and Development.

6. OECD (2010). OECD Guidelines for Chemical Testing No. 442A. Skin sensitization: Local Lymph Node assay: DA. Paris, France: Organisation for Economic Cooperation and Development; 2010.

7. OECD (2010). OECD Guidelines for Chemical Testing No. 442B. Skin sensitization: Local Lymph Node assay: BrdU-ELISA. Paris, France: Organisation for Economic Cooperation and Development; 2010.

8. Natsch (2010). The Nrf2-Keap1-ARE Toxicity Pathway as a Cellular Sensor for Skin Sensitizers-Functional Relevance and Hypothesis on Innate Reactions to Skin Sensitizers. Toxicological Sciences 113, 284-292.

9. 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.

10. Dinkova-Kostova AT, Holtzclaw WD, Kensler TW. (2005). The role of Keap1 in cellular protective responses. Chem Res Toxicol. 18, 1779-1791.

11. Kansanen E, Kuosmanen SM, Leinonen H, Levonen AL. (2013). The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer. Redox Biol. 18,1, 45-49.

12. 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.

13. Natsch, A., C.A., R., 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.

14. EURL-ECVAM Recommendation on the KeratinoSens™ assay for skin sensitisation testing (*in preparation*)

15. Natsch A, Emter R, Ellis G. (2009). Filling the concept with data: integrating data from different in vitro and in silico assays on skin sensitizers to explore the battery approach for animal-free skin sensitization testing. Toxicol Sci. 107, 106-121.

16. Ball N, Cagen S, Carrillo JC, Certa H, Eigler D, Emter R, Faulhammer F, Garcia C, Graham C, Haux C, Kolle SN, Kreiling R, Natsch A, Mehling A. (2011). Evaluating the sensitization potential of surfactants: integrating data from the local lymph node assay, guinea pig maximization test, and in vitro methods in a weight-of-evidence approach. Regul Toxicol Pharmacol. 60, 389-400.

17. Bauch C, Kolle SN, Ramirez T, Eltze T, Fabian E, Mehling A, Teubner W, van Ravenzwaay B, Landsiedel R. (2012). Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. Regul Toxicol Pharmacol. 63, 489-504.

18. Jaworska, J., Dancik, Y., Kern, P., Gerberick, F., Natsch, A., 2013. Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. J Appl Toxicol. 33, 1353-1364

19. Andres, E., Sa-Rocha, V.M., Barrichello, C., Haupt, T., Ellis, G., Natsch, A., 2013. The sensitivity of the KeratinoSens™ assay to evaluate plant extracts: A pilot study. Toxicology *In Vitro* 27, 1220-1225.

20. Gildea LA, Ryan CA, Foertsch LM, Kennedy JM, Dearman RJ, Kimber I, Gerberick GF. (2006). Identification of gene expression changes induced by chemical allergens in dendritic cells: opportunities for skin sensitization testing. J Invest Dermatol, 126, 1813-1822.

21. Ryan C.A., Gildea L.A., Hulette B.C., Dearman R.J., Kimber I. And Gerberick G.F. (2004). Gene expressing changes in peripheral blood-derived dendritic cells following exposure to a contact allergen. Toxicol. Lett. 150, 301-316.

OECD/OCDE

22. Emter R, van der Veen JW, Adamson G, Ezendam J, van Loveren H, Natsch A. (2013). Gene expression changes induced by skin sensitizers in the KeratinoSens™ cell line: Discriminating Nrf2-dependent and Nrf2-independent events. Toxicol In Vitro 27, 2225-2232.

23. DB-ALM (INVITTOX) (in preparation). Protocol 155: KeratinoSens, 17pp. Available: [http://ecvam-dbalm.jrc.ec.europa.eu/].

24. 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. Organization for Economic Cooperation and Development, Paris, France, 2005, 96 pp.

25. NAFTA (North American Free Trade Agreement) (2012). Technical Working Group on Pesticides – (Quantitative) Structure Activity Relationship ((Q)SAR) Guidance Document. 186pp. Available at: http://www.epa.gov/oppfead1/international/naftatwg/guidance/qsar-guidance.pdf

26. Aptula AO, Patlewicz G, Roberts DW, Schultz TW _(2006). Non-enzymatic glutathione reactivity and *in vitro* toxicity: a non-animal approach to skin sensitization. Toxicology *In Vitro* 20, 239-247.

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

28. Rovida C, Martin SF, Vivier M, Weltzien HU, Roggen E (2013). Advanced tests for skin and respiratory sensitization assessment. ALTEX 30, 231-352.

OECD/OCDE

ANNEX 1 - DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of "relevance." The term is often used interchangeably with "concordance", to mean the proportion of correct outcomes of a test method (24).

AOP (Adverse Outcome Pathway): sequence of events from the chemical structure of a target chemical or group of similar chemicals through the molecular initiating event to an *in vivo* outcome of interest (2).

ARE: Antioxidant response element (also called EpRE, electrophile response element), is a response element found in the upstream promoter region of many cytoprotective and phase II genes. When activated by Nfr2, it mediates the transcriptional induction of these genes.

EC 1.5: Extrapolated concentration for a 1.5 fold luciferase induction.

IC_{30}: Concentration for reduction of cellular viability by 30% as determined with the MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1) assay.

IC_{50}: Concentration for reduction of cellular viability by 50% as determined with the MTT assay.

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

IATA (Integrated Approaches to Testing and Assessment): Integrate existing knowledge based on classes of chemicals with the results of biochemical and cellular assays, computational predictive methods, exposure studies, and other sources of information to identify requirements for targeted testing or develop assessment conclusions. In some cases, the application of IATA could lead to the refinement, reduction, and/or replacement of selected conventional tests (e.g., animal toxicity tests). IATA also have the potential to further enhance the understanding of mode/mechanism of action including the consideration of relevant adverse outcome pathways (AOPs) that provide biological linkages between molecular initiating events to adverse outcomes in individual organisms and populations that are the bases for risk assessments (25).

I_{max}: Maximal induction of luciferase activity over solvent control over the complete dose-response range measured.

Keap1: Kelch-like ECH-associated protein 1, is a sensor protein that can regulate the Nrf2 activity. Under un-induced conditions the Keap1 sensor protein targets the Nrf2 transcription factor for ubiquitinylation and proteolytic degradation in the proteasome. Covalent

modification of the reactive cysteine residues of Keap 1 by small molecules can lead to dissociation of Nrf2 from Keap1 (8)(10)(11).

Mixture: A mixture or a solution composed of two or more substances in which they do not react (1).

Nrf2: nuclear factor (erythroid-derived 2)-like 2, is a transcription factor involved in the antioxidant response pathway. When Nrf2 is not ubiquitinylated, it builds up in the cytoplasm and translocates to into the nucleus, where it combines to the ARE in the upstream promoter region of many cytoprotective genes, initiating their transcription (8)(10)(11).

Positive control: A replicate containing all components of a test system and treated with a chemical known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the severe response should not be excessive.

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 (24).

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 and intra-laboratory repeatability (24).

Reproducibility: The agreement among results obtained from testing the same substance using the same test protocol (see reliability) (24).

Sensitivity: The proportion of all positive / active chemicals that are correctly classified by the test method. 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 (24).

Solvent/vehicle control: An untreated sample containing all components of a test system, including the solvent or vehicle that is processed with the test chemical-treated and other control samples to establish the baseline response for the samples treated with the test chemical dissolved in the same solvent or vehicle.

Specificity: The proportion of all negative / inactive chemicals that are correctly classified by the test method. 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 (24).

Substance: Chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition (1).

OECD/OCDE

Test Chemical: Chemical (substance or mixture) assessed in the test method.

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

Valid Test Method: A test method considered to have sufficient relevance and reliability for a specific purpose and which is based on scientifically sound principles. A test method is never valid in an absolute sense, but only in relation to a defined purpose (24).

Validated Test Method: A test method for which validation studies have been completed to determine the relevance (including accuracy) and reliability for a specific purpose. It is important to note that a validated test method may not have sufficient performance in terms of accuracy and reliability to be found acceptable for the proposed purpose (24).

OECD/OCDE

ANNEX 2 - ORGANIC FUNCTIONAL GROUPS

In Vitro Skin Sensitization: Keratinocyte-based ARE-Nrf2 Luciferase Reporter Gene Test Method

Organic functional groups (nested) representation determined using the OECD QSAR Toolbox (version 3.1.0.21) of n=210 chemicals tested in the KeratinoSens™ test method as evaluated in the independent peer review conducted by EURL ECVAM. In addition to the 210 listed test chemicals, 9 metals have been tested for which no organic functional groups can be determined.

Organic functional groups	n.	Organic functional groups	n.
Acetal	1	Benzthiazolinone/ Benzoisothiazolinone	3
Acetoxy	7	Benzoxazole/ Benzisoxazole	1
Acrylate	5	Benzyl	21
Alcohol	31	Bicycloheptane	1
Aldehyde	15	Bridged-ring carbocycles	1
Aliphatic Amine, primary	3	Carbamate	2
Aliphatic Amine, secondary	1	Carboxamide	3
Aliphatic Amine, tertiary	2	Carboxylic acid	16
Alkane, branched with tertiary carbon	10	Carboxylic acid ester	19
Alkane branched with quaternary carbon	2	Chromene	1
Alkene	9	Cycloalkane	6
Alkenyl halide	2	Cycloalkene	4
Alkoxy	10	Cycloketone	3
Alkyl halide	8	Diacyl peroxides	1
Alkyne	2	Dialdehydes	2
Allyl	24	Dihydrobenzopyranone	1
Alpha,beta unsaturated aldehyde	5	Dihydrofuran	1
Aminoaniline, meta	1	Dihydroxyl group	9
Aminoaniline, para	2	Diketone	5
Ammonium salt	1	Disulfide	1
Aniline	9	Epoxide	9
Aromatic amine	2	Ether	35
Aromatic heterocyclic halide	10	Formylamino	1
Aryl	37	Furanone/ Furanondione	1
Aryl halide	10	Fused carbocyclic aromatic	1
Benzamide	2	Glycerol and derivatives	1
Benzothiazole/ Benzoisothiazole	1	Heterocyclic Phenol	1

OECD/OCDE

Organic functional groups	n.	Organic functional groups	n.
Hydantoin	1	Precursors quinoid compounds	10
Isobenzofuran	1	Rosins	1
Isobenzofurandione	2	Sarcosine	1
Isocyanate	2	Saturated heterocyclic amine	1
Isopropyl	8	Saturted Heterocyclic fragment	14
Ketone	4	Sulfate	1
Lactone	4	Sulfonate ester	1
Maleate/ Fumarate	2	Sulfonic acid	3
Methacrylate	3	tert-Butyl	2
Naphtalene	1	Thiaazabicycloheptane, oxo	1
Nitrile	6	Thiazolidinone/ Isothiazolone	2
Nitrobenzene	13	Thioalcohol	3
Nitroso	1	Thiocarbamate	1
N/A	2	Thiocyanate	1
Overlapping groups	138	Unsatureted Heterocyclic fragment	2
Phenol	16	Urea derivatives	1
Phenothiazine	1		

OECD/OCDE

ANNEX 3 - PROFICIENCY CHEMICALS

In Vitro Skin Sensitization: Keratinocyte-based ARE-Nrf2 Luciferase Reporter Gene Test Method

Prior to routine use of a test method that adheres to this Test Guideline, laboratories should demonstrate technical proficiency by correctly identifying the sensitization hazard classification of the 10 substances recommended in Table 1. These substances were selected to represent the range of responses for skin sensitization hazards. Other selection criteria were that substances are commercially available, that there are high quality *in vivo* reference data available, and that there are high quality *in vitro* data available from the KeratinoSens™ test method.

Table 1: Recommended substances for demonstrating technical proficiency with the KeratinoSens™ Test Method

Chemical	**CASRN**	**Physical Form**	***In Vivo* Classification (13)**	***In Vitro* Classification (12)**	***$EC_{1.5}$ reference range****	***IC_{50} reference range****
Isopropanol	67-63-0	Liquid	Non-sensitizer	Non-sensitizer	> 1000	> 1000
Salicylic acid	69-72-7	Solid	Non-sensitizer	Non-sensitizer	> 1000	> 1000
Lactic acid	50-21-5	Liquid	Non-sensitizer	Non-sensitizer	> 1000	> 1000
Glycerol	56-81-5	Liquid	Non-sensitizer	Non-sensitizer	> 1000	> 1000
Cinnamyl alcohol	104-54-1	Solid	Sensitizer (weak)	Sensitizer	25 - 175	> 1000
Ethylene glycol dimethacrylate	97-90-5	Liquid	Sensitizer (weak)	Sensitizer	5 - 125	> 500
2-Mercaptobenzothiazole	149-30-4	Solid	Sensitizer (moderate)	Sensitizer	25 - 250	> 500
Methyldibromo glutaronitrile	35691-65-7	Solid	Sensitizer (strong)	Sensitizer	< 20	20 - 100
4-Methylaminophenol sulfate	55-55-0	Solid	Sensitizer (strong)	Sensitizer	< 12.5	20 - 200
2,4-Dinitro-chlorobenzene	97-00-7	Solid	Sensitizer (extreme)	Sensitizer	< 12.5	5 - 20

* Based on the historical observed values (12).

OECD/OCDE

ANNEX 4 - QUALITY CONTROL OF LUMINESCENCE MEASUREMENTS

Basic Experiment for Ensuring Optimal Luminescence Measurements in the KeratinoSens™ Assay

The following three parameters are critical to ensure obtaining reliable results with the luminometer:

- having a sufficient sensitivity giving a stable background in control wells;
- having no gradient over the plate due to long reading times; and
- having no light contamination in adjacent wells from strongly active wells.

Prior to testing it is recommended to ensure having appropriate luminescence measurements, by testing a control plate set-up as described below (triplicate analysis).

Plate setup of first training experiment

	1	2	3	4	5	6	7	8	9	10	11	12
A	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
B	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
C	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
D	**EGDMA 0.98**	**EGDMA 1.95**	**EGDMA 3.9**	**EGDMA 7.8**	**EGDMA 15.6**	**EGDMA 31.25**	**EGDMA 62.5**	**EGDMA 125**	**EGDMA 250**	**EGDMA 500**	**EGDMA 1000**	**EGDMA 2000**
E	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
F	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
G	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO
H	DMSO	DMSO	DMSO	DMSO	DMSO	DMSO	**CA 4**	**CA 8**	**CA 16**	**CA 32**	**CA 64**	**Blank**

EGDMA = Ethylene glycol dimethacrylate (CAS: 97-90-5) a strongly inducing compound
CA = Cinnamic aldehyde, positive reference (CAS: 104-55-2)

The quality control analysis should demonstrate:

- a clear dose response in row D, with the I_{max} > 20-fold above background (in most cases I_{max} values between 100 and 300 are reached);
- no dose-response in row C and E (no induction value above 1.5 (ideally not above 1.3) due to possible light contamination especially next to strongly active wells in the EGDMA row;
- no statistically significant difference between the rows A, B, C, E, F and G. (i.e. no gradient over plate); and
- variability in any of the rows A, B, C, E, F and G and in the DMSO wells in row H should be below 20% (i.e., stable background).

OECD/OCDE

ANNEX 5 - PERFORMANCE STANDARDS

For Assessment of Proposed Similar or Modified *in vitro* Skin Sensitization Keratinocyte-based ARE-Nrf2 Luciferase Reporter Gene Test Methods

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

INTRODUCTION

1. The purpose of Performance Standards (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 (24). 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 (24).

2. Similar (me-too) or modified test methods proposed for use under this Test Guideline should be evaluated to determine their reliability and relevance using Reference Chemicals (Table 1) representing the full range of *in vivo* skin sensitization 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^{TM} as described in paragraphs 9 to 13 of this Annex. The reliability of the new or modified test method, as well as its ability to correctly identify skin sensitizer chemicals should be determined prior to its use for testing chemicals.

3. 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

4. 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 (24). The essential test method components in the present PS comprise the elements described below.

5. Transgenic human keratinocyte cell lines 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. Furthermore, it should be demonstrated that the used cell line has a stable insertion of the luciferase reporter gene. Clones should be selected based on the best signal to noise ratio of the absolute light output of luciferase induction, and on the highest dynamic range of luciferase induction when cells are treated with weak sensitizers. Furthermore 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 sensitizers, including weak sensitizers.

6. 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 CHEMICALS

7. Reference Chemicals 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 (14). The recommended Reference Chemicals listed in Table 1 include substances representing the full range of *in vivo* skin sensitization 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-sensitizers and skin sensitizers 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 (9)(12)(13)(14).

8. For evaluating the between-laboratory variability of similar or modified test method only 15 of the 30 substances listed in Table 1 (in bold italics) need to be evaluated (10 skin sensitizers having various potencies and 5 non-sensitizers). On the other hand the 30 Reference Chemicals listed in Table 1 represent the minimum number of chemicals that should be used to evaluate the relevance of a proposed similar or modified test method proposed to discriminate the various skin sensitization potency (weak, moderate, strong and extreme). . The use of these Reference Chemicals for the development/optimization 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 Chemicals to further evaluate the relevance of the proposed test method.

Table 1: List of Reference Chemicals for determination of reproducibility (n=15 chemicals, in bold italics) and relevance (n=30 chemicals) of similar or modified *in vitro* skin sensitization keratinocyte-based ARE-Nrf2 luciferase reporter gene test methods

Chemical	CASRN	Physical State	*In Vivo* Categories***	VRM *In Vitro* Cat.
NON-SENSITIZER CHEMICALS based on *in vivo* results				
Glycerol*	***56-81-5***	***Liquid***	***Non-sensitizer***	***Non-sensitizer***
Isopropanol*	***67-63-0***	***Liquid***	***Non-sensitizer***	***Non-sensitizer***
Lactic acid*	***50-21-5***	***Liquid***	***Non-sensitizer***	***Non-sensitizer***
Salicylic acid*	***69-72-7***	***Solid***	***Non-sensitizer***	***Non-sensitizer***
Sulfanilamide*	***63-74-1***	***Solid***	***Non-sensitizer***	***Non-sensitizer***
4-Hydroxybenzoic acid	99-96-7	Solid	Non-sensitizer	Non-sensitizer
Chlorobenzene	108-90-7	Liquid	Non-sensitizer	Non-sensitizer
4-methoxy-acetophenone	100-06-1	Solid	Non-sensitizer	Skin sensitizer
Propylene glycol	57-55-6	Liquid	Non-sensitizer	Non-sensitizer
Tween 80	9005-65-6	Liquid	Non-sensitizer	Skin sensitizer
SKIN SENSITIZER CHEMICALS based on *in vivo* results				
Eugenol*	***97-53-0***	***Solid***	***Skin sensitizer (weak)***	***Non-sensitizer********
Ethylene glycol dimethacrylate*	***97-90-5***	***Liquid***	***Skin sensitizer (weak)***	***Skin sensitizer***
Phenyl benzoate*	***93-99-2***	***Solid***	***Skin sensitizer (weak)***	***Non-sensitizer***
Butyl glycidyl ether	2426-08-6	Liquid	Skin sensitizer (weak)	Skin sensitizer
Cinnamyl alcohol	104-54-1	Solid	Skin sensitizer (weak)	Skin sensitizer
Hydroxycitronellal	107-75-5	Solid	Skin sensitizer (weak)	Skin sensitizer

OECD/OCDE

Chemical	CASRN	Physical State	*In Vivo* Categories***	VRM *In Vitro* Cat.
2-Mercaptobenzothiazole	***149-30-4***	***Solid***	***Skin sensitizer (moderate)***	***Skin sensitizer***
Citral	***5392-40-5***	***Solid***	***Skin sensitizer (moderate)***	***Skin sensitizer***
Hexyl cinnamic aldehyde	***101-86-0***	***Solid***	***Skin sensitizer (moderate)***	***Skin sensitizer****
Resorcinol	108-46-3	Solid	Skin sensitizer (moderate)	Non-sensitizer
Tetramethylthiuram disulfide	137-26-8	Solid	Skin sensitizer (moderate)	Skin sensitizer
Thioglycerol	96-27-5	Liquid	Skin sensitizer (moderate)	Non-sensitizer
para-phenylenediamine	***106-50-3***	***Solid***	***Skin sensitizer (strong)***	***Skin sensitizer***
Methyldibromo glutaronitrile	***35691-65-7***	***Solid***	***Skin sensitizer (strong)***	***Skin sensitizer***
Glutaraldehyde	111-30-8	Liquid	Skin sensitizer (strong)	Skin sensitizer
Phthalic anhydride	85-44-9	Solid	Skin sensitizer (strong)	Non-sensitizer
2,4-Dinitrochlorobenzene	***97-00-7***	***Solid***	***Skin sensitizer (extreme)***	***Skin sensitizer***
4-Nitrobenzylbromide	***100-11-8***	***Solid***	***Skin sensitizer (extreme)***	***Skin sensitizer***
Benzoquinone	106-51-4	Solid	Skin sensitizer (extreme)	Skin sensitizer
Oxazolone	15646-46-5	Solid	Skin sensitizer (extreme)	Skin sensitizer

* The substances, sorted first by non-sensitizers followed by skin sensitizers and then by skin sensitization potency were selected from the test chemicals used in validation and in-house studies of KeratinoSens^{TM} which were peer-reviewed by EURL ECVAM (9)(12)(13)(14). The selection includes, to the extent possible, substances that: (i) are representative of the range of skin sensitization potency (e.g. non-sensitizers, weak, moderate, strong and extreme skin sensitizers) tested in the VRM; (ii) reflect the performance characteristics of the VRM; (iii) have chemical structures that are well-defined; (iv) include a variety of mechanisms of action (including pro-haptens, oxidizing chemicals, adduct forming chemicals, Michael acceptors, Schiff base formation, acyl transfer chemicals, aryl electrophile, cross linking, electrophile-H-polar chemical) (26)(27)(28); (v) include a variety of chemical categories based on their organic functional groups; (vi) induce definitive results in the *in vivo* reference test method; (vii) are commercially available; and (viii) are not associated with prohibitive disposal costs.

** In ***bold italics*** are those chemicals suggested to assess within- and between-laboratory reproducibility of the new similar or modified test method.

*** 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).

**** Chemicals which were not 100% concordant between repetitions and laboratories.

Abbreviations: CASRN = Chemical Abstracts Service Registry Number.

OECD/OCDE

DEFINED RELIABILITY AND ACCURACY VALUES

9. For purposes of establishing the reliability and relevance of proposed similar or modified test methods falling within the present Test Guidelines, the Reference Chemicals listed in Table 1 (n=15 and 30 respectively) should be tested. It is however essential that all PS-based validation studies are independently assessed by internationally recognized validation bodies, in agreement with international guidelines (24). The Reference Chemicals should be tested in three independent experiments at sufficiently spaced time points. Each experiment should consist of three concurrently tested replicates for each test chemical, negative (solvent) control and positive control.

10. The calculation of the reliability, 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:

A. Within-laboratory reproducibility (WLR) should be calculated based on concordance of classifications using at least two qualified experiments obtained with the Reference Chemicals.

B. For the calculation of between-laboratory reproducibility (BLR), 15 recommended Reference Chemicals (in bold italics in Table 1) should be tested in at least three laboratories. The final classification for each Reference Chemical in each participating laboratory should be obtained by using the arithmetic mean values over the different qualified experiments performed. BLR should be calculated based on concordance of classifications using only qualified experiments obtained with the Reference Chemicals for which at least one qualified experiment per laboratory is available.

C. The calculation of the accuracy values should be done using all qualified experiments obtained for the 30 Reference Chemical in one laboratory. The calculations should be based on the individual predictions of qualified experiments for each Reference Chemical in the testing laboratory.

In this context, a qualified experiment consists of an experiment that meets the acceptance criteria for the negative and positive control, as defined in the SOP and paragraphs 35 to 37 of the present Test Guideline. Otherwise, the experiment is considered as non-qualified.

Within-laboratory reproducibility

11. An assessment of within-laboratory reproducibility should show a concordance of classifications (sensitizer versus non-sensitizer) obtained in different, independent runs of the 15 recommended Reference Chemicals (shown in bold italics in table 1) within one single laboratory equal or higher (≥) than 90% (actual for KeratinoSens™: 93% based on the 15 Reference Chemicals, and 85% based on the validation dataset)

Between-laboratory reproducibility

12. For similar or modified test methods, the concordance of classifications (sensitizer versus non-sensitizer) between a minimum of three laboratories, obtained for the 15 recommended

Reference Chemicals (shown in bold italics in table 1), should be equal or higher (≥) than 80% (actual for KeratinoSens™: 87% based on the 15 Reference Chemicals and 86% based on the validation dataset).

Predictive capacity

13. 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 obtained with the 30 Reference Chemicals listed in Table 1 should be equal or higher (≥) than 75% (actual for KeratinoSens™: 77% based on the 30 Reference Chemicals and 75% based on the validation dataset), with a sensitivity and specificity also higher (≥) than 75% (actual for KeratinoSens™: 75% sensitivity and 80% specificity based on the 30 Reference Chemicals and 75% for both sensitivity and specificity based on the validation dataset).