European Commission

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

2016

EUR 28178 EN

Contact information
Name: João BARROSO
Address: Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
E-mail: Joao.BARROSO@ec.europa.eu
Tel.: +39 0332 78 5329

JRC Science Hub
https://ec.europa.eu/jrc

JRC103705

EUR 28178 EN

PDF ISBN 978-92-79-63200-6 ISSN 1831-9424 doi:10.2787/815737

Luxembourg: Publications Office of the European Union, 2016

How to cite: EURL ECVAM Scientific Advisory Committee (2016). 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. Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC103705.

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health, Consumers and Reference Materials
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)

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
Relevant ESAC Request No.	2016-03
Date of Opinion	24/06/2016

Table of contents

Abstract

ESAC, the EURL ECVAM Scientific Advisory Committee, advises EURL ECVAM on scientific issues. Its main role is to conduct independent peer review of validation studies of alternative test methods and to assess their scientific validity for a given purpose. The committee reviews the appropriateness of study design and management, the quality of results obtained and the plausibility of the conclusions drawn. ESAC peer reviews are formally initiated with a EURL ECVAM Request for ESAC Advice, which provides the necessary background for the peer-review and establishes its objectives, timelines and the questions to be addressed. The peer review is normally prepared by specialised ESAC Working Groups. These are typically composed of ESAC members and other external experts relevant to the test method under review. These experts may be nominated by ESAC, EURL ECVAM and partner organisations within the International Cooperation on Alternative Test Methods (ICATM). ESAC ultimately decides on the composition of these Working Groups. ESAC's advice to EURL ECVAM is formally provided as 'ESAC Opinions' and 'Working Group Reports' at the end of the peer review. ESAC may also issue Opinions on other scientific issues of relevance to the work and mission of EURL ECVAM but not directly related to a specific alternative test method.

The ESAC Opinion expressed in this report relates to the peer-review of the L'Oréal-coordinated study on the transferability and reliability of the U-SENS™ test method for skin sensitisation testing.

Ispra, 24 June 2016

ESAC Opinion

In April 2016, the EURL ECVAM Scientific Advisory Committee (ESAC) (Annex 1) received from EURL ECVAM a request for scientific advice on the L'Oréal-coordinated validation of the U-SENS™ test method for skin sensitisation testing (Annex 2). ESAC established a working group (WG) (Annex 1) which delivered an ESAC WG report dated 6 June 2016 (Annex 3).

At its 42nd meeting, held on the 9th and 10th June 2016 at EURL ECVAM, Ispra, Italy, the non-Commission members of ESAC unanimously endorsed the following statement which was based on the ESAC WG report:

The replacement of traditional animal-based test methods by alternative ones should ideally be one-to-one replacements. However, the replacement of traditional animal-based test methods for skin sensitisation by animal-free approaches is likely to require an Integrated Approach to Testing and Assessment (IATA). This combines reliable and mechanistically relevant test methods addressing the currently recognised key events in the adverse outcome pathway for skin sensitisation induction (OECD, 2012; OECD, 2016a, b).

Currently, three animal-free test methods have been validated becoming OECD Test Guidelines (TG). The DPRA (TG 442C) provides information about the capacity of the chemical to react with cysteine and/or lysine groups (key event 1). The KeratinoSens™ test method (TG 442D) addresses skin inflammation by assessing the impact of a substance on ARE mediated Nrf2 activation (key event 2). The h-CLAT (TG 442E) addresses key event 3 by measuring the upregulation of cell membrane markers CD86 and CD54 to predict a skin sensitisation potential.

The test method under evaluation (U-SENS™) addresses key event 3 by measuring CD86 upregulation as a surrogate measure for dendritic cell migration/maturation. This test method shows similarities with the h-CLAT test method, which measures the impact of substances on CD86 and CD54 expression levels. A comparison with the similar h-CLAT method in terms of its benefit would be helpful to end users enabling them to make a decision regarding which method would be most appropriate for their purpose.

The study design was reasonable. The within- and between-laboratory reproducibility are comparable, arguably better, than the validated methods listed above. However, a lower reproducibility was observed with one of the naïve laboratories. This may highlight the importance of proper training. Since proper training opportunities were provided by the test method developer, a lower reproducibility may also indicate additional challenges for a naïve laboratory trying to apply the method. The predictive capacity is good as far as the overall binary classification (Y/N) is concerned and the assay seems to have a good sensitivity, perhaps at the expense of specificity. Overall, the test method's accuracy is comparable with accepted *in vitro* methods. Where directly compared, performance against human data and LLNA data was equivalent.

Notwithstanding the above, the prediction model was found to be complex and the documentation was insufficiently clear to enable proper understanding of the conclusions drawn by the test submitter from "inconclusive" results. In particular, the application of the 'six rules' developed to solve inconclusive results was difficult to understand. This may jeopardise the application of this method by the end users. For these reasons, ESAC has the concern that the 'six rules' used to resolve inconclusive outcomes were fitting the

outliers in the absence of a reasonable biological and/or chemical rational. The presented data suggest that the 'six rules' have limited accuracy (sensitivity: 90-100 %; specificity: 33 %) and add no value to simply considering all inconclusive results to be positive by default. Indeed, considering all inconclusive results as sensitiser by default instead of applying the 'six rules' would slightly decrease the number of false negatives with only an equally small increase of false positives.

The explanations given for false positive and false negative results were not always consistent, as they appeared to be laboratory specific. Taking all the data together, a broad range of potency and chemical reactivity classes was covered. Analysis of these data revealed that outliers did not correspond with a particular potency or chemical reactivity class. The data also showed that pre-haptens were correctly identified. Overall, the presented evidence seems to indicate that volatile substances or poorly soluble substances may require additional attention in the technology transfer and training phase, as improper handling of such substances during exposure could affect the reproducibility of the results rather than their accuracy.

ESAC concludes that in terms of test performance, the U-SENS™ test method appears to be comparable to the currently available validated animal-free test methods in general, and to the h-CLAT method specifically. ESAC therefore considers that the U-SENS™ test method is ready to be considered for regulatory use in the context of an IATA for skin sensitisation and remains a useful tool for screening and early decision making during product development, within the applicability domain defined in the validation study. However, ESAC has concerns with the complexity of the prediction model and the uncertainty of the applicability domain and the predictivity of non-sensitisers (specificity of 33 %) of the 'six rules' for resolving inconclusive results. Given that the 'six rules' appear to change 'inconclusive' to 'sensitisers' in the majority of cases, the test developers are recommended to consider removing these 'six rules' from the prediction model and to consider any inconclusive result as positive by default. This would significantly simplify the prediction model and would allow standardising the maximum number of runs per experiment to three. Considering the validation dataset, modifying the prediction model and the maximum number of runs as suggested above would have little impact on the predictive capacity of the U-SENS™ test method and could lead to a decrease in testing requirements of about 8 %.

References

- 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, OECD, Paris.
- OECD (2016a). Guidance Document on the reporting of Defined Approaches to be used within Integrated Approaches to Testing and Assessment. ENV/JM/MONO(2016)28. Series on Testing and Assessment No. 255, OECD, Paris.
- OECD (2016b). 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. ENV/JM/MONO(2016)29. Series on Testing and Assessment No. 256, OECD, Paris.

Annex 1

COMPOSITION OF ESAC AND ESAC WORKING GROUP

Composition of ESAC and ESAC Working Group

EURL ECVAM Scientific Advisory Committee (ESAC)

- Dr. Neil CARMICHAEL (ESAC Chair)
- Prof. Jürgen BORLAK
- Dr. Harvey CLEWELL
- Prof. Lucio G. COSTA
- Dr. Kristina KEJLOVÁ
- Prof. David John KIRKLAND
- Prof. Annette KOPP-SCHNEIDER
- Dr. Renate KRÄTKE
- Prof. Claus-Michael LEHR
- Dr. José Maria NAVAS
- Prof. Aldert PIERSMA
- Dr. Jonathan RICHMOND
- Dr. Erwin L. ROGGEN
- Dr. Dorothea SESARDIC

ESAC Working Group (WG)

- Dr. Erwin L. ROGGEN (ESAC Member, WG Chair)
- Prof. Annette KOPP-SCHNEIDER (ESAC Member)
- Dr. David BASKETTER (EURL ECVAM nomination)
- Dr. Steve ENOCH (EURL ECVAM nomination)
- Prof. Yong HEO (Catholic University of Daegu; ICATM nomination by KoCVAM)
- Dr. Reiko ADACHI (NIHS; ICATM nomination by JaCVAM)

EURL ECVAM (Secretariat)

- Dr. João BARROSO (ESAC Coordinator)
- Dr. Silvia CASATI
- Prof. Maurice WHELAN (Head of Unit)

Annex 2

EURL ECVAM REQUEST FOR ESAC ADVICE

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health, Consumers and Reference Materials
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

ESAC Request 2016-03

EURL ECVAM Scientific Advisory Committee

(ESAC)

EURL ECVAM REQUEST FOR ESAC ADVICE

on the

L'Oréal-coordinated study on the transferability and reliability of the U-SENS™ test method for skin sensitisation testing

Title page information	
Abbreviated title of ESAC request	U-SENS™ validation
ESAC REQUEST No.	2016-03
Template used for preparing request	EP 3.02
Date of finalising request	23/03/2016
Date of submitting request to ESAC	01/04/2016
Request discussed through	Written procedure previous to ESAC 42
Opinion expected at (date)	ESAC 42 (June 2016)
File name of this request	**ER2016-03_ESAC_REQUEST_U-SENS.doc**

TABLE OF CONTENTS

1. TYPE OF REQUEST

Request Type	Identify request ("YES")
R1 ESAC Peer Review **of a Prevalidation Study or Validation Study**	**YES, external validation study (i.e. not coordinated by EURL ECVAM)**
If R1)applies please specify further:	
►Prevalidation Study	**NO**
►Prospective Validation Study	The U-SENS™ assay for skin sensitization testing (formerly known as Myeloid U937 Skin Sensitisation Test –MUSST) has been evaluated in a two-phase study coordinated by l'Oréal. The study has been designed to generate information on the test method's transferability and reproducibility to allow recommendations to be made on these two aspects of the method. In addition, data generated with the validation study's chemicals and a larger set of l'Oréal's historical data has been used to estimate the predictive capacity of the method compared to responses in the LLNA and in humans. Data generated with the U-SENS™ will be used in defined approaches to testing and assessment or in the context of a WoE assessment (i.e. within Integrated Approaches to Testing and Assessment – IATA) for determining the skin sensitization potential of chemicals.
►Retrospective Validation Study	**NO**
►Validation Study based on Performance Standards	**NO**
R2 Scientific Advice on a test method submitted to EURL ECVAM for validation **(e.g. the test method's biological relevance etc.)**	**NO**
R3 Other Scientific Advice **(e.g. on test methods, their use; on technical issues such as cell culturing, stem cells, definition of performance standards etc.)**	**NO**

2. TITLE OF STUDY OR PROJECT FOR WHICH SCIENTIFIC ADVICE OF THE ESAC IS REQUESTED

L'Oréal validation study on the U-SENS™ assay for skin sensitisation testing

3. BRIEF DESCRIPTION OF THE STUDY OR PROJECT

1) Background to skin sensitization and current regulatory adopted tests

Skin sensitisation is the toxicological endpoint associated with substances that have the intrinsic ability to cause allergic contact dermatitis (ACD) in humans.

The identification of the skin sensitisation potential represents an important component of the safety assessment of new and existing chemicals including cosmetic ingredients. Traditional regulatory predictive tests for skin sensitisation rely on the use of animals. These include: the guinea-pig tests (Buehler Test and Guinea-pig Maximisation Test) (OECD TG 406), the murine Local Lymph Node Assay (LLNA, OECD TG 429) and its non-radio-isotopic variants (OECD TG 422a and 422b).

In recent years progress has been made in the development and regulatory adoption of alternative methods for skin sensitization hazard identification. Two methods validated and/or peer-reviewed by EURL ECVAM, the direct peptide reactivity assay (DPRA) and the KeratinoSens™ have been adopted by the OECD as Test Guidelines 442C and 442D respectively. A third method, the human Cell Line Activation Test is in the final stages of the OECD adoption process at the time of writing of this request.

Methods developed in the area are addressing key chemical/biological mechanisms leading to the acquisition of skin sensitization. Such mechanisms have been described in the form of an Adverse Outcome Pathway (AOP; OECD 2010a, 2010b). Within the skin sensitisation AOP the molecular initiating event (i.e. the first key event - KE) is the covalent binding of electrophilic substances to nucleophilic centres in skin proteins. The second key event takes place in the keratinocytes and includes inflammatory responses as well as changes in gene expression associated with specific cell signalling pathways such as the antioxidant/electrophile response element (ARE)-dependent pathways. The third key event is the activation of dendritic cells (DC), typically assessed by expression of specific cell surface markers, chemokines and cytokines. The fourth key event is T-cell proliferation, which is indirectly assessed in the murine Local Lymph Node Assay (LLNA).

The U-SENS™ is proposed to address KE3 of the AOP. As for the other skin sensitization test methods evaluated by EURL ECVAM and adopted by the OECD, also the U-SENS™ does not have the potential to function as a full-replacement stand-alone method. Instead, it is proposed that a combination of in in silico, in chemico and in vitro tests, addressing the key biological events of skin sensitisation, will be needed to achieve this goal. A defined approach to testing and assessment combining U-SENS data with other information has been developed.

2) The U-SENS™ test method

The U-SENS™ addresses the role that Langerhans cell (LC) and dermal dendritic cells (DC) play in the induction of skin sensitization. These cells are important mediators in the skin sensitization process since they are capable of presenting the hapten-protein conjugate to responsive T lymphocytes in the

lymph nodes draining the site of exposure (Kimber and Cumberbatch, 1992). The maturation process of LC and DC from antigen processing cells to antigen presenting cells is considered a key event in the acquisition of skin sensitisation. This maturation process involves the modulation of the expression of cell surface phenotypic markers, those most commonly reported being CD54, CD80, CD86 and major histocompatibility complex (MHC) class II (Galvao dos Santos et al., 2009). This knowledge has been exploited in the development of in vitro tests based on the use of human monocytic immortalized cell-lines to screen the skin sensitization potential of chemicals.

The U-SENS™ measures the modulation of the CD86 protein markers on the surface of U-937 cells (human monocyte cell line) by flow cytometric analysis, following 45 hours of exposure to at least 4 concentrations of a test chemical. The concentrations used in the test are selected among concentrations predefined in the SOP. The test method is designed to discriminate between sensitising and non-sensitising chemicals whereby chemicals are classified as sensitisers if the CD86-IgG1 percent of positive cells exceeds a defined threshold (i.e. Stimulation Index ≥150) compared to the vehicle control at least to consecutive tested concentration, in at least two independent measurements (i.e. run repetitions). Cell viability is measured concurrently by Propidium Iodide (PI) staining and CD86 values are considered for the prediction only if cell viability is above 70%.

3) History of development of the test method and background to the U-SENS™ validation study

The U-SENS ™ (former MUSST) protocol and prediction model were developed within L'Oréal's laboratories using about 100 reference chemicals with animal and/or human clinical data. The protocol and the prediction model were then assessed within the Colipa's (now Cosmetics Europe) Skin Tolerance Project Team ring study with a small set of chemicals (N=10). In 2009 the test method entered the EURL ECVAM-coordinated validation study together with the DPRA and the h-CLAT. At the time of the EURL ECVAM validation study, the U-SENS was transferred to three naïve laboratories but was withdrawn at the end of the first phase of the study because of reproducibility issues derived by the fact that the prediction model allowed three final calls to be made, Sensitiser, Non-sensitiser and Inconclusive. The prediction model was then optimized within L'Oréal by the addition of a score derived from six new rules allowing only 2 final calls, Sensitiser vs Non Sensitiser to be made for the chemicals that were considered Inconclusive by using the original prediction model only.

Between 2013 and 2014 the U-SENS™ underwent a two-phase evaluation study coordinated by l'Oréal and both submitted to EURL ECVAM.

The first L'Oréal-coordinated study (study 2013) involved the participation of three laboratories (in addition to L'Oréal in-house facility), two of which (i.e. Bioassay and WIL Research (former Ricerca)) had been involved in the evaluation of the method at the time of the EURL ECVAM-coordinated study. Thus the method was newly transferred only to one of the laboratories (CiToxLab), although all of them received training on the new rules of the prediction model and on the software developed for the automated application of the prediction model.
The within-laboratory reproducibility (WLR) was evaluated in L'Oréal laboratory only by testing 21 chemicals (14 coded + 7 not coded) in two independent experiments. The between-laboratory reproducibility (BLR) was assessed on the basis of 14 coded chemicals tested once in each laboratory.

Following the evaluation by EURL ECVAM of the information submitted, l'Orèal was advised to generate additional information in order to progress the method into peer review.

A second study was coordinated by l'Oréal in 2014. In addition to l'Oréal's in-house facility the 2014 study involved the participation of the three laboratories, mentioned above, that took part also in the 2013 study. The 2014 study experimental designed foresaw the testing of 24 coded test chemicals for the evaluation of the between-laboratory reproducibility and the testing of a subset of these

chemicals (15) for the evaluation of the within-laboratory reproducibility.
The predictive capacity of the U-SENS™ method was firstly evaluated on the basis of the results obtained in the two studies of 2013 and 2014 for a total of 38 chemicals. It was further assessed using a larger number of chemicals (175) covering the entire range of skin sensitisation potency based on human and LLNA data.

Conclusions of the Test Submitter on the information submitted to EURL ECVAM

According to l'Oréal the information submitted to EURL ECVAM for entering the peer-review process indicates the potential of the U-SENS™ test method to contribute to the discrimination between sensitisers and non-sensitisers as defined by the Globally Harmonised System (GHS) of classification and labelling of substances (category 1; no category) (UN, 2011) and as implemented in the European Commission Regulation on classification, labelling and packaging (CLP) of substances and mixtures (EC, 2008), although the use of U-SENS™ as a standalone method for this purpose will not be recommended as stated as well for all preceded validated test methods addressing the skin sensitization endpoint. The U-SENS™ is foreseen to be combined with complementary information and evaluated in the context of Integrated Approaches to Testing and Assessment (IATA). In such context, the U-SENS™ test method is part of a decision strategy for skin sensitization hazard identification based on in silico, in chemico, and in vitro data. The decision strategy has been already submitted to the OECD as a case study in the context of the draft guidance document on the reporting of defined approaches to testing and assessment.
The U-SENS™ test method is also foreseen to be a part of an integrated test battery for risk and safety assessment which will be able to fully replace the in vivo test methods In such context, the U-SENS™ test method is part of The European Cosmetics Industry Trade Association current (Cosmetics Europe) program.

References

1. Anon (2003) Globally Harmonised System of Classification and Labelling of Chemicals (GHS). Part 3: Health and Environmental Hazards, pp. 151-158. New York, NY, USA, and Geneva, Switzerland: United Nations Organisation.
2. EC (2008) Regulation (EC) No 1272/2008 (16 December 2008) of the European Parliament and of the Council on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Official Journal of the European Union L 353, (31/12/2008) p. 1-1355.
3. Galvao dos Santos G, Reinders J, Ouwehand K, Rustemeyer T, Schepper RJ & Gibbs S (2009) Progress on the development of human in vitro dendritic cell based assays for assessment of the sensitizing potential of a compound. Toxicology and Applied Toxicology 236(3); 372-382.
4. Kimber I & Cumberbatch M (1992) Dendritic cells and cutaneous immune responses to chemical allergens. Toxicology and Applied Pharmacology 117; 137-46.
5. OECD 2012a. 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
6. OECD 2012b. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 2: Use of the AOP to Develop Chemical Categories and Integrated Assessment and Testing Approaches. Series on Testing and Assessment No. 168 http://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm

4. OBJECTIVES, QUESTIONS, TIMELINES

4.1 OBJECTIVE

Objective *Why does EURL ECVAM require advice on the current issue?*	The opinion of ESAC should support EURL ECVAM with respect to the development of an EURL ECVAM recommendation on the U-SENS™ assay outlining (1) the scientific basis of the assay, (2) its overall performance as assessed during the study and based on other (e.g. published) information, (3) its applicability and limitations, 4) proposed use of the method 5) additional work to be undertaken in future to further characterise the test method.

4.2 QUESTION(S) TO BE ADDRESSED

Questions *What are the questions and issues that should be addressed in view of achieving the objective of the advice?*	**1) DESIGN & CONDUCT OF STUDY**: The ESAC is requested to review whether the study was conducted appropriately in view of the objective of the study. The study objective was to assess (1) the reproducibility of the method within one laboratory (WLR) (2) its transferability to other laboratories (3) its reproducibility between laboratories (BLR) (4) An indication of the predictive capacity of the test method for distinguishing between sensitisers and non-sensitisers and, where possible, to appraise its potential to contribute to a further sub-categorisation of sensitisers into two subcategories (1A and 1B). When reviewing the design and conduct of the study, the following issues should be addressed in particular: (a) Clarity of the test definition (module 1) (b) Clarity of the definition of the study objective and study management (c) Appropriateness of the study design & execution in view of the study objectives, inter alia: - Is the number of tested chemicals sufficient for the purposes of the study? - Are the reference data used for assessing in particular the predictive capacity appropriate and of good quality? - Was the identification of chemicals conducted in an appropriate manner (i.e. presence or absence of selection criteria, justification etc.)? - Is the adverse effect range of the selected chemicals appropriate for the purpose of the study - In case of gaps (chemical class etc.) – are these justified? - Is the number of laboratories sufficient? (d) Appropriateness of the study execution (e.g. was the number of repetitions sufficient? How were exceptions / deviations handled? Were provisions specified for retesting? etc.)

	(e) Appropriateness of the statistical analysis used for analysing WLR, transferability, BLR and for providing an indication of the predictive capacity. 2) CONCLUSIONS OF STUDY: The ESAC is requested to assess whether the conclusions, as presented in the Validation Study Report, are substantiated by the information generated in the study and are plausible with respect to existing information and current views (e.g. literature). In particular: (a) Are the conclusions on reproducibility (WLR and BLR) as well as transferability justified and plausible? (b) Are the conclusions on the predictive capacity justified and plausible with respect to existing information (c) Are there possible gaps between study design and study conclusions which remain to be addressed in view of the suggested conclusions / use (see also point 3)? 3) APPLICABILITY AND LIMITATIONS OF THE TEST METHOD: The ESAC is requested to evaluate, whether the information provided in the submission, provides sufficient information on the applicability and possible limitations of the test method. 4) PROPOSED USE OF THE TEST METHOD: The ESAC is requested to evaluate, whether the information provided in the submission are sufficient to substantiate the proposed use of the method 5) FUTURE ACTIVITIES ON THE METHOD: The ESAC is requested to recommend what additional work, if necessary, should be undertaken in future to further characterise the test method and its proposed use.

4.3 TIMELINES

Timelines concerning this request *When does EURL ECVAM require the advice?*	**Timeline**	**Indication**
	Finalised ESAC Opinion required by:	June 2016
	Request to be presented to ESAC by written procedure (e.g. due to urgency) prior to the next ESAC	YES
	Request to be presented to ESAC at ESAC plenary meeting	NO

5. EURL ECVAM PROPOSALS ON HOW TO ADDRESS THE REQUEST WITHIN ESAC

5.1 EURL ECVAM PROPOSAL REGARDING REQUEST-RELATED STRUCTURES REQUIRED

Specific structures required within ESAC to address the request *Does the advice require an ESAC working group, an ESAC rapporteur etc.?*	**Structure(s) required**	**Required according to EURL ECVAM? (YES/NO)**
	S1 ESAC Rapporteur	NO
	S2 ESAC Working Group	**ESAC members** - Erwin Roggen (Chair) - Annette Kopp-Schneider **EURL ECVAM nominations** - David Basketter (DABMEB Consultancy) - Steven Enoch (Liverpool John Moores University) **ICATM nominations** - Joanna Matheson (CPSC; nominated by ICCVAM) - Yong Heo (Catholic University of Daegu; nominated by KoCVAM) - Reiko Adachi (NIHS; nominated by JaCVAM)
	S3 Invited Experts	NO
	Ad S3: If yes – list names and affiliations of suggested experts to be invited and specify whether these are member of the EEP	
	If other than above (S1-S3):	

5.2 DELIVERABLES AS PROPOSED BY EURL ECVAM

Deliverables *What deliverables (other than the ESAC opinion) are required for addressing the request?*	**Title of deliverable other than ESAC opinion**	**Required? (YES/NO)**
	D1 ESAC Rapporteur Report and draft opinion	NO
	D2 ESAC Peer Review Report and draft opinion	**YES**
	If other than above (D1-D2):	

6. LIST OF DOCUMENTS TO BE MADE AVAILABLE TO THE ESAC

Count	Description of document	Already available? (YES/NO)	File name
1	Test Submission on the U-SENS™	YES	TST U-SENS revision_11 11 2015.pdf
2	EURL ECVAM assessment report on the U-SENS™ and l'Oréal responses addressed also in the Test Submission	YES	TM2013_02_U-SENS_assessment report responses.pdf
3	Protocol(s) of the test method	YES	Attachment 1a_MUSST SOP.pdf Attachment 1c_6 rules examples.pdf Attachment 1d_SOP MUSST automation.pdf Attachment 1b_DB-ALM U-SENS.pdf
4	Test items used to assess WLR	YES	Attachment 3_WLR test items.pdf
5	Data used for WLR assessment	YES	Attachment 4_WLR assessment.pdf
6	Test items used for transferability	YES	Attachment 5_TF test items.pdf
7	Training and transfer protocols	YES	Attachment 6 & 7_ Training and Transfer protocol.pdf
8	Transfer reports to Bioassay, Ricerca, CiToxlab	YES	Attachment 8a _ Transfer report Bioassay.pdf Attachment 8b _ Transfer report Ricerca .pdf Attachment 8c_Transfer report CiToxLab.pdf
9	Test items used to assess BLR	YES	Attachment 9_BLR test items.pdf
10	Data used for BLR assessment (Tables and/or Figures)	YES	Attachment 10_BLR assessment.pdf
11	List of test items used to assess PC	YES	Attachment 11_PC test items v2.pdf

12	Data used for PC assessment	YES	Attachment 12a_PC assessment.pdf Attachment 12b_PC 175 test items v2.pdf Attachment 12c_PC assessment 175 test items.pdf Attachment 12d_PC assessment protein binding classes.pdf
13	Project plan	YES	Attachment 13_Project plan.pdf
14	List of test items used in the validation study	YES	Attachment 14_Validation study tests items.pdf
15	Coding and distribution of chemicals	YES	Attachment 18a_2014 Study product coordination report.pdf
16	Statistical reports	YES	Attachment 18b_2014 Statistical report.pdf Attachment 18m_2014 Amendment 1 to statistical report.pdf Attachment 18n_2015 Amendment 2 to statistical report.pdf
17	Results from individual laboratories	YES	Attachment 18c_Bioassay 2014 results.pdf Attachment 18d_CiToxLAB 2014 results.pdf Attachment 18e_L'Oreal 2014 results.pdf Attachment 18f_WIL Research 2014 results.pdf Attachment 18g_L'OrÃ©al manual results 2013.pdf Attachment 18h_L'OrÃ©al auto results 2013.pdf
18	Test presubmission form to EURL ECVAM	YES	Attachment 18i_ TM203-02.pdf
19	IATA case study reported in the draft OECD GD	YES	Attachment 18j_ OECD IATA submitted case study v2.pdf
20	Relevant publications	YES	Attachment 18k_ Piroird et al 2015.pdf Attachment 18l_AlÚpÚe et al 2015.pdf

7. TERMS OF REFERENCE OF THE ESAC WORKING GROUP

7.1 ESTABLISHMENT OF THE ESAC WORKING GROUP

The ESAC unanimously agreed by written procedure on the 2nd of March 2016 on the composition of a new ESAC Working Group for the review of test methods in the area of skin sensitisation.

7.2 TITLE OF THE ESAC WORKING GROUP

Full title:
ESAC Working Group on Skin Sensitisation Test Methods
Abbreviated title:

ESAC WG Sensitisation

7.3 MANDATE OF THE ESAC WORKING GROUP

The EWG is requested to conduct a scientific review of the L'Oréal-coordinated validation study concerning the U-SENS. The review needs to address the questions put forward to ESAC by EURL ECVAM.

The review should focus on the appropriateness of design and conduct of the study in view of the study objective and should provide an appraisal to which extent the conclusions of the test submitter are substantiated by the information generated during the study and how the information generated relates to the scientific background available.

7.4 DELIVERABLES OF THE ESAC WORKING GROUP

The ESAC WG is requested to deliver to the chair of the ESAC and the ESAC Coordinator a detailed **ESAC Working Group Report** outlining its analyses and conclusions and a **draft ESAC Opinion.** A reporting template has been appended (Appendix 1) intended to facilitate the drafting of the report.

The conclusions drawn in the report should be based preferably on consensus. If no consensus can be achieved, the report should clearly outline the differences in the appraisals and provide appropriate scientific justifications.

7.5 PROPOSED TIMELINES OF THE ESAC WORKING GROUP

Item	Proposed date/time	Action	Deliverable
1	17-19 May 2016	Working group meeting	Draft ESAC WG report and draft ESAC opinion
2	27 May 2016	Circulation of final WG report and draft ESAC opinion to ESAC	Final draft ESAC WG report and draft ESAC opinion
3	9-10 June 2016	Endorsement of WG report and ESAC opinion at ESAC42 meeting	Final ESAC WG report and ESAC opinion

7.6 QUESTIONS WHICH SHOULD BE ADDRESSED BY THE ESAC WORKING GROUP

The ESAC WG is requested to address the **questions posed to the ESAC** which have been broken down further in more **specific questions** by the ESAC chair, the chair of the ESAC WG and the Secretariat (see section 4.2).

When preparing the final ESAC WG report to address these questions, the ESAC WG is requested to use a pre-defined reporting template. This template (see appendix 1) follows EURL ECVAM's modular approach and addresses to which extent the standard information requirements have been addressed by the study. The template allows moreover for addressing the issues specific studies outlined in section 4.2. The Secretariat will provide guidance if necessary.

APPENDIX 1 REPORTING TEMPLATE

The appended ESAC WG template suggests a structure that is in close agreement with the EURL ECVAM information requirements ("modules") for scientific review following validation and allows at the same time for the description of the analysis and conclusions concerning more specific questions.

The template can be used for various types of validation studies (e.g. prospective full studies, retrospective studies, performance-based studies and prevalidation studies). Depending on the study type and the objective of the study, not all sections may be applicable.

However, for reasons of consistency and to clearly identify which information requirements have not been sufficiently addressed by a specific study, this template is uniformly used for the evaluation of validation studies.

The current template is

TEMPLATE_ESAC-WG_REPORT-v6.doc

Annex 3

ESAC WORKING GROUP PEER REVIEW CONSENSUS REPORT

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE
Directorate F - Health, Consumers and Reference Materials
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)

ESAC Working Group Peer Review Consensus Report

on the

L'Oréal-coordinated study on the transferability and reliability of the U-SENS™ test method for skin sensitisation testing

Title page information	
File name	ESAC_WG_Report_U-SENS.doc
Abbreviated title of ESAC request	U-SENS™ validation
Relating to ESAC REQUEST No.	2016-03
Request discussed through	Written procedure previous to ESAC 42
Report to be handed over to ESAC Chair and EURL ECVAM Coordinator by	Erwin L. Roggen

Version tracking

Date	Version	Author(s)	Description
19 May 2016	V1.0	ESAC WG	First draft agreed by the ESAC WG
27 May 2016	V2.0	ESAC WG	Second revised draft after commenting
01 June 2016	V3.0	ESAC WG	Third revised draft after commenting
06 June 2016	V4.0	ESAC WG	Final ESAC WG draft sent to ESAC for endorsement

Table of Contents

ESAC Working Group

This report was prepared by the "ESAC Working Group for Skin Sensitisation" (ESAC WG SS), charged with conducting a detailed scientific peer review of on the L'Oréal-coordinated study on the transferability and reliability of the U-SENS™ assay for skin sensitisation testing.

The ESAC WG had the following members:

ESAC members

- Erwin Roggen (Chair)
- Annette Kopp-Schneider

EURL ECVAM nominations

- David Basketter (DABMEB Consultancy)
- Steven Enoch (Liverpool John Moores University)

ICATM nominations

- Yong Heo (Catholic University of Daegu; ICATM nomination by KoCVAM)
- Reiko Adachi (NIHS; ICATM nomination by JaCVAM)

ESAC Coordination:

- João Barroso (ESAC Coordinator)
- Silvia Casati

Abbreviations used in the document

- **BLR** Between-laboratory reproducibility
- **EURL ECVAM** European Union Reference Laboratory for Alternatives to Animal Testing
- **ESAC** EURL ECVAM Scientific Advisory Committee
- **ESAC WG** ESAC Working Group
- **GLP** Good Laboratory Practice
- **IATA** Integrated Approach to Testing and Assessment
- **INC** Inconclusive
- **SOP** Standard Operating Procedure (used here as equivalent to 'protocol')
- **VMT** Validation Management Team
- **WLR** Within-laboratory reproducibility

1. Study objective and design

1.1 Analysis of the clarity of the study objective's definition

(a) ESAC WG summary of the study objective as outlined in the Test Submission

From the project plan:

Formal validation of the U-SENS™ (improved MUSST) test, focusing on its transferability and reproducibility. This validation includes an earlier 'ring trial' evaluation of the transferability and reproducibility of the test.

The above is formulated as a study goal (goal 1), as is a preliminary assessment of the accuracy of the method in discriminating sensitisers and non-sensitisers (according to UN GHS for classification and labelling of chemicals for skin sensitisation (cat. 1; no cat.) (UN, 2015) and ECR No. 1272/2008 on classification, labelling and packing of substances and mixtures (EC, 2008)) (goal 2).

(b) Appraisal of the clarity of study objective as outlined in the Test Submission

The study objective (and goals 1 and 2) is sufficiently clear to allow persons skilled in the process of validation of cell-based assays to understand.

1.2 Quality of the background provided concerning the purpose of the test method

The U-SENS™ test method is considered to be a potentially important component of the safety evaluation of chemicals. Skin sensitisation assessment represents a standard requirement of chemical legislation (UN GHS for classification and labelling of chemicals for skin sensitisation (cat. 1; no cat.) (UN, 2015) and ECR No. 1272/2008 on classification, labelling and packing of substances and mixtures (EC, 2008)).

The U-SENS™ test method is foreseen to be part of a battery or Integrated Approach to Testing and Assessment (IATA) for replacement of the existing *in vivo* assays for hazard identification.

(a) Analysis of the scientific rationale provided in the Test Submission

The scientific background of the U-SENS™ is identical to the background of the h-CLAT, which recently became an OECD test guideline (OECD, 2016). Several membrane markers, including CD86, are generally accepted to reflect dendritic cell (DC) migration/maturation. DC maturation results in migration of the DC to the lymphoid tissue and an increasing capacity of the DC to present antigen to T-cells. This sequence of events is known to result in an immune response. Sufficient reference to the relevant body of scientific literature was made.

It is well known that also skin sensitisers induce DC migration/maturation, and thus CD86 upregulation, leading to an (adverse) immune response. This sequence of events was outlined by the test submitters in a concise but precise way.

The U-SENS™ was compared with the very similar h-CLAT (measuring CD86 and CD54). The added value of the U-SENS™ (as compared to the h-CLAT) was not provided.

The provided scientific background correctly profiles the U-SENS™ as a test method covering key event 3 of the OECD Adverse Outcome Pathway for Skin Sensitisation (OECD, 2012).

(b) Analysis of the regulatory rationale provided in the Test Submission

The provided regulatory rationale builds largely on the Cosmetics Regulation, and the emergence and application of OECD test guidelines in the context of REACH.

Skin sensitisation assessment represents a standard requirement of chemical legislation (UN GHS for classification and labelling of chemicals for skin sensitisation (cat. 1; no cat.) (UN, 2015) and ECR No. 1272/2008 on classification, labelling and packing of substances and mixtures (EC, 2008)).

1.3 Appraisal of the appropriateness of the study design

The validation of the U-SENS™ comprises 2 independent studies.

1. A ring trial conducted in 2013 for a preliminary assessment of the reliability of the test (2013) (14 substances);
2. A formal validation study following EURL ECVAM recommendations (2014) (24 substances).

The results of the 2013 ring trial were submitted to EURL ECVAM. EURL ECVAM recognised that the study was performed properly, but with an insufficient number of substances, leading to a decision to initiate the 2014 validation study.

All the measures required for proper formal validation (2014 study) of a test method were taken: independent Study Product Coordinator and Study Data Coordinator; substances were blinded and statistical analysis of the reproducibility was conducted blind.

Four laboratories were involved in the study, instead of the required minimum of three. The laboratories were involved in both the 2013 and 2014 studies.

All laboratories tested the 24 substances once. Fifteen of them were tested in three independent experiments by all laboratories with three differently coded samples provided. A rationale for the number of test substances (other than adaption of the EURL ECVAM study design) was not provided in the study plan.

Three laboratories worked under GLP compliance. Also for the non/GLP laboratory, measures for assuring Quality were defined (pp. 16/17 of the TST 112015). In case of non/GLP, these measures were specified.

1.4 Appropriateness of the statistical evaluation

The Prediction Model is complex. First, two runs are performed. If they lead to concordant classification as S or NS, no more runs are performed. In case of discordance, a third run is performed. If this run leads to a classification as S or NS, the classification is based on the majority vote. U-SENS™ may also lead to inconclusive (INC) results. Concordant INC results after two, or, if it was required, three runs, trigger a decision requiring the application of a system of six decision rules on the averaged dose-response data from all runs. The prediction model is shown in a decision tree. However, when applied on the data shown in Annex A of attachment 18n (second amendment to the statistical report) of the submission to EURL ECVAM, the ESAC WG observed a number of apparent deviations from the decision tree. It was suggested that a more detailed flow chart was incorporated into the SOP.

Analysis of the performance of this complex prediction model using the Bootstrap resampling approach, although ideal, would be challenging and very time consuming. Consequently, the second amendment to the statistical report (attachment 18n of the submission to EURL ECVAM) shows a model-based Monte-Carlo simulation for the evaluation of BLR, WLR and predictive capacity. The proposed method seems to rely on the evaluation of single runs for which INC results are mapped to S or NS based on the system of six decision rules for the individual run. Each experiment was then simulated to consist of 2 or 3 runs after which a majority voting rule is applied. Hence the assessment

of BLR, WLR and predictive capacity offered in the second amendment to the statistical report does not match the prediction model shown in the study report.

2. Collection of existing data

2.1 Existing data used as reference data

The chemicals selected for the formal validation were selected from the list of reference chemicals (animal and/or human clinical data) of L'Oréal's laboratories. Both liquids and solids were evaluated.

In order to analyse the predictivity of the U-SENS™ test method a set of 175 substances has been evaluated by L'Oréal. The human data was the main criterion considered for the prediction. According to human, LLNA and EU-CLP classifications, all potency classes, from extreme sensitiser to non-sensitiser were represented.

2.2 Existing data used as testing data

See section 2.1.

2.3 Search strategy for retrieving existing data

Selection criteria for substances to be used for transferability and reproducibility assessment were:

- Availability of reliable *in vivo* LLNA reference data: sensitisers (S) and non-sensitisers (NS),
- Commercial availability,
- Diversity in terms of chemical properties with different reactivity toward proteins (as described by Aptula and Roberts (2006) and determined with the Toxtree software) and with ingredients used in Cosmetics and non-cosmetic chemicals

Selection criteria for substances to be used for predictive capacity evaluation were:

- Availability of robust *in vivo* data to allow a proper comparative evaluation of *in vitro* results. As such, availability of human (Basketter et al., 2014) and/or LLNA (OECD, 2010) *in vivo* skin sensitisation classification data was considered.

2.4 Selection criteria applied to existing data

Not specified, beyond the availability of animal and/or human clinical data.

3. Quality aspects relating to data generated during the study

3.1 Quality assurance systems used when generating the data

Three laboratories worked under GLP compliance. Also for the non/GLP laboratory, measures for assuring Quality were defined (e.g. pp. 16/17 of the Study report 112015). In case of non/GLP, these measures were specified.

Good cell culture practice was not mentioned.

3.2 Quality check of the generated data prior to analysis

No additional measures were taken beyond the quality assurance mentioned in 3.1.

4. Quality of data used for the purpose of the study (existing and newly generated)

4.1 Overall quality of the evaluated testing data (newly generated or existing)

Overall, application of the quality criteria to the test system seems to assure good quality data, but no independent data analysis before transfer to the biostatistician seems to exist.

However, the unexpected increase in non-qualified tests during the formal validation study at two of the participating laboratories does raise concerns.

4.2 Quality of the reference data for evaluating relevance[1]

The quality of the reference data was sufficient for the purpose of the study.

4.3 Sufficiency of the evaluated data in view of the study objective

The overall quality of the data is sufficient to draw conclusions on the transferability and reproducibility (within as well as between laboratories).

5. Test definition (Module 1)

The test method is adequately described, and covers the parameters that are relevant for good test performance.

1. The test description specifies the cell line used for the test and addresses the IPR issues related to commercial use of this cell line.
2. The relevance of the endpoints measured is explained on the background of the available scientific knowledge.
3. Quality and acceptance criteria for untreated cells, as well as positive and negative controls, are described.
4. Acceptance criteria for the results are outlined. It is not clear though why there is no referral to the quality and acceptance criteria for unexposed viability and CD86 levels nor the positive and negative control values.
5. The prediction model was not found to be described in sufficient detail to repeat the conclusions made by the VMT as presented in Annex A of attachment 18n (second amendment to the statistical report) of the submission to EURL ECVAM.

[1] OECD guidance document No. 34 on validation defines relevance as follows: "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 accuracy (concordance) of a test method."

6. The ESAC WG has concerns about the 'six rules', which were not alleviated by receiving details of the 60 chemicals used to develop the rules. According to the submitter, 60 chemicals were used as training set to develop the 'six rules', of which 33 were sensitisers and 27 were non-sensitisers.

   Thirteen chemicals of the training set (11 sensitisers and 2 non-sensitisers) were included in the 175 chemicals tested at L'Oréal to assess the predictive capacity of U-SENS™, whereas a total of 35 out of these 175 chemicals required the application of the 'six rules'. Therefore, only 37 % of the chemicals requiring the application of the rules in the list of 175 were part of the training set (13/35). However, of these 35 chemicals only three were non-sensitisers *in vivo* and only one of these three was correctly predicted by the 'six rules'. Indeed, for the 35 out of 175 chemicals for which the 'six rules' were applied, the sensitivity obtained was 100 % (32/32), while the specificity was only 33 % (1/3).

   Similarly, of the 38 chemicals tested in the multi-laboratory validation trial (300 experiments), only two (12 experiments) were part of the training set, whereas 17 of the 38 chemicals (38 experiments) had at least one experiment requiring application of the 'six rules'. Again, of these 17 chemicals only six were non-sensitisers *in vivo*, accounting for a total of nine experiments. Considering the 38 experiments from the multi-laboratory validation trial for which the 'six rules' were applied (17 out of 38 chemicals), only 3 out of 29 experiments with sensitisers (10 %) were misclassified by the 'six rules' as compared to 6 out of 9 experiments with non-sensitisers (67 %).

   There is therefore no evidence of any predictive value of the 'six rules'. In fact, it appears that the rules have no added value to a simple decision that all INC results should be considered positive: Of the 38 experiments in the multi-laboratory validation trial that required application of the 'six rules', 26 were true positives, 6 were false positives, 3 were true negatives and 3 were false negatives. Applying a positive outcome to all INC results in this dataset by default would lead to three extra false positive experiments but also to three less false negative experiments.

7. The expected frequency of non-qualified tests based on the experience of the test developer was defined.

6. Test materials

6.1 Sufficiency of the number of evaluated test items in view of the study objective

The number of test items is considered sufficient to assess the transferability, and within and between laboratory reproducibility of the test method.

6.2 Representativeness of the test items with respect to applicability

The 2013 ring trial included eleven sensitisers (1 extreme, 3 strong, 5 moderate, 2 weak) and three non-sensitisers. The underrepresentation of the non-sensitisers was corrected by the 2014 study.

The test substances selected (for the 2014 validation study) include strong (N=3), moderate (N=3) and weak (N=2) sensitisers, one false positive substance and fifteen clear negative substances.

While the distribution of the selected test substances may be sufficient to indicate major issues related to applicability domain and limitations, it may not be sufficient to identify minor (e.g. chemical reactivity class related) issues.

7. Within-laboratory reproducibility (WLR) (Module 2)

7.1 Assessment of repeatability and reproducibility in the same laboratory

The WLR was assessed based on data from the 2013 ring trial and the 2014 validation study.

- Ring trial 2013:
    - 21 test substances were evaluated within L'Oréal.
    - Two independent experiments by L'Oréal (reviewed by EURL ECVAM), followed by an additional independent experiment conducted by L'Oréal.
    - Based on the L'Oréal data, 20/21 substances were concordant (WLR = 95.2 %). Chlorobenzene was the outlier. A plausible explanation was provided: poor solubility and reactivity with cell membranes.
    - The within laboratory reproducibility was very good as far as the overall binary prediction (Y/N) is concerned. However, discordant results between the independent experiments may occur and may be laboratory dependent.
- Validation study 2014:
    - 15 test substances were evaluated within four laboratories.
    - Three independent experiments per laboratory, three differently coded samples provided.
    - Based on the WLRs for each of the participating laboratories (100 %, 93.3 %, 73.3 % and 100 %), an average WLR of 91.7 % was obtained.
    - Outliers were explained but the ESAC WG was not satisfied by some of the explanations (See Section 11).

7.2 Conclusion on within-laboratory reproducibility as assessed by the study

Overall, the WLR is very good, however a lower reproducibility was observed with one of the naïve laboratories. This may highlight the importance of proper training. Since proper training opportunities were provided by the test developer, a lower reproducibility may also indicate additional challenges for a naïve laboratory trying to apply the method.

In spite of this one laboratory performing weaker than the others, the ESAC WG considers the WLR comparable, arguably better, than recently validated methods listed as OECD test guidelines.

8. Transferability (Module 3)

8.1 Quality of design and analysis of the transfer phase

The test substances for the transfer phase were carefully selected following predefined criteria.

A detailed training protocol was provided. One (hands-on) training week (included discussion of the protocol) was organised at L'Oréal. In house training at the CRO was necessary. Each naïve laboratory was expected to perform at least 2 runs per week for 3-6 weeks. The criteria describing a successful transfer were predefined, and include, among others, compliance with the U-SENS™ acceptance criteria, accuracy in selecting/discarding controls and stable CD86 expression, reproducibility.

8.2 Conclusion on transferability to a naïve laboratory / naïve laboratories as assessed by the study

Based on the provided transfer report, the transfer of the test method to the participating laboratories can be considered to have been successful.

9. Between-laboratory reproducibility (BLR) (Module 4)

9.1 Assessment of reproducibility in different laboratories

The BLR was assessed based on data from the 2013 ring trial and the 2014 validation study.

- Ring trial 2013:
    - 14 test substances were evaluated, of which the majority were sensitisers with characteristics potentially interfering with test performance.
    - Two independent experiments (reviewed by EURL ECVAM) by the four laboratories, followed by an additional independent experiment conducted by L'Oréal.
    - Based on the data from the four laboratories, an average BLR of 79 % was obtained (11/14 substances concordant in all four laboratories).
- Validation study 2014:
    - 24 test substances were evaluated, the majority being non-sensitisers.
    - Three independent experiments per laboratory, three differently coded samples provided.
    - Based on the WLRs for each of the participating laboratories, an average BLR of 87.5 % (21/24 substances concordant in all four laboratories) was obtained.

Taking the data from the 2013 and 2014 studies together, the U-SENS™ exhibited a BLR of 84 % in terms of Y/N score.

Explanations were provided for most of the substances not resulting in reproducible classifications. The ESAC WG does not fully agree with these explanations (See Section 11).

9.2 Conclusion on between-laboratory reproducibility as assessed by the study

Overall, the BLR is very good as far as the overall binary classification (Y/N) is concerned. However, discordant results between the independent experiments may occur and may be laboratory dependent.

The studies also hint issues with substances of low solubility and volatile substances (see WLR studies), but also with substances with the potential to cause colour interference.

10. Predictive capacity and overall relevance (Module 5)

10.1 Adequacy of the assessment of the predictive capacity in view of the purpose

The predictive capacity of the U-SENS™ was assessed on the basis of the 2013 ring trial, the 2014 validation study and the assessment of 175 test substances by L'Oréal only.

The substances were selected on the basis of availability of robust animal and/or human clinical data (Basketter et al., 2014).

Various ingredient classes (29 % fragrances, 15 % preservatives, 8 % dyes) and protein binding (chemical reactivity) classes (65 % belonging to at least one of the five classes) were represented.

1. 2013 ring trial:

- Not provided

2. 2014 validation study:

- Reference data were LLNA data
- Predictive capacity on 24 substances:

3. 2013 ring trial + 2014 validation study

- Reference data were LLNA data
- Predictive capacity on 38 substances:
    - Sensitivity: 97 % (95 %, 100 %, 100 %, 95 %)
    - Specificity: 89 % (100 %, 74 %, 89 %, 95 %)
    - Accuracy: 93 % (97 %, 87 %, 95 %, 95 %)

4. L'Oréal 175 substances, including 35 of the substances used in the previous studies.

- 166 substances with LLNA data, 101 substances with human data and 92 substances with LLNA and human data.
- Overall performance:
    - Sensitivity: 95 %
    - Specificity: 65 %
    - Accuracy: 85 %
- The sensitivity against potent and human categories 1-4 was 100 %.
- Against human data (N=101) with human class 5 chemicals considered as sensitisers:
    - Sensitivity: 89 % (95 %)
    - Specificity: 79 % (59 %)
    - Accuracy: 88 % (83 %)

  The values in brackets show the values with human class 5 chemicals are considered as non-sensitisers as proposed by Basketter et al. (2014).
- Against substances with both human and LLNA data (N=92):
    - Sensitivity: 90 % (92 %)
    - Specificity: 60 % (55 %)
    - Accuracy: 85 % (84 %)

  Values obtained against LLNA are shown in brackets.

10.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

The data from the 2013 ring trial and the 2014 validation study suggest that the predictive capacity is good as far as the overall binary classification (Y/N) is concerned. The assay seems to have a good sensitivity perhaps at the expense of specificity. This conclusion is underpinned by the low specificity of the test method when applied to the larger dataset of 175 substances. Nevertheless, overall, the test method's accuracy is comparable with accepted *in vitro* methods.

Where directly compared, performance against human data and LLNA data was very similar.

Explanations for the false positive and false negative results were provided. False positives seem to be associated with membrane disrupting substances. The studies also lend credence to the potential for issues with substances of low solubility and volatile substances (see Section 11). These characteristics seem to affect primarily the specificity of the U-SENS™.

11. Applicability domain (Module 6)

11.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

The primary goal of the study was to assess transferability and reproducibility (WLR and BLR) of the test method. Neither the 2013 ring trial nor the 2014 validation study were designed to provide extensive information on the applicability domain of the method.

However, these studies in combination with the large study on 175 substances covered a broad range of potency and chemical reactivity classes. Analysis of these data revealed that outliers did not correspond with a particular potency or chemical reactivity class. Prehaptens were correctly identified.

The test method submitters provided a list of physico-chemical characteristics that are believed to be outside the applicability domain of the test method: substances of low solubility, rapid oxidation, volatile substances, substances disrupting cell membranes and substances with colour interference.

However, the data show that 8/8 lipophilic substances were correctly predicted (although all positive *in vivo*). Thus, the applicability domain of the test method seems to include this type of substances with reduced solubility in cell culture medium. In addition, analysis of all the data revealed that all dyes were correctly classified.

The ESAC WG disagrees with the explanation by the VMT that benzyl alcohol may be wrongly classified because it is a volatile fragrance. This was apparently a problem in one laboratory only, the same laboratory that showed a lower reproducibility. Furthermore, in order to define a characteristic (e.g. volatility) as a problem, it is necessary to look at the entire group of chemicals having the same characteristic. The VMT also fails to explain satisfactorily why this laboratory failed to accurately score lactic acid, streptomycin and polyethylene glycol, even after application of the set of 'six rules'.

Overall the presented evidence seems to indicate that volatile substances or poorly soluble substances may require additional attention in the technology transfer and training phase, as improper handling of such substances during exposure could affect the reproducibility of the results rather than their accuracy.

11.2 Quality of the description of applicability domain, limitations, exclusions

In depth analysis of the data revealed potential issues with substances of low solubility or stability in an aqueous environment, substances interfering with flow cytometry analysis, volatile substances and substances disrupting cell membranes. These limitations are likely to be very similar to other submerged cell culture assays (e.g. h-CLAT, KeratinoSens™).

Specific comments:

1. Solubility:

The ESAC WG noted the limited choice of solvents. Not all of the test substances may have been sufficiently soluble in either RPMI or DMSO, and speculates to what extent this affects (e.g. in unexperienced laboratories) the reproducibility of the test method.

2. Volatility:

The use of volatility to exclude substances from the applicability domain requires analysis of the entire data to set appropriate criteria. It is likely to apply only for a small subset of chemicals with high volatility coupled with low reactivity (e.g. methyl methacrylate).

Against that background it is unlikely that benzyl alcohol is an outlier for reasons of volatility.

3. Chemical classes used to define the 'six rules'.

It is of concern that the 'six rules' were trained to fix INC predictions towards positive results. It is not clear to what extent the training set of chemicals is representative in chemical space.

4. The ESAC WG noticed that Streptomycin is an unusual substance that is normally negative in predictive tests despite of being a well-recognised human sensitiser.

12. Performance standards (Module 7)

12.1 Adequacy of the proposed Essential Test Method Components

Not applicable.

12.2 Adequacy of proposed Reference Chemicals

Not applicable.

12.3 Adequacy of proposed performance target values

Not applicable.

13. Readiness for standardised use

13.1 Assessment of the readiness <u>for regulatory purposes</u>

The provided data (good WLR, BLR and predictive capacity) suggest that the U-SENS™ test method may be a useful component of a strategy for the regulatory testing and assessment of the skin sensitising potential of substances belonging to the applicability domain defined in the validation study.

The major concern of the ESAC WG is the complexity of the prediction model and the uncertainty of the applicability domain of the 'six rules' for deciding on INC results. In addition, the 'six rules' appear to be highly sensitive but without specificity. The ESAC WG considers these to be an obstacle for regulatory readiness.

13.2 Assessment of the readiness for other uses

The provided data (good WLR, BLR and predictive capacity) suggest that the U-SENS™ test method can be used for screening and early decision-making during product development.

13.3 Critical aspects impacting on standardised use

Proper training is required.

13.4 Gap analysis

None identified.

14. Other considerations

None.

15. Conclusions on the study

15.1 ESAC WG summary of the results and conclusions of the study

The VMT concluded that the WLR, BLR and predictive capacity are very good as far as the overall binary classification (Y/N) is concerned. The transfer of the test method to the participating laboratories was successful.

Where directly compared, performance against human data and LLNA data was comparable.

15.2 Extent to which study conclusions are justified by the study results alone

The scientific and regulatory rational of the method was clearly formulated. The study design was reasonable.

The ESAC WG agrees with the VMT that the WLR and BLR are very good, however a lower reproducibility was observed with one of the naïve laboratories. This may highlight the importance of proper training. Since proper training opportunities were provided by the test developer, a lower reproducibility may also indicate additional challenges for a naïve laboratory trying to apply the method.

Based on the analyses data published in the statistical report and its amendments submitted to EURL ECVAM (Attachments 18b, 18m and 18n), the predictive capacity is good as far as the overall binary classification (Y/N) is concerned. The assay seems to have a good sensitivity perhaps at the expense of specificity. Overall, the test method's accuracy is comparable with accepted *in vitro* methods.

The prediction model was found to be complex and the documentation was insufficiently clear to enable proper understanding of the conclusions drawn by the test submitter from INC results. The

ESAC WG has the concern that the 'six rules' used to resolve INC outcomes were fitting the outliers without having a reasonable biological and/or chemical rational. There was no evidence of the predictive value beyond the INC data set used to train the rules. The data obtained in the validation study suggest that the 'six rules' have limited accuracy (sensitivity: 90-100 %; specificity: 33 %) and have no added value to a simple decision where all INC results are considered positive. Indeed, considering all INC results as sensitiser by default instead of applying the 'six rules' would slightly decrease the number of false negatives with only an equally small increase of false positives.

Where directly compared, performance against human data and LLNA data appears comparable.

15.3 Extent to which conclusions are plausible in the context of existing information

The ESAC WG was not fully convinced about the explanations for the false positive and false negative results. For example, some of these explanations were not consistent as they appeared to be laboratory specific.

All data taken together, a broad range of potency and chemical reactivity classes was covered. Analysis of these data revealed that outliers did not correspond with a particular potency or chemical reactivity class. Pre-haptens were correctly identified. Overall the presented evidence seems to indicate that volatile substances or poorly soluble substances may require additional attention in the technology transfer and training phase, as improper handling of such substances during exposure could affect the reproducibility of the results rather than the accuracy.

The prediction model was made complex by the introduction of the 'six rules' to be applied on INC results. The ESAC WG has concerns about the rules, as the data do not provide evidence supporting any predictive value. The high sensitivity (90-100 %) and very low specificity (33 %) of the 'six rules' per se indicates that the rules have no added value to a simple decision where all INC results are considered positive.

The provided data suggest that the U-SENS™ test method may be useful as a component of an IATA for skin sensitisation or as a tool for screening and early decision-making during product development, for the assessment of chemicals belonging to the applicability domain defined in the validation study. The major concern of the ESAC WG is the complexity of the prediction model and the uncertainty of the value of the 'six rules' for deciding on INC results. The ESAC WG considers this an obstacle for regulatory readiness.

16. Recommendations

16.1 General recommendations

1. Use 'U-SENS' consistently in order to make it transparent which test has been validated.

16.2 Specific recommendations (e.g. concerning improvement of SOPs)

1. Upon request from the ESAC WG, the test developer provided a detailed flow chart showing how the decision about a substance has to be made. It is suggested to make such a flow chart an integrated part of the protocol (SOP).

2. On the basis of further analysis showing that the 'six rules' appear to change INC results to 'sensitisers' in the majority of cases, the test developers are recommended to consider removing these 'six rules' from the prediction model and to consider any inconclusive result as positive by default. This would significantly simplify the prediction model and would allow standardising the

maximum number of runs per experiment to three. Considering the validation dataset, modifying the prediction model and the maximum number of runs as suggested above would have little impact on the predictive capacity of the U-SENS™ test method and could lead to a decrease in testing requirements of about 8 %.

17. References

Aptula AO, Roberts DW (2006). Mechanistic applicability domains for non-animal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem. Res. Toxicol.* 19: 1097-1105.

Basketter DA, Alépée N, Ashikaga T, Barroso J, Gilmour N, Goebel C, Hibatallah J, Hoffmann S, Kern P, Martinozzi-Teissier S, Maxwell G, Reisinger K, Sakaguchi H, Schepky A, Tailhardat M, Templier M.(2014). Categorization of chemicals according to their relative human skin sensitization potency. *Dermatitis* 25: 11-21.

EC (2008). Regulation (EC) No. 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Official Journal of the European Union, L353.

OECD (2010). Skin Sensitization: Local Lymph Node Assay. OECD Guideline for the Testing of Chemicals No. 429, OECD, Paris.

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, OECD, Paris.

OECD (2016). *In Vitro* Skin Sensitisation: human Cell Line Activation Test (h-CLAT). OECD Guideline for the Testing of Chemicals No. 442E, OECD, Paris.

UN (2015). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Sixth revised edition, UN New York and Geneva.

LA-NA-28178-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub
EU Science Hub - Joint Research Centre
Joint Research Centre
EU Science Hub

Publications Office

doi:10.2787/815737

ISBN 978-92-79-63200-6