EUROPEAN COMMISSION
JOINT RESEARCH CENTRE
Directorate F – Health, Consumers and Reference Materials
Chemical Safety and Alternative Methods Unit (F.3)
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM Recommendation on the use of non-animal approaches for skin sensitisation assessment

Table of Contents

Executive Summary

Considerable progress has been made in non-animal methods and approaches for skin sensitisation assessment since the publication of the EURL ECVAM strategy document for this endpoint in 2013. At the OECD level, three EURL ECVAM validated and/or peer reviewed non-animal methods addressing Key Events (KE) of the skin sensitisation Adverse Outcome Pathway (AOP) have been adopted (i.e. Direct Peptide Reactivity Assay, OECD TG 442C; KeratinoSens™ OECD TG 442D and human Cell Line Activation Test, OECD TG 442E) and other *in vitro* methods are under consideration. The EURL ECVAM Scientific Advisory Committee (ESAC), following independent peer-review, recently released its opinion on two additional *in vitro* (cell-based) methods, the LuSens and the U-SENS™. Various approaches for integrating data within Defined Approaches (DAs), in particular data generated with the regulatory adopted methods, have been proposed as valid components of Integrated Approaches to Testing and Assessment (IATA) and documented in OECD Guidance Document (GD) 256 using harmonised templates for their reporting (provided in OECD GD 255). This progress was the driver of the revision of the REACH information requirements for the skin sensitisation endpoint, with *in chemico* and *in vitro* methods becoming the default route, and of the update of the REACH Guidance on Information Requirements and Chemical Safety Assessment for Skin and Respiratory Sensitisation published by ECHA.

This document provides EURL ECVAM views on the regulatory use of non-animal approaches for skin sensitisation and delivers the following key EURL ECVAM recommendations:

- The qualitative and quantitative mechanistic information generated by *in vitro* methods adopted by the OECD as Test Guidelines should be used, together with other relevant information, within DAs and IATA for assessing skin sensitisation hazard and for hazard classification purposes;

- Predictions generated using the DAs reported in Annex I to OECD GD 256 should be used where possible instead of LLNA data or in conjunction with such data if they already exist, in the context of IATA for assessing skin sensitisation hazard and for hazard classification purposes;

- DAs used for regulatory purposes should be properly documented using the templates provided in OECD GD 255;

- Future work should focus on the definition of internationally agreed standards (e.g. OECD TGs) for DAs and individual test methods that provide equivalent or better level of information than the current animal tests for skin sensitisation;

- The LuSens and U-SENS™ should be used as valid scientific methods for generating information respectively on KE2 and KE3 of the skin sensitisation AOP to be considered together with other relevant information in the context of DAs and IATA. Inclusion of the LuSens in OECD TG 442D and development of an OECD TG on the U-SENS™ is fully supported.

Introduction

Skin sensitisation is the regulatory endpoint aiming at the identification of chemicals able to elicit an allergic response in susceptible individuals. Following repeated exposure to a sensitising agent, the adverse health effect of allergic contact dermatitis (ACD) may be provoked. Thus the development of ACD is characterised by two distinct phases: a) the induction of specialised immunological memory following the initial exposure to an allergen, called sensitisation and b) elicitation of the clinical allergic response following subsequent exposure to the allergen. Skin sensitisation assessment is an important component of the safety evaluation of chemicals.

In 2013 EURL ECVAM undertook an analysis of the standard regulatory requirements for skin sensitisation within pieces of EU chemicals legislation. This analysis, reported in the EURL ECVAM skin sensitisation strategy document (EURL ECVAM, 2013a), clearly indicated that the availability of non-animal approaches capable of identifying skin sensitisation hazard and generating information that would satisfy chemicals' classification needs (i.e. potency sub-categorisation) would have the biggest impact in terms of animal saving in the area.

The EURL ECVAM strategy document also outlined the actions EURL ECVAM planned to undertake in the short (2013-2014), medium (2014-2015) and long term in order to advance progress in the area: a) to finalise the validation and peer review of non-animal test methods for skin sensitisation and lead activities for their regulatory adoption b) to develop non-animal testing strategies suitable for hazard identification and potency sub-categorisation of sensitising chemicals and c) to take a leading role at the OECD in the development of Test Guidelines (TG) and Guidance Documents (GD) that would facilitate a globally harmonised approach to skin sensitisation assessment.

Despite the fact that in the last decade regulations in the cosmetics and chemicals sectors have provided a strong impetus to assess potential toxic effects of chemicals with non-animal methods, at the time of writing of the EURL ECVAM skin sensitisation strategy the assessment of the skin sensitisation potential of chemicals still relied on the use of animal tests (i.e. mainly the Local Lymph Node Assay or LLNA) since no regulatory adopted non-animal methods were available for the purpose. Nevertheless, *in chemico* and *in vitro* (cell-based) methods, addressing mechanisms described under the first three Key Events (KEs) of the skin sensitisation Adverse Outcome Pathway (AOP) initiated by covalent binding to proteins (OECD 2012), were under development by industry and academia.

Three of these methods, namely, the Direct Peptide Reactivity Assay (DPRA), KeratinoSens™ and the human Cell Line Activation Test (h-CLAT) were being formally evaluated by EURL ECVAM through validation and/or independent peer review by the EURL ECVAM Scientific Advisory Committee (ESAC). Subsequent to the ESAC peer review and the publication of the EURL ECVAM Recommendation on the three methods (EURL ECVAM 2013b; 2014; 2015), EURL ECVAM took a leading role on behalf of the EU at the OECD in the development of the corresponding TGs. In 2015 the OECD adopted the DPRA and the KeratinoSens™ as TGs 442C and 442D respectively (OECD 2015a; 2015b) and in 2016 the h-CLAT as TG 442E (OECD 2016a).

In addition to the adopted test methods, knowledge of the skin sensitisation pathway has prompted the development of a wide range of other alternative methods (*in silico*, *in chemico*, *in vitro*), addressing specific KEs of the AOP (OECD 2012). Some of these non-animal tests: the SENS-IS (Cottez et al., 2016) addressing KE2, the U-SENS™ (Piroird et. al. 2015; Alépée et al., 2015), the IL-8 Luc Assay (Kimura et al., 2015) and the GARD (Johansson et al., 2013; 2014), the latter three all addressing mechanisms under KE3 of the skin sensitisation AOP, have been included in the OECD TG programme.

The U-SENS™ and the LuSens, a similar method to the KeratinoSens™, both of which underwent an industry-led validation study, have recently been peer-reviewed by the ESAC (EURL ECVAM Scientific Advisory Committee 2016a; 2016b).

Information generated by these methods can contribute to informing regulatory skin sensitisation assessment and hazard categorisation (e.g. according to the United Nations Globally Harmonised System for Classification and Labelling of Chemicals, GHS, Category 1, 1A and 1B) when used in combination with other relevant evidence, i.e. in the context of Integrated Approaches to Testing and Assessment (IATA) and Defined Approaches (DAs) to testing and assessment (OECD 2016b).

Such approaches are based on the integrated use of information from various sources including *in silico*, *in chemico* and *in vitro* methods. Whilst IATA necessarily include a degree of expert judgment, for example in the choice of the information sources and in their weighting, DAs are standardised both in relation to the set of information sources used and in the procedure applied to the data to derive predictions intended to be used within IATA (OECD, 2016b).

Various DAs for skin sensitisation which integrate data from *in silico*, *in chemico* and *in vitro* methods have been proposed in the past few years. Generally, they are designed to enable the use of mechanistic data from these methods together with other relevant information to predict responses in the LLNA.

In 2013, mindful of the several DAs for skin sensitisation under development, EURL ECVAM took a leading role for the EU at the OECD in the definition of guidance on the reporting of such approaches and proposed to the OECD Task Force on Hazard Assessment (TFHA), now the Working Party on Hazard Assessment (WPHA), to set up an expert group charged with the development of such guidance with the aim to bring to regulatory attention as many approaches as possible.

Discussions within the OECD expert group led to the conclusion that at that point in time, merging or harmonisation of such approaches into a single solution that would satisfy all possible regulatory needs was impossible. Instead, the diverse structures and interpretation processes underlying the different DAs was acknowledged as providing flexibility in the choice of the most appropriate approach to satisfy a specific need. It was also decided that priority should be given to the development of guidance to ensure a harmonised reporting of DAs since when used as components within IATA this would ultimately facilitate the application and evaluation of the IATA themselves for regulatory purposes.

In June 2016 the OECD TFHA endorsed two GD on the reporting of DAs. The OECD GD 255 (OECD, 2016b) provides a set of principles for the reporting of DAs and provides reporting templates to enable their structured documentation. Beside other elements, emphasis is put in the templates on the proper reporting of the limitations in the application of the DA, on their predictive performance and on the sources of uncertainty that may impact on the final prediction. The OECD GD 256 (OECD, 2016c; OECD 2016d; OECD 2016e) exemplifies how the reporting templates have been used to document a number of DAs developed in the area of skin sensitisation. It is nevertheless envisaged that these templates be applied to document DAs developed in other areas of toxicology

The international regulatory adoption of the first three non-animal test methods for skin sensitisation and proposals on how to use these methods in combination within DAs, paved the way in the EU to a substantial revision of the information requirements for skin sensitisation as laid down in Annex VII of the REACH regulation (EC, 2006; EU, 2016). In *chemico* and *in vitro* test methods have become the default requirement and the *in vivo* methods can now only be used if the non-animal tests are shown

not to be suitable for testing a specific substance or cannot be used for classification and risk assessment (EU, 2016). The revised provisions also foresee consideration of whether a substance can be presumed to have the potential to induce significant sensitisation in humans (i.e. GHS Category 1A). The amended requirements, which entered into force in October 2016, will therefore have a substantial impact in replacing and reducing animal testing in view of the 2018 REACH registration deadline.

Advances in the area have also prompted the revision of the European Chemicals Agency's (ECHA) guidance to industry on Information Requirements and Chemical Safety Assessment (Chapter R 7.a, section R.7.3 Skin sensitisation; ECHA, 2016). The revised ECHA draft guidance describes the scope and limitations of the adopted alternative methods to help registrant using them to fulfil the information requirements under REACH. In addition, it proposes a testing and assessment strategy for skin sensitisation assessment which also illustrates how information on the first three key events of the skin sensitisation AOP (i.e. information generated with the validated alternative methods) can be considered in a weight-of-evidence (WoE) approach. Although the guidance recommends the testing and assessment strategy to be followed, it acknowledges that other approaches may be more appropriate depending on the specific case.

Use of non-animal approaches

The in *chemico* and *in vitro* test methods recently adopted by the OECD address the first three KEs of the skin sensitisation AOP (OECD 2012).The DPRA (TG 442C; OECD 2015a) provides information on a chemical's reactivity towards peptides, considered to be the molecular initiating event (MIE) or KE1 within the skin sensitisation AOP. The KeratinoSens™ (TG 442D; OECD 2015b) detects activation of a relevant pathway (the Keap1/antioxidant/electrophile response element (ARE)-dependent pathway) in human-derived keratinocytes providing information on KE2 of the AOP and h-CLAT (TG 442E; OECD 2016a) addresses KE3 of the AOP by measuring, in a human monocytic leukemia cell line, the up-regulation of markers of DC activation following exposure to sensitising agents.

Based on the data generated during the validation studies and historical evidence, the three OECD-adopted test methods have been shown to be transferable to laboratories that have sufficient experience in the techniques involved and to be reproducible for positive/negative predictions in the order of 80-85%. In addition, they demonstrated considerable accuracy, of about 80%, in predicting LLNA responses despite the fact that they are not meant to be used as stand-alone replacement methods. Moreover, they have been shown in recent analyses to have the ability to correctly detect as positives the majority of chemicals that need to be air oxidised (pre-haptens) or enzymatically transformed (pro-haptens) to act as sensitisers (Casati et al., 2016; Urbisch et al., 2016; Patlewicz et al., 2016). Besides the qualitative information (positive/negative predictions), DPRA, KeratinoSens™ and h-CLAT also provide quantitative readouts that can inform hazard classification (i.e. potency categorisation).

EURL ECVAM recommends using the qualitative and quantitative mechanistic information generated by the OECD-adopted methods, together with other relevant information, in the context of DAs and IATA for assessing skin sensitisation hazard and for hazard classification purposes.

In the past few years advancements have also been made in the integration of data from different non-animal tests in the context of DAs to improve accuracy in predictions with respect to the individual methods. Twelve of these DAs are documented in Annex I to OECD GD 256 (OECD, 2016d). These DAs are based on the use of information sources addressing key mechanisms/events of the skin sensitisation AOP and make use of a variety of specific methodologies, i.e. Data Interpretation Procedures (DIP), for converting the input data into a final prediction. These DAs provide a good overview of the different set of information sources and DIP that can be used for skin sensitisation hazard assessment and/or classification. The DIP can range from very simple rule-based sequential decision steps to mathematical and statistical approaches. Besides those reported in Annex I to OECD GD 256 (OECD, 2016d), other DAs have been documented in the scientific literature (e.g. Luechtefeld et al., 2016; Macmillan et al., 2016; Strickland et al., 2017; Zang et al., 2017) and recently reviewed by Ezendam et al. (2016).

Table 1 provides an overview of the DAs in relation to their proposed use, AOP coverage, type of information sources used within, number of chemicals tested and predictive performances. Note that the information provided is meant to give a flavour of each DA and does not allow a comprehensive understanding of the DA structure and actual performance. The reader should refer to OECD GD 256 and its two annexes (OECD 2016c; 2016d; 2016e) for a detailed description of the different DAs. In addition, it is beyond the scope of this document to offer any detailed comparison of the different DAs especially in relation to their predictive performance knowing that performance statistics are very much dependent on the dataset used which differs among the reported DAs.
Nevertheless, from the information summarised in Table1 some general observations can be made:

1) All DAs make use of mechanistic data addressing one or more key events (KE) of the skin sensitisation AOP (OECD 2012). Despite non-animal experimental data on KE4 (T cell priming and proliferation) are not used in any of the DAs due to the lack of standardised alternative methods addressing this KE, the different DAs already show a high level of accuracy in predicting binary LLNA classifications;

2) Besides that derived from validated and regulatory adopted methods, other relevant information such as physicochemical properties and *in silico* predictions, contribute to skin sensitisation hazard assessment and classification;

3) Information on KE1 (i.e. the MIE in the skin sensitisation AOP) is used in all DAs (either derived with *in silico* models and/or with *in chemico, in vitro* methods) and in some cases reactivity information has proven by the underlying analyses to have the highest power in discriminating between sensitising and non-sensitising chemicals (e.g. Natsch et al., 2015; Asturiol et al., 2016).

4) All of the DAs make use of cell-based assays and/or *in silico* descriptors that account for skin metabolism and autoxidation processes. In fact, as detailed in Annex I to OECD GD 256 (OECD, 2016d), the number of pre- and pro-haptens classified as being non-sensitisers is generally limited.

5) The majority of the DAs have been developed/tested with a substantial number of chemicals (in certain cases more than 200) for which *in vivo* skin sensitisation data are available indicating that their domain of applicability should cover the main reaction mechanisms relevant to skin sensitisation ,

6) The accuracy of the different DAs specifically designed to predict binary LLNA classifications (i.e. sensitiser/non-sensitiser) is high and in the range of 79-93% (with sensitivity in the range of 79-98% and specificity in the range of 72-94%).

7) Where an evaluation of the predictive capacity against human data was performed, this shows that the DAs tend to predict human responses more accurately than the animal model (LLNA) does (e.g. Urbisch et al., 2015; Asturiol et al., 2016; Strickland at al., 2017; Zand et al., 2017).

8) When considering the accuracy of the different DAs designed for potency assessment, an important aspect to consider is the variability of the reference animal test. Consistent with what was already known about the variability of the LLNA (e.g. ICCVAM 2011), recent analyses have confirmed that this is far from being negligible (Hoffmann, 2015; Dumont et al., 2016) making it difficult to assign a chemical to a specific potency class with sufficient confidence on the basis of a single LLNA study result.

In light of the above, it is evident that some of the DAs developed in the area of skin sensitisation have comparable performance to the LLNA for the identification of skin sensitisation hazard. Moreover, they appear to be more accurate than the LLNA in predicting hazard responses in humans. Although it is recognised that further work is needed to achieve a more detailed definition of the relative potency of identified skin sensitising chemicals for risk assessment purposes, the DAs summarised in Table 1 already provide useful information for the purpose of classification and labelling.

EURL ECVAM recommends that the predictions generated using the DAs reported in Annex I to OECD GD 256 (and summarised in Table 1) be used, where possible, instead of LLNA data or in conjunction with such data if they already exist, in the context of IATA for assessing skin sensitisation hazard and for hazard classification purposes.

EURL ECVAM recommends that DAs used for regulatory purposes be properly documented using the templates provided in OECD GD 255.

Future developments

Table 1 clearly shows that some of the DAs have comparable or even better performance than the LLNA for skin sensitisation hazard assessment and classification. At international level, progress was made to guarantee harmonised reporting of DAs in view of facilitating their regulatory application. Nevertheless, harmonised reporting is not sufficient to guarantee their effective implementation and acceptance of the DAs' predictions by different jurisdictions and regions.

EURL ECVAM recommends that future work should focus on the definition of internationally agreed standards (e.g. OECD TGs) for DAs and individual test methods that provide equivalent or better level of information than the current animal tests for skin sensitisation.

To this end, work should be undertaken to develop assessment criteria to objectively and systematically evaluate the DAs reported in Annex I to OECD GD 256 as well as other candidate DAs and upcoming individual test methods. Such assessment criteria should be informed by a comprehensive understanding of the performance of the LLNA in terms of reproducibility and relevance in predicting human responses.

As part of the previous OECD activities on the documentation of DAs (OECD 2016b, 2016c), emphasis has been given to systematically report the possible sources of uncertainty associated with the application of a specific DA. For example, uncertainties can be associated with the structure of the DA itself, the information sources used within (e.g. variability of the input data) and the *in vivo* (animal and/or human) benchmark data used to assess the performance of the DA. In fact the predictive performance of the DAs listed in Table 1, including those proposed for potency categorisation, has been evaluated using as benchmark data individual LLNA predictions and potency estimates (i.e. EC3 values). Thus the calibration of the DIP associated with each DA did not take into account neither the variability of the animal test nor the variability associated with the model input parameters. The impact of the combined effect of these sources of uncertainties on the final DA prediction should be further characterised as part of future activities on the evaluation of DAs.

The LuSens test method

The LuSens is an *in vitro* test method proposed to contribute to the assessment of the skin sensitisation potential of chemicals when used in conjunction with other information (i.e. in the context of DAs and IATA).

The method quantifies luciferase gene induction as a measure of the activation of the Keap1/antioxidant/electrophile response element (ARE)-dependent pathway in a keratinocyte cell line stably transfected with a selectable plasmid. The LuSens is addressing the mechanism of induction of cyto-protective pathways in keratinocytes, covered by KE2 in the skin sensitisation AOP (OECD, 2012). The LuSens method is considered similar to the KeratinoSens™ for which an OECD TG is available (OECD TG 442D) (OECD, 2015b) supplemented by "Performance standards (PS) for assessment of proposed similar or modified *in vitro* skin sensitisation ARENrf2 luciferase test methods" (OECD, 2015c).

The LuSens underwent an industry-led PS-based validation study conducted to fulfil the requirements laid down in the OECD PS for demonstrating comparable performance to that of the validated KeratinoSens™ and adherence to the essential test methods components that would assure similarity with the validated reference method.

The validation study demonstrated that the method is easily transferable to laboratories experienced in cell culture techniques. The within-laboratory (WLR) reproducibility and the between-laboratory reproducibility (BLR), as calculated on the basis of concordant classifications for the chemicals tested (n=12 for WLR and n=20 for BLR) was 100%. The LuSens also complied with the performance standard requirements for accuracy ≥ 80% (LuSens 85%) and sensitivity ≥ 80% (LuSens 92%) but not with the one for specificity ≥ 80% (LuSens 75%). Potential differences between the LuSens and the KeratinoSens™ assays were only observed for methyl salicylate and eugenol, which resulted in apparently higher sensitivity and lower specificity.

The ESAC Opinion on the method, delivered in June 2016 (EURL ECVAM Scientific Advisory Committee, 2016a), highlighted the fact that the two substances giving different results in the two methods are borderline substances (i.e. give both positive and negative predictions in repeated runs in several methods, including LuSens, KeratinoSens™, DPRA, h-CLAT and also in the LLNA and humans).

EURL ECVAM recommends the use of the LuSens as a valid scientific method for generating information on KE2 of the skin sensitisation AOP to be used together with other relevant information in the context of DAs and IATA and fully supports the inclusion of the method into OECD TG 442D.

The U-SENS™ test method

The U-SENS™ is an *in vitro* test method proposed to contribute to the assessment of the skin sensitisation potential of chemicals when used in conjunction with other information (i.e. in the context of DAs and IATA).

The test method is based on the quantitative cytofluorometric analysis of the induction of the CD86 protein marker in U937 cells (a cell line established from a diffuse histiocytic lymphoma) after 45h exposure to the test chemical. The test method is proposed to address KE3 (dendritic cell activation) of the AOP (OECD, 2012). The U937 cells are human myeloid cells used as a surrogate model for DC. Activated upon the contact with skin sensitizers, they increase the CD86 expression. The induction of the CD86 membrane protein following exposure to skin sensitisers is one of the biomarkers indicating activation of DC that is most frequently used in *in vitro* assays.

The U-SENS™ underwent an industry-led validation study designed primarily to address the reproducibility of the method (WLR and BLR). The WLR assessed in four laboratories on the basis of concordant classifications for 15 chemicals was 73%, 93%, 100% and 100% respectively (average 91.7%) with the lowest reproducibility was observed in the laboratory less familiar with the use of the U-SENS™ method indicating that the method may require expertise and time for proper implementation. The between-laboratory reproducibility was in the order of 84% (n=38).

The accuracy of the method in discriminating between sensitisers and non sensitisers on the basis of LLNA classifications was calculated to be 93% (sensitivity 97%, specificity 89%) with the chemical tested in the validation study (n=38). In addition, performance values for a larger set of 166 substances tested in house were provided indicating an accuracy of 85% (sensitivity 95% and specificity 65%) for this larger set when evaluated against LLNA data. When evaluated against human data the U-SENS™ showed an accuracy of 83% (sensitivity 95% and specificity 59%).

The limitations of the USENS™ are likely to be very similar to other submerged cell culture assays (e.g. h-CLAT, KeratinoSens). Potential issues may be encountered with substances of low solubility or low stability in an aqueous environment, fluorescent substances interfering with flow cytometry analysis, volatile substances and substances disrupting cell membranes. The USENS™ does not seem to have any major difficulty with the detection of pre- and pro-haptens.

The ESAC Opinion on the method, delivered in June 2016 (EURL ECVAM Scientific Advisory Committee, 2016b), indicated that the application of six rules to the prediction model to resolve inconclusive results increases the complexity of the method without adding to its predictive performance since in most cases the six rules appears to convert inconclusive results into positive results. This suggestion has been taken into account by the test developer through a revision of the protocol and a supporting analysis of the validation study data and historical data showing that the elimination of the six rules does not impact on the test method's performance.

EURL ECVAM recommends the use of the U-SENS™ as a valid scientific method for generating information on KE3 of the skin sensitisation AOP to be used together with other relevant information in the context of DAs and IATA and fully supports the development of an OECD TG on the method.

References

Alépée N, Piroird C, Aujoulat M, Dreyfuss S, Hoffmann S, Hohenstein A, Meloni M, Nardelli L, Gerbeix C, Cotovio J (2015). Prospective multicentre study of the U-SENS test method for skin sensitization testing. Toxicology In Vitro, 30:373-382.

Asturiol, D., Casati, S., Worth, A (2016). Consensus classification tree model for skin sensitisation hazard prediction. Toxicology in Vitro, 36:197-209.

Casati, S., Aschberger, K., Asturiol, D., Basketter, D., Dimitrov, S., Dumont, C., Karlberg, A-T., Lepoittevin, J-P., Patlewicz, G., Roberts, D.W., Worth, A (2016). Ability of non-animal methods for skin sensitisation to detect pre- and pro-haptens: Report and recommendations of an EURL ECVAM expert meeting; EUR 27752 EN; doi:10.2788/01803

Cottrez F, Boitel E, Ourlin JC, Peiffer JL, Fabre I, Henaoui IS, Mari B, Vallauri A, Paquet A, Barbry P, Auriault C, Aeby P, Groux H (2016). SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitization potency: Reproducibility and predictivity results from an inter-laboratory study. Toxicology In Vitro, 32:248-60.

Dumont C, Barroso J, Matys I, Worth A, Casati S (2016). Analysis of the local lymph node assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. Toxicology In Vitro 34, 220-228.

EC 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal No L 396, p. 1-849.

EU 2016. Commission Regulation (EU) 2016/1688 of 20 September 2016 amending Annex VII to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards skin sensitisation. Official Journal of the European Union L 255, p.14-16.

ECHA 2016. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7a: Endpoint specific guidance. Version 5.0. December 2016.https://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf/e4a2a18f-a2bd-4a04-ac6d-0ea425b2567f

EURL ECVAM 2013a. Strategy for Replacement of Animal Testing for Skin Sensitisation Hazard Identification and Classification. https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-strategy-papers/strat-skin-sensitisation

EURL ECVAM 2013b. Recommendation on the Direct Peptide Reactivity Assay (DPRA). https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/eurl-ecvam-recommendation-on-the-direct-peptide-reactivity-assay-dpra

EURL ECVAM 2014. Recommendation on the KeratinoSens™ assay for skin sensitisation testing. https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/recommendation-keratinosens-skin-sensitisation

EURL ECVAM 2015. Recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitisation testing. https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/eurl-ecvam-recommendation-on-the-human-cell-line-activation-test-h-clat-for-skin-sensitisation-testing

EURL ECVAM Scientific Advisory Committee (2016a). ESAC Opinion on the BASF-coordinated Performance Standards-based validation of the LuSens test method for skin sensitisation testing. ESAC Opinion No. 2016-04 of 24 June 2016; EUR 28176 EN; doi:10.2787/170247. http://publications.jrc.ec.europa.eu/repository/handle/JRC103706.

EURL ECVAM Scientific Advisory Committee (2016b). ESAC Opinion on the L'Oréal-coordinated study on the transferability and reliability of the U-SENS™ test method for skin sensitisation testing. ESAC Opinion No. 2016-03 of 24 June 2016; EUR 28178 EN; doi:10.2787/815737. http://publications.jrc.ec.europa.eu/repository/handle/JRC103705.

Ezendam J, Braakhuis HM, Vandebriel RJ (2016). State of the art in non-animal approaches for skin sensitization testing: from individual test methods towards testing strategies. Arch Toxicol., 90:2861-2883. Review.

Hoffmann S (2015). LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. ALTEX, 32:379-83.

ICCVAM 2011. ICCVAM Test Method Evaluation Report: Usefulness and Limitations of the Murine Local Lymph Node Assay for Potency Categorisation of Chemicals Causing Allergic Contact Dermatitis in Humans. NIH Publication Number 11-7709. https://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/immunotoxicity/llna-potency/tmer/index.html

Johansson H, Albrekt AS, Borrebaeck CA, Lindstedt M (2013). The GARD assay for assessment of chemical skin sensitizers. Toxicology In Vitro,27:1163-9.

Johansson H, Rydnert F, Kühnl J, Schepky A, Borrebaeck C, Lindstedt M (2014). Genomic allergen rapid detection in-house validation-a proof of concept. Toxicol Sci., 139:362-70.

Kimura Y, Fujimura C, Ito Y, Takahashi T, Nakajima Y, Ohmiya Y, Aiba S (2015). Optimization of the IL-8 Luc assay as an in vitro test for skin sensitization. Toxicology In Vitro, 29:1816-30.

Luechtefeld T, Maertens A, McKim JM, Hartung T, Kleensang A, Sá-Rocha V (2016). Probabilistic hazard assessment for skin sensitization potency by dose-response modeling using feature elimination instead of quantitative structure-activity relationships. J Appl Toxicol., 35:1361-71.

Macmillan DS, Canipa SJ, Chilton ML, Williams RV, Barber CG (2016). Predicting skin sensitisation using a decision tree integrated testing strategy with an in silico model and in chemico/in vitro assays. Regul Toxicol Pharmacol. ,76:30-8.

Natsch A, Emter R, Gfeller H, Haupt T, Ellis G (2015). Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: global versus domain-based assessment. Toxicol Sci., 143:319-32.

OECD 2007. Guidance document on the validation of (quantitative) structure-activity relationships [(q)sar] models. ENV/JM/MONO(2007)2. Series on Testing and Assessment No. 69. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)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. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

OECD 2015a. Guidelines for the Testing of Chemicals Test No. 442C. In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA). http://www.oecdilibrary.org/docserver/download/9744011e.pdf?expires=1462264428&id=id&accname=guest&checksum=A189F220E372F008D7D9D3483AEE3E6D

OECD 2015b. Guidelines for the Testing of Chemicals Test No. 442D. In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. http://www.oecdilibrary.org/docserver/download/9744021e.pdf?expires=1462264529&id=id&accname=guest&checksum=9E031143681D2BBA085579C20EF11305

OECD 2015c. Series on Testing and Assessment: Performance Standards No. 213: Performance Standards for Assessment of Proposed Similar or Modified In Vitro Skin Sensitisation ARE-NRF2 Luciferase Test Methods. http://www.oecd.org/env/ehs/testing/seriesontestingandassessmentperformancestandards.htm

OECD 2016a. OECD Guidelines for the Testing of Chemicals Test No. 442E. In Vitro Skin Sensitisation: human Cell Line Activation Test (h-CLAT). http://www.oecd.org/env/ehs/testing/151216-Draft-h-CLAT-TG-After-Expert-Meeting-%28clean%29-Final.pdf (in publication)

OECD 2016b. Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment. Series on Testing & Assessment No. 255. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)28&doclanguage=en

OECD 2016c. Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used Within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation . Series on Testing & AssessmentNo. 256. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)29&doclanguage=en

OECD 2016d. Annex I: Case Studies to the Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used Within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation. Series on Testing & AssessmentNo. 256. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)29/ann1&doclanguage=en

OECD 2016e. Annex II: Information Sources Used Within the Case Studies to the Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used Within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation. Series on Testing & AssessmentNo.256.

http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)29/ann2&doclanguage=en

Patlewicz G, Casati S, Basketter DA, Asturiol D, Roberts DW, Lepoittevin JP, Worth AP, Aschberger K (2016). Can currently available non-animal methods detect pre and pro-haptens relevant for skin sensitization? Regul Toxicol Pharmacol., 82:147-155.

Piroird C, Ovigne JM, Rousset F, Martinozzi-Teissier S, Gomes C, Cotovio J, Alépée N (2015). The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. Toxicology In Vitro, 29:901-16.

Strickland J, Zang Q, Paris M, Lehmann DM, Allen D, Choksi N, Matheson J, Jacobs A, Casey W, Kleinstreuer N (2017). Multivariate models for prediction of human skin sensitization hazard. J Appl Toxicol.. 37:347-360.

UN (2011). Globally Harmonised System of Classification and Labelling of Chemicals (GHS, Rev.6). Part 3: Health and Environmental Hazards. New York, NY, USA, and Geneva, Switzerland. United Nations Economic Commission for Europe. http://www.unece.org/trans/danger/publi/ghs/ghs_rev06/06files_e.html#c38156

Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul Toxicol Pharmacol., 71:337-51.

Urbisch D, Becker M, Honarvar N, Kolle SN, Mehling A, Teubner W, Wareing B, Landsiedel R (2016). Assessment of Pre- and Pro-haptens Using Nonanimal Test Methods for Skin Sensitization. Chem Res Toxicol., 16;29:901-13.

Zang Q, Paris M, Lehmann DM, Bell S, Kleinstreuer N, Allen D, Matheson J, Jacobs A, Casey W, Strickland J (2017). Prediction of skin sensitization potency using machine learning approaches. J Appl Toxicol. 2017 Jan 10. doi: 10.1002/jat.3424. [Epub ahead of print].

Table 1: Overview of the Defined Approaches documented in Annex I to OECD GD 256

	Defined Approach	Proposed use	AOP KEs addressed[1]	Information sources used[2]	DIP	Number of chemicals tested	Predictive capacity parameters (%) evaluated against LLNA and/or human responses[3]	Comments
1	AOP –based "2 out of 3" weight of evidence / integrated testing strategy ("2 out of 3 – Sens ITS") approach to skin hazard identification (BASF)	Hazard identification	1,2,3	OECD TG 442C (KE1) OECD TG 442D or LuSens (KE2) mMUSST or OECD TG 442E (KE3)	Integrated Testing Strategy (ITS) in which concordant results for two KEs drive the prediction	213 (151 S[4]) (62 NS[4])	Against LLNA (n=126-180): Accuracy 79-84 Sensitivity 79-84 Specificity 72-84 Against human (n=75-101): Accuracy 88-91 Sensitivity 84-90 Specificity 89-100	

[1] The Key Events (KE) reported in the column does not necessary imply that all of them are addressed each time a substance is tested with the DA

[2] The information sources are not necessarily listed in the table in the order they are used within the DAs

[3] Predictive capacity parameters are those documented in Annex I to OECD GD 256 (OECD, 2016d) and, where possible, are presented as ranges when more than one value is reported (e.g. in case these have been calculated using different subsets of data)

[4] S=sensitiser, NS= non-sensitiser on the basis of the reference *in vivo* data

* Values calculated by EURL ECVAM

2	Sequential Testing Strategy (STS) for hazard identification of skin sensitisers (RIVM)	Hazard identification	1,2,3,4	OECD TG 442C (KE1) OECD TG 442D and HaCaT gene signature (KE2) OECD TG 442E (KE3) Bayesian QSAR approach (MultiCASE, CAESAR, DEREK and OECD QSAR Toolbox) (KE4)	Sequential Testing Strategy (TST) in which decision criteria are applied after each tier	41 (27 S) (14 NS)	Against human: Accuracy 95 Sensitivity 96 Specificity 93 LLNA against human Accuracy 78 Sensitivity 93 Specificity 64	Performance of the LLNA in predicting human responses for the same set of chemicals, as reported in van der Veen et al. (2014)
3	A non-testing Pipeline approach for skin sensitisation (G. Patlewicz)	Primarily hazard identification. In certain cases allows sub-categorisation of sensitisers into GHS subcategories 1A and 1B	1,2,3,4,AO	Various physicochemical properties Various *in silico* simulators for abiotic or enzymatic activation Various *in silico* methods (KE1, KE2 and KE4) TG 442C and glutathione depletion assay (KE1) OECD TG 442D (KE2) OECD TG 442E and U-SENS™ (KE3) OECD TG 429 (KE4) OECD TG 406 (AO) Others	Weight-of-evidence	100 (55 S) (45 NS)	Against LLNA, GPMT, Buhler test: Accuracy 88 Sensitivity 89 Specificity 86	This approach represents an IATA workflow rather than a DA. It was used to show that the template provided in OECD GD 255 is flexible enough to be used to document IATA besides DAs.

4	Stacking meta-model for skin sensitisation hazard identification (L'Oréal)	Hazard identification	1,2,3	Various physicochemical properties Various *in silico* methods and OECD TG 442C (KE1) OECD TG 442D (KE2) U-SENS™ (KE3)	Meta-model stacking five different statistical methods (Boosting, Naïve Bayes, Support Vector Machine (SVM), Sparse PLS-DA and Expert Scoring). The model provides a probabilistic output	165 113 training set (66 S) (47 NS) 52 test set (31 S) (21 NS)	Against LLNA for training set: Accuracy 93 Sensitivity 95 Specificity 90 Against LLNA for test set: Accuracy 92 Sensitivity 93 Specificity 90	Overall accuracy values not reported
5	Integrated decision strategy for skin sensitisation hazard (ICCVAM)	Hazard identification	3,4,AO	Various physicochemical properties OECD TG 442E (KE3) OECD Toolbox (KE4, AO)	Support vector machine	120 (87 S) (33 NS)	Against LLNA: Accuracy 88 Sensitivity 85 Specificity 94	Accuracy values for training and test set reported in the Annex I to OECD GD 256
6	Classification consensus model of decision trees based on in silico descriptors to predict skin sensitisation hazard (EC- JRC)	Hazard identification	1	Various in silico descriptors generated with TIMES-SS (KE1) and DRAGON software packages	Consensus model of two decision trees	269 (170 S) (99 NS)	Against LLNA (n=269): Accuracy 93 Sensitivity 98 Specificity 85 Against human (n=99): Accuracy 81 Sensitivity 90 Specificity 64	Accuracy values for training and test set reported Annex I to OECD GD 256

7	Sensitizer potency prediction based on Key event 1 + 2: Combination of kinetic peptide reactivity data and KeratinoSens™ data (Givaudan)	Hazard identification Prediction of a continuous likely LLNA or human potency value Potency sub-categorisation in the two GHS potency classes (1A and 1B)	1,2	Various physicochemical properties TIMES-SS and Cor1C420-assay (KE1) OECD TG 442D (KE2)	Reactivity domain-based regression equations Global regression equation for chemicals that cannot be assigned to a specific domain	244 (133 S) (111 NS)	Against LLNA for binary classification: Accuracy 83* Sensitivity 89* Specificity 75* Against LLNA GHS potency classes (NS, 1B, 1A) Accuracy: Overall for 3 classes 75 NS 76* 1B 77* 1A 70*	The accuracy values reported are calculated combining results from the global and domain models.
						Human GHS class 1B (44 S) Human GHS class 1A (27 S)	a) Predicted by in vitro data: 1B 55 1A 70 b) Predicted by LLNA data: 1B 72 1A 52	Classifications derived by estimating human DSA_{05} values (induction dose per skin area, in µg/cm2, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population using a) in vitro data and b) LLNA data

8	The artificial neural network model for predicting LLNA EC3 (Shiseido)	Hazard identification potency sub-categorisation in 3 potency classes: non-sensitisers (N), combined weak and moderate (M), combined strong (S) and extreme (E) sensitisers	1,2,3	Log P cell-surface thiol test (SH test) (KE1) Antioxidant Response Element (ARE) assay (KE2) OECD TG 442E (KE3)	Artificial neural network	62 (48 S) (14 NS)	Against 3 LLNA potency classes (NS, W/M, S/E) Accuracy: Overall for 3 classes 79 NS 64* W/M 93* S/E 67*	
9	Sensitizer potency prediction based on Key event 1+2+3: Bayesian Network ITS/DS for hazard and potency identification of skin sensitizers (P&G)	LLNA potency probabilistic distribution (pEC3), for 4 potency classes: non-sensitisers (N), weak (W), moderate (M), and combined strong (S) and extreme (E) sensitisers	1,2,3	Various parameters for bioavailability In silico simulators for abiotic or enzymatic activation (TIMES) TIMES-SS and OECD TG 442C (KE1) OECD TG 442D (KE2) OECD TG 442E (KE3)	Bayesian Network, the model provides a prediction with either all or partial data inputs	207 (154 S) (53 NS)	Against LLNA for binary classification: Accuracy 96 Against 4 LLNA potency classes (NS, W, M, S/E): Accuracy: Overall 74-89 NS 87-100 W 83-90 M 45-75 S/E 60-87	The accuracy reported is calculated for the test set and considering data inputs and omission of each one of the KE assays Potency probabilistic distributions are documented in Annex I to OECD GD 256

10	Sequential testing strategy (STS) for sensitising potency classification based on in chemico and in vitro data (Kao Corporation)	Hazard identification LLNA potency sub-categorisation in 3 potency classes: non-sensitisers (N), weak (W) (combined W and M sensitisers in the LLNA), strong (S) (combined S and E sentisers in the LLNA)	1,3	OECD TG 442C (KE1) OECD TG 442E (KE3)	Sequential Testing Strategy (TST) in which decision criteria are applied after each tier	139 (1O2 S) (37 NS)	Against LLNA for binary classification: Accuracy 81 Sensitivity 90 Specificity 54 Against LLNA potency classification: Strong (EC3<1% in the LLNA) and Weak (EC3≥1% in the LLNA): Accuracy Overall 69 NS 54* Weak 78* Strong 66*	
11	Integrated testing strategy (ITS) for sensitising potency classification based on in silico, in chemico, and in vitro data (Kao Corporation)	Hazard identification LLNA potency sub-categorisation in 3 potency classes: non-sensitisers (N), weak (W) (combined W and M sensitisers in the LLNA), strong (S) (combined S and E sentisers in the LLNA)	1,3	DEREK Nexus and OECD TG 442C (KE1) OECD TG 442E (KE3)	Integrated Testing Strategy (ITS) based on the integration of input parameters converted into scores	139 (1O2 S) (37 NS)	Against LLNA for binary classification: Accuracy 84 Sensitivity 89 Specificity 70 Against LLNA potency classification Strong (EC3<1% in the LLNA) and Weak (EC3≥1% in the LLNA): Accuracy: Overall 71 NS 70* Weak 78* Strong 52*	

12	DIP for skin allergy risk assessment (SARA) (Unilever)	Potency prediction expressed as probability that a specific CD8+ T cell response will be induced following a given skin exposure to a direct-acting sensitising chemical	1,3,4	Modified OECD TG 428 to derive information on skin bioavailability kinetics and protein haptenation kinetics (KE1) Prediction of Class I MHC processing & presentation of haptenated skin protein by Dendritic cells (DC) (KE3) Prediction of the extent of human naïve CD8+ T cell activation (KE4)	Ordinary differential equation	1	Not applicable	