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

27-Oct-2016

English - Or. English

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

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

JT03403926

ENV/JM/MONO(2016)29/ANN2
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing & Assessment

No. 256

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

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris, 2016

ABOUT THE OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in 11 different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

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

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

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

or contact:

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

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

Funded by the

TABLE OF CONTENTS

ANNEX II: INFORMATION SOURCES USED WITHIN THE CASE STUDIES

1. Direct Peptide Reactivity Assay (DPRA) - OECD TG 442C

Name of the information source	Direct Peptide Reactivity Assay (DPRA) - OECD TG 442C
Mechanistic basis including AOP coverage	The DPRA measures in chemico the binding of test chemicals to model synthetic peptides containing either lysine or cysteine. Within the skin sensitisation AOP the covalent binding of electrophilic chemicals with nucleophilic sites of amino acids in skin proteins is postulated to be the molecular initiating event (MIE) (i.e. key event 1 – protein binding reactions) leading to skin sensitisation. In skin proteins many amino acids contain electron-rich groups capable of reacting with sensitisers. Lysine and cysteine are those most often quoted but others such as arginine, histidine, methionine and tyrosine can react with electrophilic chemicals.
Description	Solutions of cysteine and lysine containing synthetic heptapeptides are incubated with a 100mM solution of the test chemical at 1:10 and 1:50 ratio respectively for 24-hours at room temperature. At the end of the incubation period unreacted peptide concentration is measured by high-performance liquid chromatography (HPLC) with gradient elution and UV detection at 220 nm. Each test chemical is tested at a single concentration in triplicate. The positive control cinnamic aldehydeis tested concurrently and the positive control results are used as one of the run acceptance criteria. Solvent is used as the negative control. From the determined concentration of unreacted cysteine- and lysine-containing peptides the percent peptide depletion, relative to unreacted peptide control samples is calculated (OECD, TG 442C).
Response(s) measured	Direct peptide reactivity, expressed as: % cysteine depletion % lysine depletion.
Prediction model	The mean cysteine and lysine peptide percent depletion value of 6.38 is used to discriminate between peptide non-reactive and peptide reactive chemicals (OECD TG 442C). Within structured approaches to data integration the % cysteine and % lysine depletion values or the % of unreacted peptides are often directly used as input parameters instead of the reactivity prediction derived as described above.
Metabolic competence (if applicable)	No metabolic competent system.

Status of development, standardisation, validation	Evaluated in a validation study for reliability (EURL ECVAM, 2013) and officially adopted test method (OECD TG 442C).
Technical limitations and limitations with regard to predictivity	*Technical limitations* - The method is not suitable for testing highly hydrophobic chemicals. - Test chemicals must be stable under the test conditions (e.g. DPRA uses highly alkaline conditions for lysine reactivity). - Test chemicals having the same retention time as the cysteine and the lysine peptides provide inconclusive results. - The method cannot be used for the testing of complex mixtures of unknown composition or for substances of unknown or variable composition, complex reaction products or biological materials (i.e UVCB substances) due to the defined molar ratios of test chemicals and peptides. *Limitations with regard to predictivity* - Test chemicals requiring to be metabolically activated to act as sensitisers (pro-haptens) cannot be detected as being reactive in the DPRA. - Metals are considered outside the applicability of the DPRA since they react with proteins with mechanisms different than covalent binding.
Weaknesses and Strengths	*Strengths* - Evaluated in a validation study for reliability (EURL ECVAM, 2013) and detailed protocol publicly available at: http://ecvam-dbalm.jrc.ec.europa.eu/ (DB-ALM protocol N°154). - Large dataset (N>150) publicly available (e.g. Natsch et al., 2013). - Implemented and in use by several industry laboratories. - Relatively cheap and easy to perform by personnel experienced with HPLC analysis. *Weaknesses* - Since a single concentration of the test chemical is assessed at a single time point, reaction kinetic information cannot be derived. - Evaluation of the reactivity of the electrophile is limited to cysteine and lysine. Test chemicals with selective reactivity towards other nucleophiles may not be detected by the assay. - Test chemicals requiring to be abiotically activated to act as sensitisers (pre-haptens) may not always be correctly identified.
Reliability (within and between laboratories) (if applicable)	The reproducibility in predictions (reactive/non-reactive) that can be expected from the method is in the order of 85% within-laboratories an 80% between-laboratories (OECD TG 442C).

Predictive capacity (if applicable)	Results generated in the validation study (EURL ECVAM, 2013) and published studies (Natsch et al., 2013) overall indicate that the accuracy of the DPRA in discriminating sensitisers (i.e. UN GHS cat. 1) from non-sensitisers is 80% (N=157) with a sensitivity of 80% (88/109) and specificity of 77% (37/48) when compared to LLNA results. False negative predictions in the DPRA generally concern pro-haptens and chemicals showing a low to moderate sensitisation potency in vivo. It has to be noted that the DPRA is not proposed as a stand-alone replacement method and therefore the predictive performance values are reported for indication only.
Proprietary aspects	The test method does not have proprietary elements.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within a Defined Approach. For the purpose of certain regulations a positive DPRA prediction can be used to classify a chemical into UN GHS category 1. DPRA data can be used within a Defined Approach to support potency prediction.
Potential role within the Defined Approach	See specific descriptions of the role of the information source in case studies I, II, III, IV, IX, X and XI.

2. KeratinoSens™ - OECD TG 442D

Name of the information source	KeratinoSens™ - OECD TG 442D
Mechanistic basis including AOP coverage	The KeratinoSens™ test method addresses one of the biological mechanisms described under key event 2 (events in keratinocytes) of the skin sensitisation AOP by measuring the activation of the Keap1-Nrf2-ARE pathway. The Keap1-Nrf2-ARE regulatory pathway is reported to be a major regulator of cyto-protective responses to electrophile and oxidative stress by controlling the expression of detoxification, antioxidant and stress response enzymes and proteins. Small electrophilic substances such as skin sensitisers can act on the sensor protein Keap1 (Kelch-like ECH-associated protein 1), by e.g., covalent modification of its cysteine residue, resulting in its dissociation from the transcription factor Nrf2 (nuclear factor-erythroid 12-related factor 2). The dissociated Nrf2 can then activate ARE-dependent genes such as those coding for phase II detoxifying enzymes. The KeratinoSens™ is performed using an immortalised adherent cell line derived from HaCaT human keratinocytes stably transfected with a selectable plasmid containing the luciferase gene under the transcriptional control of a constitutive promoter fused with an ARE element. The quantitative measurement by luminescence detection of the luciferase gene induction is used as an indicator of the activity of the Nrf2 transcription factor in cells following exposure to electrophilic test chemicals.
Description	Cells are exposed to 12 concentrations of the test chemical for 48 hours. At the end of the incubation period quantification of luciferase gene induction is performed by luminescence analysis. Each test chemical is tested in three parallel replicate plates and a fourth replicate plate is used for cytotoxicity determination (with the MTT assay). The positive control cinnamic aldehyde is tested concurrently and the positive control results are used as one of the run acceptance criteria. Solvent is used as the negative control. Test chemicals are considered positive in the KeratinoSens™ if they induce a statistically significant induction of the luciferase gene above a given threshold (i.e. >1.5 fold) over solvent negative controls, at a concentration which does not significantly affect cell viability and below the concentration of 1000 µM.
Response(s) measured	− EC1.5 corresponding to the concentration needed for a statistically significant luciferase gene induction above the 1.5-fold threshold. − Imax corresponding to the maximal fold induction of the luciferase gene over solvent control. − % cytotoxicity.

Prediction model	Test chemicals are identified as potential skin sensitisers if the I_{max} is statistically significantly higher than 1.5-fold as compared to the basal luciferase activity and the EC 1.5 value is below 1000 µM in at least two out of the three repetitions. In addition at the lowest concentration with a gene induction above 1.5 fold the cellular viability should be above 70% and the dose-response for luciferase induction should be similar between the repetitions (OECD TG 442D).
Metabolic competence (if applicable)	Limited metabolic capacities.
Status of development, standardisation, validation	Evaluated in a validation study for reliability (EURL ECVAM, 2014) and officially adopted test method (OECD TG 442D).
Technical limitations and limitations with regard to predictivity	*Technical limitations* − The test method is not applicable to the testing of chemicals which are not soluble or do not form a stable dispersion either in water or DMSO. − Highly cytotoxic chemicals cannot always be reliably assessed. − Test chemicals that strongly interfere with the luciferase enzyme cannot be reliably tested. *Limitations with regard to predictivity* − Test chemicals with cLogP above 7 fall outside the known applicability of the method.
Weaknesses and Strengths	Strengths: − Validated method for reliability (EURL ECVAM, 2014) and detailed protocol publicly available at: http://ecvam-dbalm.jrc.ec.europa.eu/ (DB-ALM protocol N°155). − Large dataset (N> 300) publicly available (e.g. Natsch et al., 2013; EURL ECVAM, 2014). − Provides dose-response information. − Easy to perform. − Implemented and in use by several industry laboratories. Weaknesses: − Because of the limited metabolic capacity of the cell line and the experimental conditions, test chemicals requiring enzymatic activation (pro-haptens) or requiring autoxidation to act as sensitisers (pre-haptens) may provide negative predictions. − Substances with an exclusive reactivity towards lysine-residues are likely to give negative results in the KeratinoSensTM. − Test chemical stressors other than electrophilic chemicals may activate the Keap1-Nrf2-ARE pathway leading to false positive predictions in the KeratinoSensTM.

Reliability (within and between laboratories) (if applicable)	The reproducibility in predictions (positive/negative that can be expected from the method is in the order of 85% within- and between-laboratories (OECD TG 442D).
Predictive capacity (if applicable)	The accuracy of the KeratinoSens™ (EURL ECVAM, 2014) in discriminating sensitisers (i.e. UN GHS cat. 1) from non-sensitisers is 77% (N=201) with a sensitivity of 78% (71/91) and a specificity of 76% (84/110) when compared to LLNA results. False negative predictions in the KeratinoSens™ generally concern pro-haptens or chemicals showing low to moderate skin sensitisation potency in vivo. It has to be noted that the KeratinoSens™ is not proposed as a stand-alone replacement method and therefore the predictive performance values are reported for indication only.
Proprietary aspects	The KeratinoSens™ is a proprietary method for which a license agreement is needed. It is now widely offered by CRO's. The plasmid encoding for the luciferase gene is proprietary to Promega, but a license for use in sensitisation assessment is included in the MTA of KeratinoSens™.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitizsng chemicals within a Defined Approach. For the purpose of certain regulations KeratinoSens™ prediction can be used to classify a chemical into UN GHS category 1. KeratinoSens™ data can be used within a Defined Approach to support potency prediction.
Potential role within the Defined Approach	See specific descriptions of the role of the information source in case studies I, II, III, IV, VII and IX.

3. LuSens assay (Ramirez et al., 2016)
ARE-Nrf2-Luciferase Test Method (OECD TG 442D)

Name of the information source	LuSens assay (Ramirez et al., 2016); ARE-Nrf2-Luciferase Test Method (OECD TG 442D).
Mechanistic basis including AOP coverage	The LuSens test method addresses one of the biological mechanisms described under key event 2 (events in keratinocytes) of the skin sensitisation AOP by measuring the activation of the Keap1-Nrf2-ARE pathway. It employs the reporter gene for luciferase which is placed under the control of an antioxidant response element (ARE) and hence monitors Nrf-2 transcription factor activity. Keratinocytes respond to electrophilic haptens in that the modification of the cysteine groups of the keap1 protein, which lies associated with Nrf2 in the cytoplasm, leads to the dissociation of Nrf2 from keap1 and its translocation to the nucleus. Nrf2 then binds to the ARE response elements and activates the transcription of various downstream protective genes, e.g. glutathione (GSH). The keap1 protein therefore constitutes an intracellular sensor protein for cysteine reactive substances. The LuSens assay addresses this pathway to identify sensitisers by coupling the ARE-response element to a luciferase gene. The luciferase activity triggered by a substance is then used as a measure for the sensitisation potential.
Description	The ARE-Nrf2 luciferase-based test method, LuSens, is an assay utilising the same principle as the KeratinoSens™ assay described in OECD 442D utilises an immortalised human keratinocyte-based cell line stably transfected with the reporter gene construct. The reporter gene construct is composed of the luciferase reporter gene under the control of a rat ARE element. The luciferase signal reflects the activation of endogenous Nrf2 dependent genes. The quantitative measurement by luminescence detection of luciferase gene induction is used as an indicator of the activity of the Nrf2 transcription factor in cells following exposure to electrophilic test chemicals. Cells are exposed to series of concentrations of the test chemical for 48 hours. The assay comprises at least two concordant or a maximum of three independent repetitions in total. In a valid repetition (i.e. meeting all acceptance criteria), sensitising potential of the substance is indicated if the luciferase activity equals or exceeds a 1.5 fold induction compared to the vehicle control at concentrations that do not reduce cell viability to more than 70%.

Response(s) measured	A test compound is considered to have sensitising potential when the luciferase induction is above or equal to 1.5 fold compared to the vehicle control in 2 (or more than) consecutive non-cytotoxic tested concentrations whereby at least three tested concentrations must be non-cytotoxic.
Prediction model	For the assessment of the predictive capacity of the LuSens assay, the data obtained from the in vitro assay were compared to human or LLNA data from the literature using Cooper statistics. From this analysis the following predictivity values were calculated: Sensitivity of 83% or 74%, specificity of 81% or 74% and an overall accuracy of 83% or 74% when compared to human or LLNA data, respectively (Ramirez et al., 2014; 2016).
Metabolic competence (if applicable)	Limited, e.g. similar enzyme activities as primary keratinocytes are observed for FMO, ADH, ALDH and NAT1 but not for UGT (Fabian et al., 2013).
Status of development, standardization, validation	An intra- and interlaboratory study was conducted and the LuSens method was submitted to ECVAM early 2015 for evaluation and has now progressed to the ESAC review stage (status April 2016). The results of the study were published in Ramirez et al. 2016. The study indicates a very good reproducibility of the assay as tested by laboratories from different parts of the world. The principles of the method are described in OECD TG 442D. Studies have demonstrated that this method can be used interchangeably with the KeratinoSens™ assay in integrated testing strategies (ITS), e.g. the 2 out of 3 WoE ITS (Urbisch et al., 2015). When comparing accuracies for the set of 69 substances for which data was available for both methods, use of the LuSens/KeratinoSens™ assays in the a '2 out of 3' approach with DPRA and h-CLAT data resulted in an accuracy of 83/85% or 93%/91%, when comparing the predictions to LLNA or human data, respectively.
Technical limitations and limitations with regard to predictivity	- As the LuSens assay assesses the activation of the ARE dependent gene expression in keratinocytes by modification of a cysteine in the Nrf2 protein, some substances with an exclusive reactivity towards lysine-residues may give negative results, for instance the acyl transfer agents (Urbisch et al., 2015). - As is the case with most cell-based methods, solubility and cytotoxicity of the substance can limit the applicability as the cells are cultured in aqueous medium. - The metabolic capacity of the cells which is required to activate certain pro-haptens is not always identical to the metabolic capacity of native skin. - Chemical stressors other than electrophilic chemicals may activate the Keap1-Nrf2-ARE pathway leading to false positive predictions in the LuSens assay.

	- Substances which interfere with the detection systems, e.g. luciferase, may lead to false predictions.
Weaknesses and Strengths	Strengths - Published data on 74 chemicals (Ramirez et al., 2014 and Urbisch et al., 2015). - Very good reproducibility in all laboratories participating in the validation study (Ramirez et al., 2016). - Applicable to chemicals covering for testing a large range of chemicals inluding ketones, aldehydes, and aromatic compounds, physico-chemical properties, and are that are used in a variety of application fields (e.g. fragrances, preservatives, solvents) . - Nonanimal test. - Interlaboratory validation conducted using the performance standards of the OECD TG 442D. The study indicates high robustness of the method, showing 100% within and between laboratory reproducibility - Gives dose-response information. - The LuSens cell-line can be readily obtained by laboratories that would like to perform the assay. - A detailed protocol is publicly available (Ramirez et al., 2014) and a training video was produced in Q2 2015 and is freely available. Weaknesses - Because of the limited metabolic capacity of the cell line and the experimental conditions, test chemicals requiring enzymatic activation (pro-haptens) or requiring autoxidation to act as sensitisers (pre-haptens) may provide negative predictions in the LuSens cell-line. - Potency not yet covered. - (see also technical limitations above).
Reliability (within and between laboratories) (if applicable)	The validation study showed a very good within and between laboratory reproducibility of 100% and an accuracy of over 80% to identify skin sensitisers.
Predictive capacity (if applicable)	Results generated in the in house-validation study (Bauch et al., 2012) indicate that the accuracy of the mMUSST in discriminating sensitisers (i.e., UN GHS Cat. 1) from non-sensitisers is 74% and 86% with a sensitivity of 64% and 75% and specificity of 94% and 100% when compared to LLNA and human data, respectively. The extended data set (Urbisch et al., 2015) results in 75% or 84% accuracy, in 68% and 70% sensitivity, and 92% and 100% specificity when compared to LLNA and human data, respectively.
Proprietary aspects	Use of the luciferase reporter gene plasmid is subject to a license agreement with Promega, which will be readily granted for use of the LuSens assay.

Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals for classification and labelling purposes such as GHS. For the purpose of certain regulations (e.g. for read-across approaches) a positive LuSens prediction can be used to classify a chemical into UN GHS category 1. However, given the complexity of the sensitisation process, a combination of tests should be used to achieve reliable predictions of the skin sensitisation potential of a substance.
Potential role within the Defined Approach (see case study I)	Contributes to hazard prediction for classification for GHS and/or REACH in the context of a weight of evidence and/or data integration approach for hazard identification. In this study, the method was used to address key events in the AOP - based "2 out of 3" integrated testing strategy (ITS) approach to skin hazard identification ("2 out of 3 – Sens ITS"; BASF).

4. Human Cell Line Activation Test (h-CLAT) - OECD TG 442E

Name of the information source	Human Cell Line Activation Test (h-CLAT)
Mechanistic basis including AOP coverage	The h-CLAT quantifies in vitro changes in the expression of the CD86 and CD54 membrane phenotypic markers in a human monocytic leukemia cell line (THP-1 cells). THP-1 cells are monocyte-derived cells that have shown to produce DC-like responses following exposure to skin sensitising chemicals, including upregulation of surface markers (e.g. CD86 and CD54) and cytokine production (e.g. TNF-α). The CD86 (a co-stimulatory molecule) and the CD54 (an adhesion molecule) are upregulated in activated Dendritic Cells (DC) and play a critical role in DC presentation of antigens to T cells (T-cell priming). By studying the potential of test chemicals to up-regulate markers of DC activation, the h-CLAT generates information addressing key event 3 (dendritic cell activation) of the skin sensitisation AOP.
Description	Qualified THP-1 cells are exposed for 24 hours to eight serial concentrations of test chemicals selected on the basis of a predetermined CV75 (concentration of test chemical yielding 75% cells survival). At the end of the incubation period, cells are stained with FITC-labelled anti-CD86, anti-CD54 and mouse IgG1 antibodies (for measurement of non-specific background signal). Changes of CD86 and CD54 surface markers expression are measured by flow cytometry analysis. Each chemical is tested in singlicate in at least two independent runs to derive a positive or negative prediction. The positive control 2,4-dinitrochlorobenzene (DNCB) is tested concurrently at a single concentration yielding approximately 70-90% of cell viability and positive control's results are used as one of the run acceptance criteria. Solvent is used as the negative control. Cytotoxicity is measured in parallel (with propidium iodide staining). The calculated relative fluorescence intensity (RFI) is used as indicator of CD86 and CD54 expression.
Response(s) measured	CD86 relative fluorescence intensity. CD54 relative fluorescence intensity. % cell viability.
Prediction model	An h-CLAT prediction is considered positive if: the RFI of CD86 is equal to or greater than 150% at any tested dose (with cell viability ≥ 50%) in at least two independent runs or if the RFI of CD54 is equal to or greater than 200% at any tested dose (with cell viability ≥ 50%) in at least two independent runs or the RFIs of both markers exceed the respective thresholds at any tested dose (with cell viability ≥ 50%) in at least two independent runs. For test chemicals predicted as positives, two Effective Concentrations (EC) values, the EC150 for CD86 and EC200 for CD54, i.e. the concentration at which the test chemicals induced a RFI of 150 or 200, can be calculated.

Metabolic competence (if applicable)	Limited metabolic capacities (Fabian et al., 2013).
Status of development, standardisation, validation	Evaluated in a validation study for reliability (EURL ECVAM, 2015) and officially adopted test method (OECD TG 442E).
Technical limitations and limitations with regard to predictivity	*Technical limitations:* − The method is not applicable to the testing of chemicals which are not soluble or do not form a stable dispersion in a solvent compatible with the experimental system. − Highly cytotoxic chemicals cannot be tested. − Strong fluorescent test chemicals emitting at the same wavelength as FITC may interfere with the flow-cytometry light-signal acquisition. *Limitations with regard to predictivity* − Test chemicals with a Log Kow of greater than 3.5 tend to produce false negative results. Negative results with these test chemicals should be considered as inconclusive.
Weaknesses and Strengths	*Strengths* − Validated method for reliability (EURL ECVAM, 2015) and detailed protocol publicly available at: http://ecvam-dbalm.jrc.ec.europa.eu/ (DB-ALM protocol N°158). − Large dataset (N>140) publicly available (e.g. Takenouchi et al., 2013). − Provides dose-response information. − Implemented and in use by several industry laboratories. *Weaknesses* − Because of the limited metabolic capacity of the cell line and the experimental conditions, test chemicals requiring enzymatic bioactivation (pro-haptens) or autoxidation (pre-haptens) to induce sensitisation may produce false negative results. − Need of expensive instruments.
Reliability (within and between laboratories) (if applicable)	The reproducibility in predictions (positive/negative) that can be expected from the method is in the order of 80% within- and between-laboratories (EURL ECVAM, 2015).
Predictive capacity (if applicable)	Results generated in the validation study (EURL ECVAM, 2015) and published studies (Takenouchi et al., 2013) overall indicate that the accuracy of the h-CLAT in discriminating sensitisers (i.e. UN GHS cat. 1) from non-sensitisers is 85% (N=142) with a sensitivity of 93% (94/101) and a specificity of 66% (27/41) when compared to LLNA results. Published data indicate and accuracy of 83% (N=66) in predicting responses in humans (Nukada et al., 2011). The relatively low rate of false negative predictions in the h-CLAT generally concern pro-haptens or chemicals showing low to moderate skin sensitisation potency in vivo. It

	has to be noted that the h-CLAT is not proposed as a stand-alone replacement method and therefore the predictive performance values are reported for indication only.
Proprietary aspects	The test method has intellectual property rights protected by Patent N. 4270702 only in Japan.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within a Defined Approach. For the purpose of certain regulations a positive h-CLAT prediction can be used to classify a chemical into UN GHS category 1. h-CLAT data can be used within a Defined Approach to support potency prediction.
Potential role within the Defined Approach	See specific descriptions of the role of the information source in case studies I, II, III, V, VIII IX, X and XI.

5. Modified Myeloid U937 Skin Sensitisation Test (mMUSST)

Name of the information source	Modified Myeloid U937 Skin Sensitisation Test (mMUSST)
Mechanistic basis including AOP coverage	The modified Myeloid U937 Skin Sensitisation Test (mMUSST) addresses the third key event, namely dendritic cell activation, of the skin sensitisation AOP by quantifying changes in the expression of cell surface markers associated with dendritic cell activation (i.e. CD86) following exposure to the test substance. CD 86 is a costimulatory factor which is upregulated in mature DCs and plays a pivotal role in antigen presentation and subsequent T cell priming.
Description	The mMUSST is performed using the human myeloid cell line U937 as a surrogate for dendritic cells (Ade et al., 2006). The change in the expression of the cell surface marker CD86, which is indicative for DC activation, is measured by flow cytometry following 48 hours of exposure to the test substance. Differences in the measured expression levels of CD86 between the vehicle control and the test substance are then used to support the discrimination between skin sensitisers and non-sensitisers.
Response(s) measured	Expression level of CD86 (in relation to concurrent solvent control); a substance is considered to be a sensitiser if CD86 expression is increased by 1.2 fold at test substance concentrations with viabilities exceeding 70% compared to the vehicle control.
Prediction model	Expression level of CD86 (in relation to concurrent solvent control); a substance is considered to be a sensitiser if CD86 expression is increased by 1.2 fold at test substance concentrations with viabilities exceeding 70% compared to the vehicle control (Bauch et al., 2012).
Metabolic competence (if applicable)	Limited (Fabian et al., 2013).
Status of development, standardization, validation	In house validation; over 65 chemicals tested and compared to human and/or LLNA data.
Technical limitations and limitations with regard to predictivity	- Highly cytotoxic chemicals or chemicals that interfere with the detection systems (e.g. flow cytometric analysis) cannot always be reliably tested. - Due to the aqueous nature of the cell medium, solubility issues can occur when testing lipophilic substances. - Applicable to test chemicals that are soluble or that form a homogenous suspensions.

Weaknesses and Strengths	Strengths - Published data on 65 chemicals (Urbisch et al., 2015). - Applicable to chemicals covering a variety of organic functional groups, reaction mechanisms, skin sensitisation potency (as determined in *in vivo* studies) and physico-chemical properties. - Gives dose-response information. - Nonanimal test. Weaknesses - Because of the limited metabolic capacity of the cell line and the experimental conditions, test chemicals requiring enzymatic activation (pro-haptens) or requiring autoxidation to act as sensitisers (pre-haptens) may provide negative predictions in U937 cells. - Use for potency not yet evaluated. - (see also technical limitations above).
Reliability (within and between laboratories) (if applicable)	The within-laboratory reproducibility during the initial implementation phase of the test method was determined to be 76%; the interlaboratory reproducibility has not yet been determined.
Predictive capacity (if applicable)	Results generated in the in house-validation study (Bauch et al., 2012) indicate that the accuracy of the mMUSST in discriminating sensitisers (i.e., UN GHS Cat. 1) from non-sensitisers is 74% and 86% with a sensitivity of 64% and 75% and specificity of 94% and 100% when compared to LLNA and human data, respectively. The extended data set (Urbisch et al., 2015) results in 75% or 84% accuracy, in 68% and 70% sensitivity, and 92% and 100% specificity when compared to LLNA and human data, respectively.
Proprietary aspects	The test method does not have proprietary elements. Restrictions to cell line use for commercial purposes apply. The method is described in Bauch et al. 2012.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals for classification and labelling purposes such as GHS. For the purpose of certain regulations (e.g. for read-across approaches) a positive mMUSST prediction can be used to classify a chemical into UN GHS category 1. However, given the complexity of the sensitisation process, a combination of tests should be used to achieve reliable predictions of the skin sensitisation potential of a substance.
Potential role within the Defined Approach (see case study I)	Contribute to hazard prediction for classification for GHS and/or REACH in the context of a weight of evidence and/or data integration approach for hazard identification. In this study, the method was used to address key events in the AOP - based "2 out of 3" integrated testing strategy (ITS) approach to skin hazard identification ("2 out of 3 – Sens ITS"; BASF).

6. HaCaT gene signature

Name of the information source	HaCaT gene signature
Mechanistic basis including AOP coverage	The HaCaT gene signature test method addresses one of the biological mechanisms described under key event 2 (events in keratinocytes) of the skin sensitisation AOP by measuring the activation of a predictive gene signature consisting of 10 genes. The results of a DNA microarray in the HaCaT cell line was used to select the most predictive genes by using three classifier algorithms: Random Forest (RF), Support Vector Machine (SVM) and PAM-R using a leave-one-compound-out cross-validation. The most common genes across the algorithms were selected resulting in a signature of 10 genes (Table 1). These genes not only are highly predictive but most of them can be linked to the pathways that are significantly regulated by skin sensitisers, such as pathways involved in innate and inflammatory responses and the Keap1-Nrf2-ARE regulatory pathway. The latter is reported to be a major regulator of cyto-protective responses to electrophile and oxidative stress by controlling the expression of detoxification, antioxidant and stress response enzymes and proteins. Small electrophilic substances such as skin sensitisers can act on the sensor protein Keap1 (Kelch-like ECH-associated protein 1), by e.g., covalent modification of its cysteine residue, resulting in its dissociation from the transcription factor Nrf2 (nuclear factor-erythroid 12-related factor 2). The dissociated Nrf2 can then activate ARE-dependent genes such as those coding for phase II detoxifying enzymes.
Description	The test method is performed using an immortalised adherent cell line derived from HaCaT human keratinocytes. Cells are exposed for 4 hours to a concentration that causes a 20% decrease in cell viability (CV80). At the end of the incubation period cells are lysed and RNA is isolated. The samples were analysed using RT-PCR assays for the 10 biomarker genes and the housekeeping gene HPRT1. Gene expression data was log2 transformed and normalised against the housekeeping gene.
Response(s) measured	− % cytotoxicity. − Gene expression of the biomarker genes.
Prediction model	The prediction model is based on the gene expression of the 10 genes. Classification is performed using the environment for statistical computing of R. The data obtained from the microarray study is used as a training set (Van der Veen et al., 2013) and is accessible at Array Express (http://www.ebi.ac.uk/arrayex- 192); accession number 943-MTAB-E. This training set is used to train the three different algorithms and classify: • Random forests (RF) is based on the creation of prediction trees. • Support Vector Machine (SVM) uses the radial kernel on scaled data, creates a separating hyper plane. • The Prediction analysis for Microarrays in R (PAM-R) uses shrunken centroids to classify samples. Each algorithm generates three predictions per sample. A substance is classified based on the prediction of the triplicate samples generated by the three algorithms, with a total of nine predictions. Majority voting was

	used when the prediction for the replicate samples was discordant. The substance is classified according to the prediction of the majority of the samples.
Metabolic competence (if applicable)	Limited metabolic capacities.
Status of development, standardisation, validation	This test method is an in-house model that has been validated in one independent experiment performed at the RIVM. The test method is not validated or transferred to other laboratories.
Technical limitations and limitations with regard to predictivity	*Technical limitations* − The test method is not applicable to the testing of chemicals which are not soluble or do not form a stable dispersion either in water or DMSO. − Highly cytotoxic chemicals cannot always be reliably assessed. *Limitations with regard to predictivity* − Test chemicals with cLogP above 7 fall outside the known applicability of the method. − Some chemicals that were misclassified in the LLNA (false-positive or false-negative) were wrongly predicted in this assay as well (e.g. SDS, nickel). However, maleic acid and triisobutylphosphate, respectively false-positive and false-negative in the LLNA, were correctly classified by the gene signature.
Weaknesses and Strengths	Strengths: − The gene signature consists of genes that are involved both in pathways that regulate stress responses as well as in inflammatory responses. As such, genes of the signature cover multiple pathways that are relevant to skin sensitisers. These genes were regulated after *ex vivo* exposure of fresh human skin to skin sensitisers, showing their relevance in humans as well (Van der Veen et al., 2015). Weaknesses: − Because of the limited metabolic capacity of the cell line and the experimental conditions, test chemicals requiring enzymatic activation (pro-haptens) or requiring autoxidation to act as sensitisers (pre-haptens) may provide negative predictions. − Test chemical stressors other than electrophilic chemicals may activate the Keap1-Nrf2-ARE pathway leading to false positive predictions (Van der Veen et al., 2013).
Reliability (within and between laboratories) (if applicable)	Reproducibility within and between laboratories has not been assessed.

Predictive capacity (if applicable)	The accuracy of the HaCaT gene signature in discriminating sensitisers (i.e. UN GHS cat. 1) from non-sensitisers is 90.2% (n=41), with a sensitivity of 100% (27/27) and a specificity of 71.4% (10/14). The Positive Predictive Value (PPV) was 79.8% and the Negative Predictive Value (NPV) 100% (Van der Veen, 2014). False-positive predictions concern substances that are known false-positives in the LLNA (Benzalkonium chloride, Sodium Dodecyl Sulfate, hexaethylene glycol monodecyl ether).
Proprietary aspects	Not applicable.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within a Defined Approach to measure key event 2. The assay does not provide data that can be used for potency assessment.
Potential role within the Defined Approach (see case study II)	HaCaT gene signature is used in Tier 2 to test the substances that are rated negative in Tier 1.

Table 1: Genes and function of the HaCaT gene signature. *Entrez Gene NCBI's database for gene-specific information (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene).

Entrez ID*	**Gene name**		**Function**
1789	DNA (cytosine-5-)-methyltransferase 3 beta	DNMT3b	DNA methylation
3162	Hemeoxygenase 1	HMOX1	Oxidative stress
8614	Stanniocalcin 2	STC2	
133	Adrenomedullin	ADM	
140809	Sulfiredoxin 1 homolog	SRXN1	
2353	FBJ murine osteosarcoma viral oncogene homolog	FOS	Development and inflammation
8061	FOS-like antigen 1	FosL1	
10181	RNA binding motif protein 5	RBM5	Alternative splicing
51755	Cyclin-dependent kinase 12	CDK12	
353322	Ankyrin repeat domain 37	ANKRD37	Unknown

7. Read-across for skin sensitisation using QSAR Toolbox V3.2

Name of the information source	Read-across for skin sensitisation using QSAR ToolboxV3.2
Mechanistic basis including AOP coverage	The read-across protocol uses skin penetration and metabolism information, as well as all key events of the AOP, because it uses the *in vivo* data from analogs to predict skin sensitisation hazard. The *in vivo* data include LLNA, which assesses the AOP up to Key Event 4, activation and proliferation of T-cells; and guinea pig and human tests, which include all the events in the AOP, including the adverse outcome.
Description	The read-across method starts with the MIE, Key Event 1, by assessing the target substance for protein binding alerts. If the target substance has no protein binding alerts, the protocol predicts auto-oxidation products and skin metabolites, which are then evaluated for protein binding alerts. If the products/metabolites have no protein binding alerts, the target substance is predicted to be a non-sensitiser. If the target substance or its products/metabolites have protein binding alerts, a group of similar chemicals with *in vivo* skin sensitisation data are identified. This group, which is similar to the target substance in structure and protein binding mechanism, is used to make a read-across prediction for the skin sensitisation outcome of the target substance.
Response(s) measured	Skin sensitiser or non-sensitiser.
Prediction model	The read-across algorithm uses the skin sensitisation outcome that appears most often for the five chemicals (in the group of similar chemicals) nearest the target substance, based on log K_{ow}, to predict the skin sensitisation hazard of the target substance.
Metabolic competence (if applicable)	If the target compound has no protein binding alerts, QSAR Toolbox is used to predict auto-oxidation products and skin metabolites.
Status of development, standardisation, validation	Read-across has not been standardized or validated, but is recommended as a method to fill toxicity data gaps in the assessment of chemical hazards. It is evaluated on a case-by-case basis for regulatory applications. It can be used as a stand-alone prediction or as part of a weight-of-evidence approach.
Technical limitations and limitations with regard to applicability	Read-across using QSAR Toolbox is not applicable to substances that have no associated chemical structure such as substances of unknown composition. Results for predicted auto-oxidation products or skin metabolites may rely on those that are not biologically important (i.e., the relative amounts of products/metabolites produced *in vivo* are unknown)

Weaknesses and Strengths	Weaknesses: − The read-across predictions depend on the availability of *in vivo* skin sensitisation data for mechanistic and structural analogs. − The responsibility for quality control of the *in vivo* data has been left to the data submitters; the quality control measures used during data generation are not known. − The read-across predictions depend heavily on the grouping of chemicals by similar mechanism of protein binding and similarity of structure and assumes that the categorization schemes offered in QSAR Toolbox are sufficient. Strengths: − Uses publicly available software that is supported by OECD. − The read-across prediction uses *in vivo* data, thus accounting for all of the processes in the AOP, including skin absorption and metabolism.
Reliability (within and between laboratories) (if applicable)	A preliminary evaluation of reproducibility using 10 substances and two analysts found that between-analyst reproducibility was 80%. The differences in outcomes were due to differences in the application of the protein binding alert system applied to subcategorise analogs to refine the category grouping. The protocol has been modified to clarify that only endpoint-specific protein binding alert system, protein binding alerts for skin sensitisation by OASIS v1.2, should be used to subcategorise analogs.
Predictive capacity (if applicable)	For the database of 120 chemicals used for the Integrated Decision Strategy for Skin Sensitisation Hazard, the performance of the read-across prediction, with respect to predicting LLNA results was: accuracy = 77% (92/120), sensitivity = 77% (67/87), and specificity = 76% (25/33) (Strickland et al., 2016).
Proprietary aspects	This read-across method uses publicly available software, QSAR Toolbox, which is supported by OECD.
Proposed regulatory use	Read-across can be used as a stand-alone prediction or as part of a weight-of-evidence approach for skin sensitisation hazard (ECHA, 2014).
Potential role within the Defined Approach (see case study V)	Accounts for skin absorption and metabolism as it identifies potential skin sensitisers or non-sensitisers. Of the information sources in this integrated decision strategy, a random forest analysis showed that the importance of the read-across prediction was just below that of the h-CLAT (Strickland et al., 2016).

8. TIMES-SS

Name of the information source	TIMES-SS
Mechanistic basis including AOP coverage	Chemical reactivity of xenobiotics (and their metabolies) with proteins can be predicted from their chemical structure as is the molecular initiating event of skin sensitisation and Key event 1 of the AOP.
Description	TIMES-SS is a software package to predict skin sensitisation.
Response(s) measured	i. Amount of protein-hapten adduct formation ii. Total Structural domain
Prediction model	Automatic prediction of the amount of protein-hapten adduct formation per mole of hapten.
Metabolic competence (if applicable)	*In silico* predicted metabolism and abiotic oxidation.
Status of information source development, standardisation, validation	Commercially available software, compliant with the OECD principles for QSAR validation (OECD, 2004).
Technical limitations and limitations with regard to applicability	A defined chemical structure is needed. Less reliable predictions for chemicals falling outside the applicability domain of the model. This is indicated by the output of the software in each prediction. However, our results show that the defined approach is not affected by the applicability domain of TIMES-SS.
Weaknesses and Strengths	Strengths: - Includes prediction of metabolism, indicates whether molecule is within applicability domain. High predictive capacity. - 100% reproducibility - Fast - No high expertise needed - Can be used on any computer Weakness: - Cannot calculate mixtures, metals, polymers, and natural products.
Reliability	Not applicable
Predictive capacity (if applicable)	According to Patlewicz et al. 2007, the skin sensitisation prediction of the model performs as shown below. However, the skin sensitisation prediction readout was not used in the defined approach, but the amount of protein-hapten. Accuracy (75%, 30/40) Sensitivity (56%, 9/16) Specificity (87.5%, 21/24) In our dataset, if we assigned a positive prediction to the chemicals predicted by TIMES to be reactive to proteins and viceversa, the predictive power of the "amount of protein-hapten" was the following: All comp. (269) comp. not in training set of TIMES(92). Accuracy= 87% 80% Sensitivity= 92% 86% Specificity= 78% 70%

Proprietary aspects	Need for a License; TIMES-SS may be replaced in the defined approach by an *in vitro*/*in chemico* assay that accounts for skin metabolism and protein binding.
Proposed regulatory use	- To support the discrimination between sensitising and non-sensitising chemicals within the defined approach. - The structural alerts also included in the readouts of the software package can contribute to classification of chemicals into mechanistic domains to support read-across.
Potential role within the Defined Approach (see case study VI)	- The main discriminatory node corresponds to a readout of TIMES-SS. The defined approach is mostly based on this descriptor.

9. DRAGON

Name of the information source	DRAGON
Mechanistic basis including AOP coverage	Not applicable
Description	DRAGON is a software package to calculate chemical descriptors.
Response(s) measured	i. **Dragon-Mor32s:** 3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) are derived from Infrared spectra simulation using a generalized scattering function. This descriptor corresponds to signal 32 weighted by l-state. ii. **Dragon-SpDiam_EA(bo):** Spectral diameter from edge adjacency matrix weighted by bond order. iii. **Dragon-O-056:** Presence of alcohol (-OH) groups. iv. **Dragon-Eig08_AEA(bo):** Eigenvalue n. 8 from augmented edge adjacency matrix weighted by bond order. v. **Dragon-HATS4e:** Leverage-weighted autocorrelation of lag 4 / weighted by Sanderson electronegativity. The GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptors are molecular descriptors derived from the Molecular Influence Matrix (MIM). vi. **Dragon-Ds**: D total accessibility index / weighted by I-state (WHIM descriptors are based on the statistical indices calculated on the projections of atoms along principal axes[18,19]. They are built in such a way as to capture relevant molecular 3D information regarding the molecular size, shape, symmetry and atom distribution with respect to invariant reference frames. The algorithm consists of performing a Principal Components Analysis on the centred Cartesian coordinates of a molecule by using a weighted covariance matrix obtained from different weighing schemes for the atoms). l-state the Electro topological State S_i of the i^{th} atom in a molecule, also called the E-state index gives information related to the electronic and topological state of the atom in the molecule. vii. **Dragon-H-052:** H attached to C(sp3) with 1 heteroatom attached to the next C. viii. **Dragon-HATS6i:** Leverage-weighted autocorrelation of lag 6 / weighted by ionization potential. The GETAWAY (GEometry, Topology, and Atom-Weights Assembly) descriptors are molecular descriptors derived from the Molecular Influence Matrix (MIM). ix. **Dragon-Mor24u:** 3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) are derived from Infrared spectra simulation using a generalized scattering function. This descriptor corresponds to signal 24 un-weighted.
Prediction model	Automatic prediction of the descriptors.

Metabolic competence (if applicable)	No.
Status of information source development, standardisation, validation	Not applicable.
Technical limitations and limitations with regard to applicability	A defined chemical structure is needed
Weaknesses and Strengths	Strengths: - 100% reproducibility - Fast - No high expertise needed - Can be used on any computer Weakness: - Cannot calculate mixtures, metals, polymers, and natural products. - The values of some descriptors can depend on the optimization process of the 3D structure of the chemical compounds
Reliability	Not applicable
Predictive capacity (if applicable)	Not applicable
Proprietary aspects	Need for a License. Some descriptors can be calculated free at www.vcclab.org.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within the defined approach.
Potential role within the Defined Approach (see case study VI)	The different descriptors confirm and modify the prediction of Key event 1. The weight of DRAGON descriptors on the defined approach is relatively low.

10. LC-MS and fluorescence-based kinetic peptide reactivity assay

Name of the information source	LC-MS and fluorescence-based kinetic peptide reactivity assay (Cor1C420-assay)
Mechanistic basis including AOP coverage	The Cor1C420-assay measures *in chemico* the binding of test chemicals to a model synthetic peptides containing both lysine and cysteine residues. Within the skin sensitisation AOP the covalent binding of electrophilic chemicals with nucleophilic sites of amino acids in skin proteins is postulated to be the molecular initiating event (MIE) (i.e. key event 1 – protein binding reactions) leading to skin sensitisation. In skin proteins many amino acids contain electron-rich groups capable of reacting with sensitisers, lysine and cysteine are those being most often quoted.
Description	In the Cor1C420-assay, solutions of peptide containing cysteine and lysine residues (0.1 mM; peptide Cor1C420, derived from a reactive hotspot in the human proteome) are incubated with a 1 mM solution of the test chemical for 24-hours at 36°C. Remaining concentration of the peptide following incubation is determined. Relative peptide concentration is measured by high-performance liquid chromatography (HPLC) with gradient elution and LC-MS detection. Percent peptide depletion is used as quantitative information. In parallel molecular information on formed adducts is collected to verify peptide modification and gain mechanistic information. For chemicals with a high reactivity at 24 h (> 50% peptide depletion), the assay is repeated at shorter incubation and lower test chemical concentration to determine the true reaction rate. These latter measurements are made with fluorescent derivatisation of the sulfhydryl group in the peptide, as this allows to stop the reaction at a precise time which is not possible in the HPLC assessment. Assay is described in detail in Natsch et al. 2008.
Response(s) measured	- Direct peptide reactivity, expressed as: % peptide depletion ate different time points. - Rate constant for peptide depletion ($min^{-1}mM^{-1}$). - Qualitative information whether depletion is due to covalent modification or peptide oxidation. - Molecular information on formed adducts (Mass of [peptide + bound (part of) hapten]).
Prediction model	For hazard assessment, chemicals are rated positive in the assay if covalent modification of the peptide can be observed. Within structured approaches to data integration the and for potency assessment, the continuous scale rate constant is entered into the DIP and no thresholds do apply.
Metabolic competence	No metabolic competent system.
Status of development, standardisation, validation	No validation studies performed. Fully standardized protocol published (Natsch et al., 2008). Good correlation of rate constant obtained with the Cor1C420 assay and the DPRA-Cys peptide in the (not validated) kinetic DPRA (Natsch et al., 2015).

Technical limitations and limitations with regard to applicability	*Technical limitations* − The method is not suitable for testing highly hydrophobic substances. − The method cannot be used for the testing of complex mixtures of unknown composition or for substances of unknown or variable composition, complex reaction products or biological materials (i.e UVCB substances) due to the defined molar ratios of test chemicals and peptides. *Limitations with regard to predictivity* − Substances requiring to be metabolically activated to act as sensitisers (pro-haptens) cannot be detected by the Cor1C420-assay because of the lack of a metabolic system. − Substances requiring to be oxidised to act as sensitisers (pre-haptens) are often, but not always, detected by the method. − Metals are considered outside the applicability of the Cor1_C420 assay since they react with proteins with mechanisms different than covalent binding.
Weaknesses and Strengths	Data for more than 300 chemicals in Natsch et al. 2015. Data available for these chemicals indicate the Cor1C420-assay is applicable to chemicals covering a variety of organic functional groups, reaction mechanisms, skin sensitisation potency (as determined in *in vivo* studies) and physico-chemical properties.
Reliability (within and between laboratories) (if applicable)	High intralaboratory reproducibility. Protocol is very close to the validated DPRA protocol (Different test peptide, lower concentration of chemicals and peptide for improved solubility, different detection system, but equal incubation conditions and equal assay principle; thus practical handling is equivalent to DPRA).
Predictive capacity (if applicable)	With a prediction model of only rating chemicals positive with direct adducts with the peptide, the method has a very high specificity, and a limited sensitivity. Improved sensitivity but reduced specificity is obtained based on depletion values.
Proprietary aspects	The test method does not have proprietary elements. The protocol is published (Natsch et al., 2008).
Proposed regulatory use	- To support the discrimination between sensitising and non-sensitising chemicals within a DIP or a Defined Approach. - The molecular information from adduct formation can contribute to classification of chemicals into mechanistic domains. - The kinetic rate constants are used in a DIP / Defined Approach to support potency prediction.
Potential role within the Defined Approach (see case study VII)	Useful for molecular characterization of MIE and generating quantitative kinetic data which can be used in Defined Approach and DIP for potency prediction.

11. Derek Nexus (version 2.0 from Lhasa Limited)

Name of the information source	Derek Nexus (version 2.0 from Lhasa Limited)
Mechanistic basis including AOP coverage	The skin sensitisation alerts that are given by Derek Nexus are mainly giving an indication of the reactivity potential/behavior of the tested chemical derived from its structure. Reactivity determines the capacity of the substance to modify/haptenize skin proteins, which is the molecular initiating event defined in the AOP (Langton et al., 2006)
Description	*In silico* knowledge-based toxicity alerting software comprising alerts on skin sensitisation.
Response(s)	Mechanistic alerts for Skin Sensitisation. Binary conclusions: Positive alert (=Probable, Plausible, Equivocal, doubted alerts) or Inconclusive (absence of alert).
Prediction model	Derek Nexus is a knowledge based expert system designed to alert on the toxicity of a chemical from its structure. An alert is given if a structural feature or toxicophore associated with the occurrence of skin sensitisation has been recognized. To each alert there is a certainty level is associated. Chemicals with a skin sensitisation alert with a "probable", "plausible", "equivocal" or "doubted" certainty level are conservatively regarded as potential sensitisers.
Metabolic competence (if applicable)	Not applicable.
Status of information source development, standardisation, validation	Commercially available software, no official validation. Derek Nexus skin sensitisation alerts follow OECD *in silico* models' validation principles (OECD, 2004). The approach is published in peer-reviewed journals.
Technical limitations and limitations with regard to applicability	The method can only be applied to chemicals with a defined structure (no mixtures, no polymers). Its domain mostly covers small organics, rarely inorganics. To each alert there is a certainty level is associated. Chemicals with a skin sensitisation alert with a "probable", "plausible", "equivocal" or "doubted" certainty level are conservatively regarded as potential sensitisers. Alerting system, not prediction model (i.e. no identification of "negatives" in our case "non-sensitisers" possible).
Weaknesses and Strengths	Strengths: Mechanism based alerts; the results are extensively documented; the approach is published in peer-reviewed journals; transparency of the algorithms used to generate data; only the chemical structure is needed as input. Weaknesses: Commercial software; no calculations on structurally unidentified substances and mixtures possible; alerting system, not prediction model (i.e. no identification of "negatives" in our case "non-sensitisers" possible).
Reliability	Not applicable
Predictive capacity (if applicable)	Alerting system, not prediction model (i.e. no identification of "negatives" in our case "non-sensitisers" possible).

Proprietary aspects	A license agreement is needed for Derek Nexus, commercially available software from Lhasa Limited.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within a Defined Approach. The alerts can contribute to classification of chemicals into mechanistic domains to support read-across.
Potential role within the Defined Approach (see case study III)	The Derek Nexus alerts are foreseen to be combined with complementary information and evaluated in the context of Defined Approach. In such context, the Derek Nexus alerts are part of the integrated strategy for skin sensitisation hazard identification based on *in silico*, *in chemico*, and *in vitro* data analysed using a statistic "staking" meta-model (Gomes et al., 2012).

12. OECD TG 428 modified to include time course and free/bound measurements

Name of the information source	OECD TG 428 modified to include time course and free/bound measurements
Mechanistic basis including coverage of the AOP	Our skin bioavailability and protein haptenation kinetics data aim to quantify the free & irreversibly bound concentration of chemical throughout the different layers of human skin over time to allow us to predict the extent of protein haptenation within the viable layers of the skin (i.e. layers that are 'sampled' by dendritic cells). Skin penetration is defined within the AOP as penetration through the stratum corneum, however we hypothesise that quantitative kinetic information on the amount of chemical present in the different layers of viable skin are required to allow an accurate prediction of the sensitiser-induced T cell response to be made. Haptenation of skin protein is the molecular initiating event (MIE) defined within the Skin Sensitisation AOP. Our skin haptenation kinetics data aims to accurately characterise this event *in vitro* through measuring the protein haptenation rate of the sensitising chemical in the context of actual *ex vivo* human skin. In this sense the reaction rate is assumed to be more representative of the actual *in vivo* reaction rate than those provided by model peptides or cell-based assays. However, we are exploring reactivity data obtained using model peptides and cell lysates in order to determine whether these systems provide similar reaction rates to those we have found in skin. The major assumption implicit within our reliance on these measurements is that following extraction of free chemical the radio-labelled chemical found within the protein fraction is covalently (irreversibly) bound to the skin protein and not non-covalently associated.
Description	Skin bioavailability kinetics and protein haptenation kinetics data are obtained via a modification to OECD Skin Penetration test guideline 428 that has previously been documented (Pendlington et al., 2008; Davies et al., 2011). Briefly, radio-labelled chemical is applied to the top layer of *ex vivo* human skin; incubations are then stopped at predetermined time points (0.5, 1, 2, 4, 8 and 24 hours) by removing and separating the skin samples to determine the extent of the free and irreversibly bound chemical in each of the different skin layers. The readout for skin bioavailability kinetics and protein haptenation kinetics is the measured amount of test item in each of the following: stratum corneum, epidermis (free and bound), dermis (free and bound), receptor fluid and other measurements appropriate to determine full mass balance. The test can be run such that it returns the total amount in each compartment (Pendlington et al., 2008) or further analysis can be performed to determine the free and irreversibly bound amounts. Where free and irreversibly bound amounts of test item are determined, this is achieved by homogenising the skin layer and extracting the free test item, to allow measurement of what is irreversibly bound to protein.
Response(s) measured	The test method has been developed to characterise the skin bioavailability kinetics (AOP step 1) and protein haptenation kinetics (AOP step 3 and 4, key event 1) of chemical sensitisers.

Prediction model	N/A
Biological relevance of the test system used	*Ex vivo* human skin is the test system used in the *in vitro* skin absorption component of the bioavailability measurement. As such it is therefore directly relevant to the *in vivo* situation, with the following caveat: *in vivo*, materials passing into the skin meet the systemic circulation at the level of the microvasculature (capillary bed) that lies at the epidermal/dermal junction; in the *in vitro* skin absorption assay, the receptor fluid flows below the lower surface of the skin. The skin is dermatomed to remove most of the dermis (total thickness stratum corneum + epidermis + dermis approximately 400µm) in an effort to redress this difference.
Metabolic competence (if applicable)	It is also assumed that the *ex vivo* skin is not metabolically active.
Status of information source development, standardization, validation	The skin bioavailability kinetics and protein haptenation kinetics data is a modification of OECD Test Guideline 428.
Technical limitations and limitations with regard to applicability	The current input data measurement systems have been selected to allow the SARA model to be applied for organic chemicals that do not require auto-oxidation or skin metabolism to become protein-reactive. The test items need to be soluble in a suitable vehicle.
Weaknesses and Strengths	- As far as we are aware, these represent the most relevant approaches for direct measurement of bioavailability and kinetics of protein haptenation by a sensitising chemical in human skin. - Applicability domain: need to radiolabel the chemical of interest prior to experimental data generation - Metabolic capacity: does not allow the assessment of pro- or pre-haptens
Reliability (with and between laboratories (if applicable))	There is inherent variability in results obtained using the *in vitro* skin absorption method: both inter- and intra-skin donor. To take this into account skin from multiple donors is used in each experiment, with skin from multiple donors being used for each time point and a full time course obtained for each donor. We have in-house data that indicate that the method is transferable between laboratories (the method has been performed by ourselves, a CRO, and Unilever colleagues at a different site).
Predictive Capacity (if applicable)	Parameters are inferred from the skin bioavailability and protein binding data to inform the mathematical model and relate to partitioning between skin compartments and rates of diffusion, evaporation and haptenation. The parameters are inferred by Bayesian parameter estimation using markov chain monte carlo (Gilks et al., 1996) and computation performed in Python using packages numpy and scipy (Python Software Foundation. Python Language Reference, Python version 3.3.5, numpy version 1.8.1 and scipy version 0.14.0. Available at http://www.python.org). Standard model checking procedures (Gelman et al., 2013) are used to ensure that the model generates plausible posterior predictive simulation data on comparison with actual experimental data.
Proprietary aspects	To date the method has been performed by ourselves, a CRO, and Unilever

	colleagues at a different site, however a manuscript detailing the method has recently been submitted for publication (Reynolds et al., 2016).
Proposed regulatory use	Input data for skin sensitisation risk assessment.
Potential role within the Defined Approach (see case study XII)	Skin bioavailability and skin protein haptenation rate are used in the SARA model as input data to predict the rate of human, naïve $CD8^+$ T cell receptor triggering using; these datasets are generated to closely mimic the human *in vivo* experimental or consumer product exposure scenario that is being risk assessed.

REFERENCES

Ade N, Martinozzi-Teissier S, Pallardy M, Rousset F. (2006). Activation of U937 cells by contact sensitizers: CD86 expression is independent of apoptosis. J Immunotoxicol, 3:189-97

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

EURL ECVAM (2013). DPRA validation study report. 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

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

Fabian E, Vogel D, Blatz V, Ramirez T, Kolle S, Eltze T, van Ravenzwaay B, Oesch F, Landsiedel R. (2013). Xenobiotic metabolizing enzyme activities in cells used for testing skin sensitization in vitro. Arch Toxicol, 87:1683-96.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. (2013). Bayesian Data Analysis, Third Edition. Taylor & Francis.

Gilks WR, Richardson S, Spiegelhalter, D. (1996). Markov Chain Monte Carlo in Practice, Chapman & Hall/CRC Interdisciplinary Statistics Series.

Langton K, Patlewicz G, Long A, Marchant C and Basketter D. (2006). Structure–Activity Relationships for Skin Sensitization: Recent Improvements To Derek For Windows; Contact Dermatitis, 55:342-7.

Natsch A, Gfeller H. (2008). LC-MS-based characterization of the peptide reactivity of chemicals to improve the in vitro prediction of the skin sensitization potential. Toxicol Sci, 106:464-78.

Natsch A, Ryan CA, Foertsch L, Emter R, Jaworska J, Gerberick F, Kern P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. J Appl Toxicol, 33:1337-52.

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.

Nukada Y, Ashikaga T, Sakaguchi H, Sono S, Mugita N, Hirota M, Miyazawa M, Ito Y, Sasa H, Nishiyama N. (2011). Predictive performance for human skin sensitizing potential of the human cell line activation test (h-CLAT). Contact Dermatitis, 65:343-53.

OECD, 2004. OECD principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationships models. http://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf

Patlewicz G, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MI, Aptula AO, Phillips RD, Niemelä J, Madsen C, Wedebye EB, Roberts DW, Bailey PT, Mekenyan OG. (2007). TIMES-SS-A promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. Regul Toxicol Pharmacol, 48:225-39.

Ramirez T, Mehling A, Kolle SN, Wruck CJ, Teubner W, Eltze T, Aumann A, Urbisch D, van Ravenzwaay B, Landsiedel R. (2014). LuSens: a keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification. Toxicol In Vitro, 28:1482-97.

Ramirez T, Stein N, Aumann A, Remus T, Edwards A, Norman KG, Ryan C, Bader JE, Fehr M, Burleson F, Foertsch L, Wang X, Gerberic, F, Beilstein P, Hoffmann S, Mehling A, van Ravenzwaay B, Landsiedel R. (2016). Intra- and inter-laboratory reproducibility and accuracy of the LuSens assay: A reporter gene-cell line to detect keratinocyte activation by skin sensitizers. Toxicol In Vitro, 32:278-86.

Strickland J, Zang Q, Kleinstreuer N, Paris M, Lehmann DM, Choksi N, Matheson J, Jacobs A, Lowit A, Allen D, Casey W. (2016) Integrated decision strategies for skin sensitization hazard. J Appl Toxicol, 36:1150-62.

Takenouchi O, Miyazawa M, Saito K, Ashikaga T, Sakaguchi H. (2013). Predictive performance of the human Cell Line Activation Test (h-CLAT) for lipophilic chemicals with high octanol-water partition coefficients. J Toxicol Sci, 38:599-609.

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 2:337-51.

Van der Veen JW, Pronk TE, van Loveren H, Ezendam J. (2013). Applicability of a keratinocyte gene signature to predict skin sensitizing potential. Toxicol In Vitro, 27:314-22.

Van der Veen JW, Rorije E, Emter R, Natsch A, van Loveren H, Ezendam J. (2014). Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. Regul Toxicol Pharmacol, 69:371-9.

Van der Veen JW, Hodemaekers H, Reus AA, Maas WJM, Van Loveren H, Ezendam J. (2015). Human relevance of an in vitro gene signature in HaCaT for skin sensitization. Toxicol in Vitro, 29:81-4.