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

GUIDANCE DOCUMENT ON THE REPORTING OF DEFINED APPROACHES AND INDIVIDUAL INFORMATION SOURCES TO BE USED WITHIN INTEGRATED APPROACHES TO TESTING AND ASSESSMENT (IATA) FOR SKIN SENSITISATION

Series on Testing & Assessment
No. 256

The corresponding annexes are available in the following cotes:
ENV/JM/MONO(2016)29/ANN1 and ENV/JM/MONO(2016)29/ANN2

JT03403923

Complete document available on OLIS in its original format
This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

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

OECD Environment, Health and Safety Publications

Series on Testing & Assessment

No. 256

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

© OECD 2016
Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

FOREWORD

With a view to assisting the evaluation of integrated approaches to testing and assessment (IATA) in regulatory decision-making within OECD Member Countries, this guidance document provides guidance on the reporting of defined approaches to testing and assessment. A defined approach consists of a fixed data interpretation procedure (DIP) (e.g. statistical, mathematical models) applied to data (e.g *in silico* predictions, *in chemico*, *in vitro* data) generated with a defined set of information sources to derive a prediction. In contrast to the assessment process within Integrated Approaches to Testing and Assessment (IATA), that necessarily involves some degree of expert judgment, predictions generated with defined approaches are rule-based and can either be used on their own if they are deemed to fit-for-purpose or considered together with other sources of information in the context of IATA.

The template for reporting defined approaches to testing and assessment based on multiple information sources and the template for reporting individual information sources are provided in guidance document ENV/JM/MONO(2016)28 and they have been used by an ad-hoc expert group to document a number of defined approaches developed in the area of skin sensitisation using the adverse outcome pathway (AOP) as a conceptual framework. These defined approaches are proposed for hazard and/or potency prediction. It is not the intent of this document to seek for endorsement of any specific defined approach provided in the case studies, but rather provide a perspective of how individual information sources and defined approaches developed for skin sensitisation assessment should be reported in a harmonised way and to illustrate what forms these may take, whether they are statistically derived, or qualitative in nature, and serve different purposes (i.e. hazard versus potency prediction). A harmonised approach in the reporting of the different elements used within IATA is critical to ensure consistency in the use of IATA-derived predictions/assessments for regulatory decisions and to promote mutual acceptance of such assessments. The present document was endorsed by the Task Force on Hazard Assessment in June 2016.

This document is being published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology, which has agreed that it be declassified and made available to the public.

TABLE OF CONTENTS

1. INTRODUCTION

This document provides guidance on the reporting of defined approaches to testing and assessment. A defined approach consists of a fixed data interpretation procedure (DIP) (e.g. statistical, mathematical models) applied to data (e.g *in silico* predictions, *in chemico*, *in vitro* data) generated with a defined set of information sources to derive a prediction. In contrast to the assessment process within Integrated Approaches to Testing and Assessment (IATA), that necessarily involves some degree of expert judgment, predictions generated with defined approaches are rule-based and can either be used on their own if they are deemed to fit-for-purpose or considered together with other sources of information in the context of IATA.

This document is not intended to endorse any specific defined approach exemplified in the case studies. The case studies are provided as examples of the level of information needed to facilitate a harmonised approach to the reporting of defined approaches that can be used as elements within IATA specifically in the field of skin sensitisation. A harmonised approach in reporting the different IATA elements is critical to ensure consistency in the use of IATA-derived predictions/assessments for regulatory decisions and to promote mutual acceptance of such assessments.

2. BACKGROUND

Allergic contact dermatitis (ACD) is the clinical manifestation of a changed responsiveness of the adaptive immune system following repeated exposure to a sensitising substance. The development of ACD is characterised by two distinct phases: 1) the induction of specialised immunological memory following the initial exposure to the allergen, termed sensitisation and 2) elicitation of the visible, clinical allergic response following subsequent exposure to the allergen.

Historically, predictive tests to identify and characterise substances causing ACD have used animals. The standard and accepted skin sensitisation test methods, for which OECD guidelines are available, include the guinea-pig maximisation test (GPMT) according to Magnusson and Kligman and the occluded patch test of Buehler (TG 406), where the endpoint measured is elicitation (i.e. the organism response/adverse outcome); and the mouse local lymph node assay (LLNA; TG 429) and its non-radioactive variants (TG 442a and TG 442b) where the endpoint measured is cell proliferation in the lymph node (i.e. organ response/induction).

There is general agreement on the key chemical and biological events underlying skin sensitisation (e.g. Karlberg et al., 2008; Vocanson et al. 2009; Adler et al., 2010; Martin et al., 2010; Kimber et al., 2011), and this knowledge has now been summarised by the OECD in the report entitled: "The Adverse Outcome Pathway (AOP) for skin sensitisation initiated by covalent binding to proteins" (OECD, 2012a, 2012b) to facilitate the development of toxicological assays and strategies to assess this toxicological endpoint.

The skin sensitisation AOP identifies four key events (KEs) with KE_1, the covalent binding to skin proteins (termed haptenation) either of the parent substance or of its reactive derivatives following abiotic/metabolic activation, which is postulated to be the molecular initiating event (MIE), followed by KE_2, the activation of epidermal keratinocytes, KE_3, the activation (maturation) and mobilisation of Langerhans cell and dermal dendritic cells (DC), and KE_4, the DC-mediated antigen presentation to naïve T-cells and proliferation /activation of allergen specific T-cells (Figure 1).

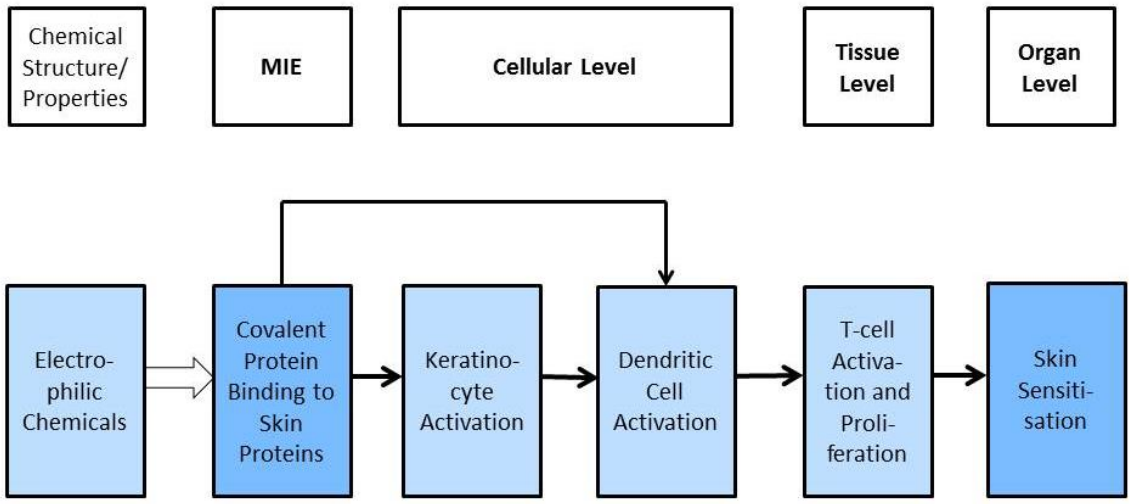


Figure 1: Flow diagram of the pathways and the intermediate steps associated with skin sensitisation (adapted from OECD, 2012a).

Knowledge of the skin sensitisation pathway has prompted the development of alternative methods (*in silico*, *in chemico*, *in vitro*) addressing specific KEs. Information generated by these methods can contribute to the assessment of the skin sensitisation potential and potency of chemicals when used as information sources within defined approaches and IATA. Within such AOP-informed defined approaches/IATA, the different information sources would target KEs along the defined toxicity pathway and the results could be used to inform a regulatory decision.

Non-testing and testing methods are available to estimate penetration, simulate metabolism or abiotic transformation processes as well as identify electrophilic features and quantify their reactivity. *In chemico* and *in vitro* assays are also available to measure reactivity, informing about the ability of a substance to activate the MIE. *In vitro* assays are available to characterise keratinocyte inflammatory responses and to measure markers of dendritic/monocytic cell activation. These methods are able to characterise a number of the KEs in the skin sensitisation AOP and in doing so form the basis of AOP-informed defined approaches and IATA. Exposure considerations and an understanding of bioavailability may also inform the defined approach or IATA, though these components fall outside of the definition of an AOP (OECD, 2013).

The availability of non-animal methods for skin sensitisation favoured in recent years the development of defined approaches to testing and assessments which, in most cases, are designed to predict an existing line of evidence (i.e. responses in animal models or in humans). Within such defined approaches data generated with selected sources of information (i.e. physicochemical properties, *in silico*, *in chemico*, *in vitro* data etc.) are converted into predictions by applying a DIP. Examples of DIP include mathematical and statistical models.

Predictions generated with defined approaches can be used on their own if considered adequate for a specific regulatory application or may be evaluated together with other information sources in the assessment process within IATA. In such a case a defined approach would be considered as an IATA component.

3. MAPPING OF INFORMATION SOURCES THAT CAN BE USED WITHIN DEFINED APPROACHES AND IATA FOR SKIN SENSITISATION BY APPLYING THE AOP AS A FRAMEWORK

Depending on the final purpose (e.g. hazard or risk), the assessment of skin sensitisation can include: consideration of the expected exposure to the substance being evaluated, an understanding of dermal bioavailability including skin penetration and metabolism, information on KEs and any other supporting information, i.e. information from non-testing and testing methods designed to address other health or environmental endpoints that nevertheless may inform skin sensitisation assessment. The possible elements and information sources that can be used within defined approaches and IATA for skin sensitisation assessment are listed in Table 1. Some of the elements, highlighted in grey in Table 1, address KEs within the skin sensitisation AOP. Note that this is not an exhaustive list and does not imply any judgement about the suitability of any of the listed information sources for a specific assessment.

It has to be noted that the elements addressed within a specific defined approach or IATA and the type of information sources used to populate each individual element may vary depending on the scope and the specific regulatory requirement. This implies that for certain regulatory purposes (e.g. hazard identification) the assessment may be conducted by addressing fewer elements than in the case of more complex regulatory needs (e.g. risk assessment). It is therefore envisaged that different defined approaches and IATA solutions may be possible depending on the chemical under investigation, the regulatory need and the specific regulatory requirements in the different regions.

Table 1: Elements and examples of information sources that can be used within defined approaches and IATAs for skin sensitisation

Elements	**Information sources addressing each element**
Exposure consideration	• Applied dose • Frequency of exposure • Formulation effects • *In vitro* to *in vivo* extrapolation
Chemical descriptors	**Chemical structure** **Physico-chemical properties** • Molecular Weight • Physical state • pKa • Log Kow • Evaporation rate/Vapour pressure • Water solubility
Dermal bioavailability − **Skin penetration**	**Non-testing methods** • Characterisation of skin absorption (e.g. physiologically based-pharmacokinetic (PBPK) models **Testing methods** • TG 428 (Skin absorption: *in vitro* method) • TG 427 (Skin absorption: *in vivo* method) • TG 428 modified to include time course (Pendlington et al., 2008; Davies et al., 2011)
− **Skin metabolism**	**Non testing methods:** • *In silico* e.g. structure-metabolism rules encoded in the expert system TIMES-SS, Meteor Nexus, simulators for skin metabolism and autoxidation within the OECD Toolbox **Testing methods** • Incubation with S9 or microsomes from skin or surrogate systems (e.g. liver) • Peroxidase-peroxide system

AOP key event 1: Covalent interaction with cellular proteins	
	Non-testing methods • Protein binding/reactivity alerts (e.g. OECD Toolbox, Derek Nexus, Toxtree, TIMES-SS)[1] **Testing methods** • TG 442C (Direct Peptide Reactivity Assay) • Adduct formation or relative reactivity rate, with or without metabolic activation, e.g: − Cor1C420 assay (Natsch and Gfeller, 2008) − PPRA (Gerberick et al., 2009) − Kinetic DPRA (Roberts and Natsch et al., 2009) − Glutathione depletion assay (Aptula et al., 2006; Schultz et al., 2005) − TG 428 modified to include free/bound measurements (Pickles et al., submitted) − Allergen-protein interaction assay (APIA; Dietz et al., 2013) − Amino acid Derivative Reactivity Assay (ADRA; Yamamoto et al., 2015) − SH test (Suzuki et al., 2009)
AOP key event 2: events in Keratinocytes	
Activation of biochemical pathways	**Testing methods** • TG 442D (ARE-Nrf2 Luciferase Test Method-KeratinoSens™) • LuSens (Ramirez et al., 2014, 2016) • AREc32 cell line assay (Natsch and Emter, 2008).
Pathways-associated gene expression	• SENS-IS (Cottrez et al., 2015, 2016) • HaCaT gene signature (van der Veen et al., 2013) • SenCeeTox (McKim et al., 2012) • Epidermal Sensitization Assay (EpiSensA; Saito et al., 2013)
Pathways-associated protein expression	• Proteomic signature in keratinocytes (Thierse et al., 2011)
Release of pro-inflammatory mediators	• RhE-IL-18 (Gibbs et al., 2013)
AOP key Event 3: Events in Dendritic cell	
Expression of co-stimulatory and adhesion molecules in dendritic / monocytic cells	**Testing methods** • h-CLAT (Ashikaga et al., 2010; TG 442E) • U-SENS™ (Piroird et al., 2015) • modified MUSST (Bauch et al., 2012) • PBMDC (Reuter et al., 2011)
Pathways-associated protein	• MUTZ SensiDerm (Thierse et al., 2011)

expression in dendritic / monocytic cells **Pathways-associated gene expression in dendritic / monocytic cells**	• IL-8 Luc assay **(**Takahashi et al., 2011) • GARD (Johansson et al., 2013) • VitoSens (Hooyberghs et al., 2008)
AOP key event 4: Events in Lymphocytes	
	Testing methods • Human T cell priming/proliferation assay (hTCPA) (Moulon et al., 1993; Krasteva et al., 1996; Dietz et al., 2010; Martin et al., 2010, Richter et al., 2013; Popple et al., 2015) **(Existing) animal data** • TG 429 (LLNA) • TG 442A (LLNA: DA) • TG 442B (LLNA: BrdU-ELISA)
AOP Adverse Outcome	
	(Existing) human data • Human Repeat Insult Patch Test (HRIPT) • Human Maximisation Test • Clinical data • Data from occupational exposure • Epidemiological data **(Existing) animal data** • TG 406 (Guinea-pig Maximisation Test; Buehler Test)
Others	• Skin corrosion (e.g. OECD TG 430,431,435, 404) • Skin irritation (e.g. OECD TG 439, 404) • Genotoxicity (e.g. OECD TG 471) (see Wolfreys and Basketter, 2004; Patlewicz et al., 2010; Mekenyan et al., 2010)

[1] Note Derek Nexus and TIMES-SS are expert systems that aim to provide a prediction of likely skin sensitisation hazard and potency drawing on knowledge captured in SARs and in the case of TIMES-SS additionally underpinned by QSARs. As such their scope is broader than simply providing insight of potential electrophilic reaction centres indicative of protein binding potential which itself defines the MIE.

The sorts of (Q)SAR models that are available for skin sensitisation are provided in Table 2 for illustrative purposes. For more information, reviews describing the available *in silico* approaches for skin sensitisation include Patlewicz and Worth (2008) and more recently Sharma et al. (2012).

Table 2: QSARs models for skin sensitisation

Model	Type	Chemical coverage	Availability	Anchor point in the AOP	Endpoint predicted	Role in IATA	References
Relative alkylation index (RAI) approach	Local QSAR approach	Various RAI derived for specific chemical classes e.g. sulfonate esters, sulfones, primary alkyl bromides, acrylates, aldehydes and diketones	Published in the literature	KE4, AO	Most of the RAI models aim to predict the EC3 value in the LLNA, a few predict the outcome in guinea pig tests	Hazard identification and characterisation	Examples include: Roberts and Williams, (1982), Roberts et al., (1983, 1991, 2007a), Roberts, (1987, 1995), Roberts and Basketter, (1990, 1997, 2000), Patlewicz et al., (2002), Patlewicz et al., (2004), Roberts et al., (1999), Roberts and Patlewicz, (2002)
QMM approach which is an extension of the RAI approach	Local QSAR approach	Developed on the basis of Reaction mechanistic domains (Schiff base formers, Michael addition, Acylating agents, SN2)	Published in the literature	KE4	EC3 in the LLNA	Hazard identification and characterisation	Examples are: Roberts et al., (2006, 2011), Roberts and Natsch, (2009); Roberts and Aptula, (2014).
Various e.g. Estrada et al., (2003)	Global models	Mainly based on the Gerberick et al. (2005) dataset hence cover a broad coverage of chemicals	Variable	KE4	Potency categorisation as defined by EC3 values in the LLNA	Hazard identification – semi-quantitative assessment of potency	Many were reviewed in Roberts et al. (2007b)
TOPKAT	Expert system (statistical)	Based mainly on the datasets published by Cronin and Basketter (1994) hence reasonably broad coverage of chemicals	Commercial	AO	Binary model to predict likelihood of sensitisation and additional model to estimate qualitatively the potency as defined in the GPMT	Hazard identification – semi-quantitative assessment of potency	http://www.accelrys.com/products/topkat/
MCASE Suite of models to predict each of the KEs in the AOP	Expert system (statistical)	Broad coverage of chemicals	Commercial	KE1 (MIE), KE2, KE3, KE4, AO	Models to predict the outcome of the DPRA, ARE activation, n-CLAT, EC3 potency bands and overall binary sensitisation outcome	Hazard identification – semi-quantitative assessment of potency	http://www.multicase.com/case-ultra-models#skin_eye_tox_bundle
Derek	Expert	Broad coverage	Commercial	KE4, AO	Qualitative	Hazard	http://www.lhasalimited.org/index.php

Nexus	system (Knowledge based)	of chemicals			likelihood of skin sensitisation potential	identification	
TIMES-SS	Expert system (Hybrid)	Broad coverage of chemicals	Commercial	AO	Based on data from LLNA, GPMT and Human	Hazard identification – semi-quantitative assessment of potency	Dimitrov et al., (2005); Patlewicz et al., (2007, 2014)

(Q)SAR predictions may be gathered from databases (in which the predictions have already been generated and documented) or generated *de novo* through the available models. Most (Q)SARs do not account for transformation of a substance explicitly. Some expert systems such as TIMES-SS incorporate simulators for metabolism so that predictions for parent compounds and their metabolites are considered at the same time in making an overall prediction of activity. Derek Nexus can be linked to its Meteor Nexus metabolism program to make predictions of parent compounds and their estimated metabolites. The OECD toolbox incorporates simulators for metabolism and degradation such that a parent chemical and its expected metabolites can be profiled together for the purposes of forming chemical categories to facilitate data gap filling.

Conclusions about the likely properties of a substance can also be based on the knowledge of the properties of one or more similar chemicals, by applying chemical grouping methods.

The OECD guidance document, Series on Testing and Assessment No. 194 provides information on the use of chemical grouping and read-across approaches (OECD, 2014). As with (Q)SARs, grouping approaches can be used to indicate either the presence or the absence of an effect.

4. DEFINED APPROACHES TO TESTING AND ASSESSMENT AND THEIR ROLE WITHIN IATA FOR SKIN SENSITISATION

In the area of skin sensitisation the availability of a suite of non-animal *in silico*, *in chemico* and *in vitro* methods has prompted the development of defined approaches based on the integration of readouts from these methods. As defined in the OECD guidance document ENV/JM/MONO(2016)28, defined approaches to testing and assessment are based on a fixed set of information sources and a fixed data interpretation procedure (DIP) to convert inputs from the different information sources into a prediction.

The DIP within defined approaches can range from simple rule-based decision steps to mathematical and statistical models. In contrast to the WoE process, in a defined approach the weighting of the different information is fixed and does not leave room for subjective interpretation. The final prediction can be used on its own if fit-for-purpose to satisfy a specific regulatory need or can be used as a component within IATA and thus considered in the WoE assessment together with other relevant information (see Table 1).

In contrast to an IATA that is customised for the chemical/class of chemicals under investigation and the specific regulatory need, defined approaches are generally designed to be applicable to a larger chemical space and most of those available in the area of skin sensitisation have been developed to predict an existing line of evidence (e.g. LLNA hazards or potency).

An overview of the defined approaches, documented in more details in Annex I (ENV/JM/MONO(2016)29/ANN1), is provided in Table 3.

Table 3: Defined approaches to testing and assessment documented in Annex I.

Case study		**Purpose**
I	An Adverse Outcome Pathway-based "2 out of 3" integrated testing strategy approach to skin hazard identification (BASF)	Hazard identification
II	Sequential Testing Strategy (STS) for hazard identification of skin sensitisers (RIVM)	Hazard identification
III	A non-testing Pipeline approach for skin sensitisation (G. Patlewicz)	Hazard identification
IV	Stacking meta-model for skin sensitisation hazard identification (L'Oréal)	Hazard identification
V	Integrated decision strategy for skin sensitisation hazard (ICCVAM)	Hazard identification
VI	Consensus of classification trees for skin sensitisation hazard prediction (EC- JRC)	Hazard identification

VII	Sensitizer potency prediction based on Key event 1 + 2: Combination of kinetic peptide reactivity data and KeratinoSens® data (Givaudan)	Potency prediction
VIII	The artificial neural network model for predicting LLNA EC3 (Shiseido)	Potency prediction
IX	Bayesian Network DIP (BN-ITS-3) for hazard and potency identification of skin sensitizers (P&G)	Potency prediction
X	Sequential testing strategy (STS) for sensitising potency classification based on in chemico and in vitro data (Kao Corporation)	Potency prediction
XI	Integrated testing strategy (ITS) for sensitising potency classification based on in silico, in chemico, and in vitro data (Kao Corporation)	Potency prediction
XII	DIP for skin allergy risk assessment (SARA) (Unilever)	Potency prediction

The intent of this guidance document is to exemplify how these defined approaches and the information sources used within (see Annex II in ENV/JM/MONO(2016)29/ANN2) should be documented to facilitate a harmonised methodology in their reporting, critical to ensure consistency in the use of IATA-derived predictions/assessment for regulatory decisions.

The case studies documented in this guidance document do not imply acceptance or endorsement by any Member Country or OECD. They are intended only to provide a perspective of how individual information sources and defined approaches, used on their own or within an IATA for skin sensitisation, should be reported and to illustrate what forms these may take, whether they are statistically derived, or qualitative in nature, and serve different purposes (i.e. hazard versus potency prediction).

REFERENCES

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

Aptula AO, Patlewicz G, Roberts DW, Schultz TW. (2006). Non-enzymatic glutathione reactivity and *in vitro* toxicity: a non-animal approach to skin sensitization. Toxicol In Vitro, 20:237-47.

Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y, Miyazawa M, Ito Y, Nishiyama N, Itagaki H. (2010). A comparative evaluation of *in vitro* skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). Altern Lab Anim, 38:275-84.

Cottrez F, Boitel E, Auriault C, Aeby P, Groux H. (2015). Genes specifically modulated in sensitized skins allow the detection of sensitizers in a reconstructed human skin model. Development of the SENS-IS assay. Toxicol In Vitro, 29:787-802.

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

Davies M, Pendlington RU, Page L, Roper CS, Sanders DJ, Bourner C, Pease CK, MacKay C. (2011). Determining epidermal disposition kinetics for use in an integrated nonanimal approach to skin sensitization risk assessment. Toxicol Sci, 119:308-18.

Dietz L, Esser PR, Schmucker SS, Goette I, Richter A, Schnölzer M, Martin SF, Thierse HJ. (2010). Tracking human contact allergens: from mass spectrometric identification of peptide-bound reactive small chemicals to chemical-specific naive human T-cell priming. Toxicol Sci, 117:336-47.

Dietz L, Kinzebach S, Ohnesorge S, Franke B, Goette I, Koenig-Gressel D, Thierse HJ. (2013). Proteomic allergen-peptide/protein interaction assay for the identification of human skin sensitizers. Toxicol In Vitro, 27:1157-62.

Dimitrov SD, Low LK, Patlewicz GY, Kern PS, Dimitrova GD, Comber MH, Phillips RD, Niemela J, Bailey PT, Mekenyan OG. (2005). Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. Int J Toxicol. 24:189-204.

Estrada E, Patlewicz G, Chamberlain M, Basketter D, Larbey S. (2003). Computer-aided knowledge generation for understanding skin sensitization mechanisms: the TOPS-MODE approach. Chem Res Toxicol, 16:1226-35.

Gerberick GF, Troutman JA, Foertsch LM, Vassallo JD, Quijano M, Dobson RLM, Goebel C, Lepoittevin JP. (2009). Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. Toxicol Sci, 112:164-74.

Gibbs S, Corsini E, Spiekstra SW, Galbiate V, Fuchs HW, Degeorge G, Troese M, Hayden P, Deng W, Roggen E. (2013). An epidermal equivalent assay for identification and ranking potency of contact sensitizers. Toxicol Appl Pharmacol, 272:529-41.

Hooyberghs J, Schoeters E, Lambrechts N, Nelissen I, Witters H, Schoeters G, van den Heuvel RA. (2008). Cell-based *in vitro* alternative to identify skin sensitisers by gene expression. Toxicol Appl Pharmacol, 231:103-11.

Johansson H, Albrekt AS, Borrebaeck CA. (2013). The GARD assay for assessment of chemical skin sensitisers. Toxicol In Vitro, 37:1163-69.

Karlberg AT, Bergström MA, Börje A, Luthman K, Nilsson JL. (2008). Allergic contact dermatitis-formation, structural requirements, and reactivity of skin sensitizers. Chem Res Toxicol, 21:53-69.

Kimber I, Basketter DA, Gerberick GF, Ryan CA, Dearman RJ. (2011). Chemical allergy: translating biology into hazard characterization. Toxicol Sci, 120:238-68.

Krasteva M, Peguet-Navarro J, Moulon C, Courtellemont P, Redziniak G, Schmitt D. (1996). In vitro primary sensitization of hapten-specific T cells by cultured human epidermal Langerhans cells - a screening predictive assay for contact sensitizers. Clin Exp Allergy, 26:563-70.

Martin SF, Esser PR, Schmucker S, Dietz L, Naisbitt DJ, Park BK, Vocanson M, Nicolas JF, Keller M, Pichler WJ, Peiser M, Luch A, Wanner R, Maggi E, Cavani A, Rustemeyer T, Richter A, Thierse HJ, Sallusto F. (2010). T-cell recognition of chemicals, protein allergens and drugs: towards the development of in vitro assays. Cell Mol Life Sci, 67:4171-84.

McKim JM, Keller DJ, Gorski JR. (2012). An *in vitro* method for detecting chemical sensitisation using human reconstructed skin models and its applicability to cosmetic, pharmaceutical, and medical device safety testing. Cutan Ocul Toxicol, 31:292-305.

Mekenyan O, Patlewicz G, Dimitrova G, Kuseva C, Todorov M, Stoeva S, Kotov S, Donner EM. (2010). Use of genotoxicity information in the development of integrated testing strategies (ITS) for skin sensitization. Chem Res Toxicol, 23:1519-40.

Moulon C, Peguet-Navarro J, Courtellemont P, Redziniak G, Schmitt D. (1993). In vitro primary sensitisation and restimulation of hapten-specific T cells by fresh and cultured human epidermal Langerhans' cells. Immunology, 80:373-79.

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

Natsch A, Emter R. (2008). Skin sensitizers induce antioxidant response element dependent genes: application to the *in vitro* testing of the sensitization potential of chemicals. Toxicol Sci, 102:110-19.

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

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. ENV/JM/MONO(2012)10/PART2 http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part2&doclanguage=en

OECD 2013. Guidance Document on Developing and Assessing Adverse Outcome Pathways. Series on Testing and Assessment No. 184. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en

OECD 2014. Guidance on Grouping of Chemicals, Second edition. Series on Testing and Assessment No.194 http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&doclanguage=en

Patlewicz GY, Wright ZM, Basketter DA, Pease CK, Lepoittevin JP, Gimenez Arnau E. (2002). Structure-activity relationships for selected fragrance allergens. Contact Dermatitis, 47:219-26.

Patlewicz GY, Basketter DA, Pease CK, Wilson K, Wright ZM, Roberts DW, Bernard G, Arnau EG, Lepoittevin JP. (2004). Further evaluation of quantitative structure--activity relationship models for the prediction of the skin sensitization potency of selected fragrance allergens. Contact Dermatitis, 50:91-7.

Patlewicz G, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MIH, Aptula AO, Phillips RD, Niemela 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 characterisation with respect to the OECD Validation Principles for (Q)SARs. Reg Toxicol Pharmacol, 48:225-39.

Patlewicz G, Worth A. (2008) Review of Data Sources, QSARs and Integrated Testing Strategies for Skin Sensitisation. EUR 23225 EN Available at: https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_23225_EN.pdf

Patlewicz G, Mekenyan O, Dimitrova G, Kuseva C, Todorov M, Kotov S, Stoeva S, Donner EM. (2010). Can mutagenicity information be useful in an Integrated Testing Strategy (ITS) for skin sensitization? SAR QSAR Environ Res, 21:619-56.

Patlewicz G, Kuseva C, Mehmed A, Popova Y, Dimitrova G, Ellis G, Hunziker R, Kern P, Low L, Ringeissen S, Roberts DW, Mekenyan O. (2014). TIMES-SS--recent refinements resulting from an industrial skin sensitisation consortium. SAR QSAR Environ Res, 25:367-91.

Pendlington RU, Minter HJ, Stupart L, MacKay C, Roper CS, Sanders DJ, Pease CK. (2008). Development of a modified in vitro skin absorption method to study the epidermal/dermal disposition of a contact allergen in human skin. Cutan Ocul. Toxicol, 27:283-94.

Pickles J, Cubberley R, Moore C, Pendlington R, Saib O, Blackstock C, Vinall J. Development of a modified in vitro skin absorption method with determination of the epidermal/dermal binding profile of three model sensitisers. Toxicol In Vitro, Submitted

Popple A, Williams J, Maxwell G, Gellatly N, Dearman RJ, Kimber I. (2016). The lymphocyte transformation test in allergic contact dermatitis: New opportunities. J Immunotoxicol, 13:84-91.

Piroird C, Ovigne JM, Rousset F, Teissier SM, Gomes C, Cotovio J, Alépée N. (2015). The myeloid U937 skin sensitization test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitisation. Toxicol In Vitro, 24:901-16.

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, Gerberick 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.

Reuter H, Spieker J, Gerlach S, Engels U, Pape W, Kolbe L, Schmucker R, Wenck H, Diembeck W, Wittern KP, Reisinger K, Schepky AG. (2011). In vitro detection of contact allergens: development of an optimized protocol using human peripheral blood monocyte-derived dendritic cells. Toxicol In Vitro, 25:315-23.

Richter A, Schmucker SS, Esser PR, Traska V, Weber V, Dietz L, Thierse HJ, Pennino D, Cavani A, Martin SF. (2013). Human T cell priming assay (hTCPA) for the identification of contact allergens based on naïve T cells and DC-IFN-γ and TNF-α readout. Toxicol in Vitro, 27:1180-85.

Roberts DW, Williams DL. (1982). The derivation of quantitative correlations between skin sensitisation and physico–chemical parameters for alkylating agents and their application to experimental data for sultones. J Theor Biol, 99:807-25.

Roberts DW, Goodwin BFJ, Williams DL, Jones K, Johnson AW, Alderson CJE. (1983). Correlations between skin sensitisation potential and chemical reactivity for p-nitrobenzyl compounds. Food Chem Toxicol, 21:811-13.

Roberts DW. (1987). Structure–activity relationships for skin sensitisation potential of diacrylates and dimethacrylates. Contact Dermatitis, 17:281-89.

Roberts DW, Basketter DA. (1990). A quantitative structure–activity/dose relationship for contact allergenic potential of alkyl group transfer agents. Contact Dermatitis, 23:331-35.

Roberts DW, Fraginals R, Lepoittevin JP, Benezra C. (1991). Refinement of the relative alkylation index (RAI) model for skin sensitization and application to mouse and guinea-pig test data for alkyl alkanesulphonates. Arch Dermatol Res, 283:387-94.

Roberts DW. (1995). Linear free energy relationships for reactions of electrophilic halo- and pseudohalobenzenes, and their application in prediction of skin sensitisation potential for SNAr electrophiles. Chem Res Toxicol, 8: 545-51.

Roberts DW, Basketter DA. (1997). Further evaluation of the quantitative structure-activity relationship for skin-sensitizing alkyl transfer agents. Contact Dermatitis, 37:107–12.

Roberts DW, York M, Basketter DA. (1999). Structure-activity relationships in the murine local lymph node assay for skin sensitization: alpha,beta-diketones. Contact Dermatitis, 41:14-7.

Roberts DW, Basketter DA. (2000). Quantitative structure-activity relationships: sulfonate esters in the local lymph node assay. Contact Dermatitis, 42:154-61.

Roberts DW, Patlewicz G. (2002). Mechanism based structure-activity relationships for skin sensitisation–the carbonyl group domain. SAR QSAR Environ Res, 13:145-52.

Roberts DW, Aptula AO, Patlewicz G. (2006). Mechanistic applicability domains for non-animal based prediction of toxicological endpoints. QSAR analysis of the schiff base applicability domain for skin sensitization. Chem Res Toxicol, 19:1228-33.

Roberts DW, Williams DL, Bethell D. (2007a). Electrophilic reactions of skin-sensitizing sultones. Chem Res Toxicol, 20:61-71.

Roberts DW, Aptula AO, Cronin MT, Hulzebos E, Patlewicz G. (2007b). Global (Q)SARs for skin sensitisation: assessment against OECD principles. SAR QSAR Environ Res, 18:343-65.

Roberts DW, Natsch, A. (2009). High throughput kinetic profiling approach for covalent binding to peptides: application to skin sensitization potency of Michael acceptor electrophiles. Chem Res Toxicol, 16:592-603.

Roberts DW, Aptula AO, Patlewicz GY. (2011). Chemistry-based risk assessment for skin sensitization: quantitative mechanistic modelling for the S(N)Ar domain. Chem Res Toxicol, 24:1003-11.

Roberts DW, Aptula AO. (2014). Electrophilic reactivity and skin sensitisation potency of SNAr electrophiles. Chem Res Toxicol, 17:240-46.

Saito K, Nukada Y, Takenouchi O, Miyazawa M, Sakaguchi H, Nishiyama N. (2013). Development of a new in vitro skin sensitization assay (Epidermal Sensitization Assay; EpiSensA) using reconstructed human epidermis. Toxicol In Vitro, 27:2213-24.

Schultz TW, Yarbrough JW, Johnson EL. (2005). Structure-activity relationships for reactivity of carbonyl-containing compounds with glutathione. SAR QSAR Environ Res, 16:313-22.

Sharma NS, Jindal R, Mitra B, Lee S, Li L, Maguire TJ, Schloss R, Yarmush ML. (2012). Perspectives on Non-Animal Alternatives for Assessing Sensitization Potential in Allergic Contact Dermatitis. Cell Mol Bioeng, 5:52-72.

Suzuki M, Hirota M, Hagino S, Itagaki H, Aiba S. (2009). Evaluation of changes of cell-surface thiols as a new biomarker for in vitro sensitization test. Toxicol in Vitro, 23:687-96.

Takahashi T, Kimura Y, Saito R, Nakajima Y, Ohmiya Y, Yamasaki K, Aiba S. (2011). An *in vitro* test to screen skin sensitizers using a stable THP-1-derived IL-8 reporter cell line, THP-G8. Toxicol Sci, 124:359-69.

Thierse HJ, Budde P, Dietz L, Ohnesorge S, Eikelmeier S, Conde M, Zucht HD, Schulz-Knappe P. (2011). Proteomic identification of allergen-regulated proteins and allergen-protein interaction networks in assisting biomarker and assay development. In: Roggen E.L., Weltzien H.U., Helma H. (Eds.), Progress to worlds novel testing strategies for *in vitro* assessment of allergens. Transworld Research Network, Kerala, India, 145-166.

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.

Vocanson M, Hennino A, Rozières A, Poyet G, Nicolas JF. (2009). Effector and regulatory mechanisms in allergic contact dermatitis. Allergy, 64:1699-714.

Wolfreys AM, Basketter DA. (2004). Mutagens and sensitizers - An unequal relationship? Cutan Ocul Toxicol, 23:197-205.

Yamamoto Y, Tahara H, Usami R, Kasahara T, Jimbo Y, Hioki T, Fujita M. (2015). A novel in chemico method to detect skin sensitizers in highly diluted reaction conditions. J Appl Toxicol, 35:1348-60.

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 I: CASE STUDIES TO THE GUIDANCE DOCUMENT ON THE REPORTING OF DEFINED APPROACHES AND INDIVIDUAL INFORMATION SOURCES TO BE USED WITHIN INTEGRATED APPROACHES TO TESTING AND ASSESSMENT (IATA) FOR SKIN SENSITISATION

Series on Testing & Assessment
No. 256

JT03403925

Complete document available on OLIS in its original format
This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

ENV/JM/MONO(2016)29/ANN1
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing & Assessment

No. 256

ANNEX I: CASE STUDIES TO THE GUIDANCE DOCUMENT ON THE REPORTING OF DEFINED APPROACHES AND INDIVIDUAL INFORMATION SOURCES TO BE USED WITHIN INTEGRATED APPROACHES TO TESTING AND ASSESSMENT (IATA) FOR SKIN SENSITISATION

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

© **OECD 2016** Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

TABLE OF CONTENTS

CASE STUDY I

An Adverse Outcome Pathway - based "2 OUT OF 3" weight of evidence / integrated testing strategy ("2 out of 3 – Sens ITS") approach to skin hazard identification

1. Summary

The defined approach presented in this document describes an integrated testing strategy (ITS) for the identification of the skin sensitisation hazard of a substance primarily for the purposes of classification and labelling without the use of animal testing. The data integration process (DIP) is currently not designed to provide information on the potency of a sensitiser. The combination of test methods used covers the first three key events (KEs) of the adverse outcome pathway (AOP) leading to skin sensitisation as formally described by the OECD: KE 1: protein binding (e.g. via the direct peptide reactivity assay (DPRA); OECD TG 442C); KE 2: keratinocyte activation (e.g. via the KeratinoSens™ or LuSens assay; OECD TG 442D); and dendritic cell activation [e.g. via the human cell line activation test (h-CLAT); OECD TG 442E or the modified Myeloid U937 Skin Sensitisation Test (mMUSST)]. The prediction model entails that two concordant results obtained from methods addressing different steps of first three KEs of the AOP, determine the final classification. Performance and classifications derived from the "2 out of 3 - Sens ITS" of 213 substances were compared to both high quality animal and human data. Depending on the combination of tests used, the "2 out of 3 - Sens ITS" prediction model generally achieved accuracies slightly exceeding those of the murine local lymph node assay (LLNA) when compared to human data. These results compellingly verify the applicability of this easy to understand integrated testing approach (ITS) for a wide range of chemicals.

2. General information

2.1 Identifier:

An AOP - based "2 out of 3" weight of evidence (WoE)/ integrated testing strategy (ITS) approach to skin hazard identification ("2 out of 3 – Sens ITS"; BASF)

2.2 Date:

15 April 2016

2.3 Author(s) and contact details:

Dr. Robert Landsiedel: BASF SE, GB/TB-Z470, 67056 Ludwigshafen, Germany

Email: robert.landsiedel@basf.com; Tel: +49 621 60-56203

Dr. Annette Mehling: BASF PCN GmbH, Henkelstr. 67, 40589 Duesseldorf, Germany

Email: annette.mehling@basf.com; Tel: +49 211 7940-9209

2.4 Template update(s): Version 1.

2.5 Reference to main scientific papers:

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.

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.

Key paper:

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

2.6 Proprietary aspects:

The antioxidant response element (ARE) based assays use luciferase activity as a measure of the sensitisation response. The plasmid encoding the luciferase gene is proprietary to Promega. The terms of use and/or license for the KeratinoSens^{TM} and LuSens cell-lines can be obtained and include the use conditions for the luciferase reporter gene construct. Intellectual property rights protect the h-CLAT in Japan.

3. Endpoint addressed

3.1 Endpoint:

The endpoint being assessed is the inherent skin sensitisation potential of a substance for the purpose of hazard identification and classification primarily for regulatory purposes such as the Globally Harmonized System of Classification, Labelling and Packaging of Chemicals (GHS; UN GHS rev 1, 2005) and/or for Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)

3.2 Species:

The species of interest is humans. As hazard identification is primarily based on animal data, results of this ITS were compared to both human data (where available) and animal data (OECD TG 406 and/or OECD TG 429).

3.3 Additional information about the endpoint:

The adverse outcome pathway (AOP) for skin sensitisation has been formally described by the OECD (ENV/JM/MONO(2012)10/PART1 and /PART2) and the key events (KEs) described (see also section 5).

An AOP in the AOP-WIKI titled "Skin Sensitisation Initiated by Covalent Binding to Proteins" can be found under: https://aopkb.org/aopwiki/index.php/Aop:40). This defined approach uses non-animal methods to address three KEs defined by the AOP, namely protein binding, keratinocyte activation and dendritic cell activation.

4. Definition of the purpose and regulatory relevance

The "2 out of 3 - Sens ITS" was developed to allow the classification of substances as sensitisers or those that do not need to be classified as such, for regulatory purposes such as the Globally Harmonized System of Classification, Labelling and Packaging of Chemicals (GHS; UN GHS rev 1, 2005) and/or for REACH without the need to use animal-based test methods. Traditionally, hazard identification and classification for this endpoint is covered by animal tests [OECD TG 406 (induction and elicitation)/OECD TG 429 (induction only)]. The ITS is not designed to provide information on skin sensitisation potency and subcategorization of sensitisers into CLP/GHS categories 1A and 1B.

5. Rationale underlying the construction of the defined approach

Skin sensitisation is the result of a complex multifactorial sequence of events and has long been the focus of research. As a consequence, the chemical and biological pathways involved are relatively well characterised. Eleven steps and four KEs (protein binding, keratinocyte activation, dendritic cell activation, and antigen-specific T-cell proliferation with the generation of memory T-cells) leading to the AOP for skin sensitisation have been formally described and published by the OECD (ENV/JM/MONO(2012)10/PART1 and /PART2). The rationale for constructing this ITS, was to use well developed test methods and integrate data that covered the first three KEs of the AOP as these are essential for the sensitisation process, and subsequent events are dependent on these three events. A further prerequisite for use was also good quality animal and human data was available and that details had been published in peer-reviewed journals.

The molecular initiating event is defined as the covalent binding of the hapten to skin proteins. This step is evaluated using the Direct Peptide Reactivity Assay (DPRA; OECD TG 442C). Inflammatory and protective responses by the first cells coming into contact with the substance, the keratinocytes, are essential for downstream events to take place. Keratinocyte activation is evaluated via the Nrf2-ARE-based KeratinoSens^{TM} and LuSens assays (OECD TG 442D). Dendritic cells (DCs) transport the hapten to the regional lymph nodes, present the hapten on the cell surface and, when activated (mature DCs), are able to present the antigen in the proper context (upregulated cell surface markers, e.g. CD86) to activate naïve T-cells thereby triggering their proliferation. The potential of a substance to cause DC activation is assessed using the h-CLAT (OECD TG 442E) or (m)MUSST/U-SENS assays (OECD TG draft). These three events are essential to allow a DC to activate naïve T-cells, which then leads to KE4 – T cell proliferation and the development of effector T-cells (antigen presentation by mature DC resulting in sensitisation) or regulatory T cells (antigen presentation by immature DC resulting in tolerance). T-cell activation (KE 4) is dependent on the activation/maturation state of the antigen presenting cells (e.g. dendritic cells), i.e. sensitisation only takes place if the antigen presenting cells (APCs) presenting the antigen are activated, and this is dependent on the first three key steps. The read-outs and cut-offs to define a positive or negative result are used as described by the method developers and/or OECD TGs. The results are then used as a binary yes/no input to obtain the final prediction. In general, the results of two tests drives the prediction and conducting two of three tests, each addressing a separate KE, can be considered sufficient if the results are concordant.

Additional information can be gleaned from the third test, in particular when borderline results are obtained in at least one of the first two tests.

6. Description of the individual information sources used (see Annex II)

- **Direct Peptide Reactivity Assay** (DPRA; OECD TG 442C; KE1): Skin sensitisers are generally electrophilic and react with the nucleophilic moieties of proteins. The DPRA measures depletion of two peptides containing either cysteine or lysine residues due to covalent binding. The prediction model describes in OECD TG 442C is used to identify positive and negative results.

- **KeratinoSens™ and/or LuSens assay** (In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method OECD TG 442D; KE2); Keratinocytes harbouring a reporter gene construct react to possible sensitisers via the Nrf2-Keap1 pathway. Substances are considered to be positive in the KeratinoSens™ if they induce a statistically significant induction (1.5fold at viabilities $\geq$ 70%) of the luciferase gene over the vehicle controls. The prediction model describes in OECD TG 442C is used to identify positive and negative results.

- **(modified) Myeloid U936 Skin Sensitisation Test** [(m)MUSST; KE3] Activation of antigen presenting cells (APCs) is characterized by the up-regulation of CD86. The (m)MUSST is considered to be positive if CD86 induction exceeds 1.5fold at viabilities $\geq$ 70%.

- **Human cell-line activation test** (h-CLAT; OECD TG 442E; KE 3): Activation of antigen presenting cells (APCs) is characterized by the up-regulation of CD86 and/or CD54. The h-CLAT is considered to be positive if CD86 induction exceeds 1.5fold and CD54 exceeds 2.0fold at viabilities $\geq$ 50% when compared to the vehicle control.

- **OECD Toolbox Vers. 3.2** (possible peptide reactivity mechanisms)

7. Data interpretation procedure (DIP) applied

Within the ITS, three methods are used reflecting the first three KEs in the sensitisation process (protein binding, keratinocyte activation, DC activation). The DIP applied uses the readout of the prediction models of each of the individual test methods and/or information sources as defined by the method developer or test guideline (see also Annex II). The substance is then classified as a sensitiser or not for that specific method and KE (binary answer: yes or no; see also the reporting of the individual information sources attached for more details). The prediction model of the DIP then defines that two concordant results addressing two different KEs indicate the sensitising potential, i.e. two positive results indicate a sensitiser, two negative results indicate a non-sensitiser. This also implies that if the first two tests conducted yield a) concordant results, the third does not need to be performed, or b) discordant results in the first two tests necessitates conducting a third test addressing a third KE (Figure I.1).

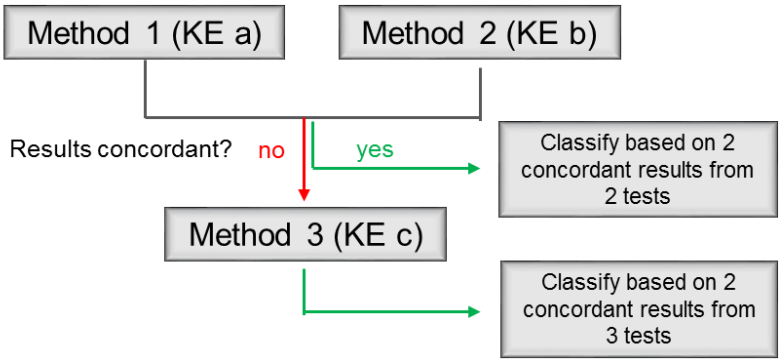


Figure I.1: Schematic representation of the "2 out of 3 – Sens ITS".

As there is no differential weighting of the individual test methods used and no predefined sequential order of testing, the order and information source from which data is obtained is not defined. Due to the higher complexity and resources needed (e.g. flow cytometer needed) to conduct the tests used for KE 3, the DPRA (KE1) and Nrf2-ARE-based tests (KE2) will usually be conducted first. The "2 out of 3 - Sens ITS" is quantitative in that the values from measurements are obtained in each of the individual test methods used. It is qualitative as these values play a secondary role as only binary (positive/negative) results drive the final prediction. The results from the "third" test can then be useful, e.g. to obtain a two out of three prediction when the first two tests are not concordant, or to further corroborate a prediction when assessing when borderline results are achieved.

8. Chemicals used to develop and test the DIP

8.1. Availability of training and test sets:

The "2 out of 3" approach is based on the predictivities achieved over time with an ever growing number of chemicals and not based on an in silico model or similar (similar to how the LLNA was validated; ICCVAM 1999). Therefore, the data is not separated into test sets and training sets. The initial test set of 54 chemicals was published in Bauch et al., 2012. This data set was expanded to encompass 145 chemicals in Natsch et al. 2013. In 2015, Urbisch et al. further increased the total number of evaluated chemical to 213. Data and predictivities of the individual tests and the "2 out of 3 - Sens ITS" approach were compared to available human and animal data. All relevant information can be easily obtained as open access Excel files at the publisher's website of the Urbisch et al. (2015) paper (http://www.sciencedirect.com/science/article/pii/S0273230014003092).

8.2. Selection of the training set and test set used to assess the DIP:

The test substances were selected based on the availability of high quality human and animal sensitisation data, the general availability of the test substances, coverage of a range of sensitisation potencies observed in vivo, coverage of a wide range of physicochemical, structural properties and substances classes (see also section 8.1), and care was also taken to include the performance standards used in the Local Lymph Node Assay (OECD TG 429).

8.3. Supporting information on the training and test sets:

The full set of substances (currently 213 substances) is publically available. Both the paper and Excel tables (supplementary data) with information on the 213 substances evaluated (Urbisch et al., 2015), such as CAS numbers, chemical structures, protein reactivity alerts based on the OECD QSAR toolbox (vers. 3.2), along with the results of the different tests and "2 out of 3 - Sens ITS" can be downloaded free of charge (open access) at the Regulatory Toxicology and Pharmacology Journal website (http://www.sciencedirect.com/science/article/pii/S0273230014003092#MMCvFirst)

8.4. Other information on the training and test sets:

For 208 substances, high quality LLNA data were described in the literature or data bases; LLNA data for a further 5 substances not previously published were also used. Human data was derived from published literature (e.g. Basketter et al., 2014). Of the 213 substances evaluated by Urbisch et al. (2015) 151 (71%) are considered to be sensitisers and 62 (29%) to be non-sensitisers according to available LLNA data. Human data were available for 114 of these substances. For all substances within this data compilation, results were available for at least two of the investigated non-animal test methods. The set of non-animal test data comprises data originating from the DPRA (results for 199 substances), KeratinoSens™ assay (results for 195 substances), LuSens assay (results for 77 substances), h-CLAT (results for 166 substances), MUSST (now renamed as USens; results for 145 substances) and mMUSST data (results for 65 substances). The following substance classes were assessed: Fragrances (n = 53), preservatives/disinfectants (n = 26), dyes (n = 12), monomers (n = 15), pesticide active ingredients (n = 9), solvents (n = 8), cosmetics ingredients (n = 14), pharmaceutical ingredients (n = 9), surfactants (n = 5), plasticizers (n = 3), food/feed ingredients (n = 4), other uses (n = 55).

9. Limitations in the application of the defined approach

The strengths and limitations are described in greater detail in the publications on the individual methods and for the ITS in Bauch et al., 2012, Natsch et al., 2013 and in particular in Urbisch et al., 2015 (open access; http://www.sciencedirect.com/science/article/pii/S0273230014003092). The following are the most prominent strengths and limitations:

Strengths:

- Non animal ITS
- Large reference data set now available (n=213; Urbisch et al., 2015)
- When compared to available human data, the ITS even achieves slightly better predictivities than LLNA when comparing to human data (accuracies: "2 out of 3 - Sens ITS": 88-91%; LLNA: 82%)
- Lower quantities of test substance are needed for testing compared to animal tests
- AOP based ITS, more mechanistic data less "black box" animal data

Limitations:

Technical limitations:

- Technical limitations exist, e.g. if substances or precipitates that are formed interfere with the detection system (bubbles formed by surfactants can interfere with flow cytometric detection in some cytometers, depletion of peptides not due to adduct formation, pigments could interfere with viability readouts)

- Physical state may preclude testing e.g. gases, highly lipophilic substances (cell culture). Substances with a high logP (e.g. exceeding 3.5 in the h-CLAT and 5.0 in the KeratinoSens assay) may pose problems due to the aqueous nature of the cell culture medium and solubility issues.

- Substances must be stable under test conditions e.g. the DPRA uses high alkaline conditions for lysine reactivity

- Complex mixtures, e.g. plant extracts or formulations are difficult to evaluate as molecular weights or molar equivalents are used in some tests

- Peptide depletion due to adduct formation cannot be differentiated from peptide depletion due to dimerisation or oxidation of the peptide

Substance related limitations:

- Substances with high cytotoxicity cannot always be tested to a sufficiently high concentration

- Prohaptens are identified if simultaneously prehaptens but not always reliably predicted if not (e.g. cells have a xenobiotic metabolism but these may be limited) (Fabian et al., 2013). In the "2 out of 3 – Sens ITS" pre- and prohaptens are identified with an accuracy of 81% (Urbisch et al., 2016). An expert group also reported that unless a substance is exclusively a prohapten, identification of sensitisers is sufficient (Casati et al., 2016).

- Depending on the protein binding mechanisms, the individual assays may have varying predictivities (see Urbisch et al., 2015 and also point 10 below)

- Substances that only react with lysine and not with cysteine can lead to false negative predictions as both the DPRA and KeratinoSens^{TM} use cysteine reactivity as a read-out.

10. Predictive capacity of the defined approach

The predictive capacities of this approach are discussed in detail in the Urbisch et al., 2015 paper (free article; http://www.sciencedirect.com/science/article/pii/S0273230014003092). Some information is given in Table I.1. Table I.2 gives the predictivities of the individual assays and combinations thereof compared to human data and animal data. The number of test substances evaluated varies depending the availability of the data sets.

Table I.1: Sensitisers and non-sensitisers among substances with LLNA data and with human data on skin sensitisation.

Chemical set and reference data	LLNA data	Human data
Sensitisers	151	75
Non-sensitisers	62	36
Borderlines	0	3
Total	**213**	**114**

Table I.2: Predictivities of the individual assays and combinations compared to human data and animal data.

Compared to human data		n	Cooper statistics [%]				
			Sensitivity	Specificity	Positive predictive value	negative predictive value	Accuracy
animal test	LLNA	111	91	64	84	77	82
"2 out of 3 – Sens ITS"	DPRA + KeratinoSens + h-CLAT	101	90	90	96	79	90
	DPRA + KeratinoSens + (m)MUSST	95	84	100	100	70	88
	DPRA + LuSens + h-CLAT	90	90	89	95	80	90
	DPRA + LuSens + (m)MUSST	75	87	100	100	75	91

Compared to LLNA data		n	Cooper statistics [%]				
			Sensitivity	Specificity	Positive predictive value	negative predictive value	Accuracy
"2 out of 3 – Sens ITS"	DPRA + KeratinoSens + h-CLAT	180	82	72	89	59	79
	DPRA + KeratinoSens + (m)MUSST	171	79	77	90	59	78
	DPRA + LuSens + h-CLAT	133	83	78	91	64	82
	DPRA + LuSens + (m)MUSST	126	84	84	93	69	84

The interchangeability of the test methods addressing a specific KEwas also analyzed. A set of 69 substances was tested in the KeratinoSens™ (Givaudan) and the LuSens assay (BASF). The same overall outcome for 61 of the 69 substances resulted in an interchangeability of 88%. In the "2 out of 3 – Sens ITS" the predictivities were almost identical indicating that these two methods can be used interchangeably. Regarding the h-CLAT (Kao and Shiseido) and the (m)MUSST (BASF and P&G), a common set of 105 substances was tested to investigate dendritic cell activation and revealed an lower interchangeability of these two tests of 72%. More details can be found in the Urbisch et al. 2015 paper.

In order to gain more insight into the predictive capacity of the methods and ITS for identifying sensitisers based on protein binding mechanisms, the OECD Toolbox version 3.2, which is freely available on the OECD website (http://toolbox.oasis-lmc.org/?section=download&version=latest) was used to define different mechanistic domains by probable protein-binding mechanisms. This approach shows that Michael acceptors, substances reacting in nucleophilic substitutions and quinone precursors were predicted with the

highest accuracies. In the domain of Schiff ‘base formers as well as in the group of substances with a lack of obvious alerts for peptide reactivity, accuracies were slightly decreased. In the domain of acylating agents, the ARE based assays show mechanistically justifiable decreased predictivities due to lack of Cys reactivity (Urbisch et al., 2015).

11. Consideration of uncertainties associated with the application of the defined approach

11.1 Sources of uncertainty

- This DIP has the purpose of skin sensitisation hazard identification for classification and labelling purposes. The use of this approach for potency assessments is not yet possible. The “2 out of 3 - Sens ITS” testing approach described is an AOP-based approach and addresses the first three KEs; KE4 is missing and no validated test methods are available for this endpoint. However, T-cells can develop as either T-effector cells (leading to sensitisation) or as T-regulatory cells (leading to tolerance). Therefore, the assumption is made that the events leading to antigen presenting cell (dendritic cell) activation is sufficient to drive the T-effector response leading to sensitisation. As the “2 out of 3 - Sens ITS” testing approach used is based on data obtained from over 200 chemicals and predictivities and there is no weighting of the results obtained by the individual methods or KEs covered. For this data set, when compared to human data, the accuracy of the “2 out of 3 - Sens ITS” testing approach is better than the typically used animal test, the LLNA, namely 88-91% depending on the combination of methods used and available data vs. 82% for the LLNA (see section 9 and Urbisch et al., 2015). Therefore, the confidence in the prediction for hazard identification is high when taking the limitations into account.

- The “2 out of 3 - Sens ITS” testing approach used in this defined approach has been used to test a wide range of different chemical types and mechanistic protein-binding domains. As stated in section 9 (limitations), the metabolic capacities of the cell-based methods is limited but not non-existent and a number of major xenobiotic enzymes are present. Still uncertainties exist when assessing prohaptens. There are also uncertainties with chemicals acting as acylating agents as Cys reactivity can be decreased (e.g. in favor of Lys reactivity) and this can lead to false negative results in the OECD TG 442C and OECD TG 442D methods.

- A further uncertainty is the quality of the reference data. Biological systems are prone to variations and mice or guinea pigs are not humans.

11.2 The information sources used within the defined approach

All test methods used have been shown to be reliable and the reproducibility is good. This is reflected in that the DPRA, KeratinoSens and h-CLAT assays are now described by OECD TGs 442C, 442D and 442E, and the LuSens assay has undergone inter and intralaboratory trials, was submitted to EURL JRC for assessment and is now in the ESAC review stage. The U937-based assays are published in peer-reviewed journals and a method similar to the mMUSST, the U-SENS assay has been submitted to ECVAM and the OECD.

11.3 Benchmark data used

The benchmark data that was used to develop the test methods is primarily data obtained from the murine LLNA. The LLNA has in cases been reported to lead to false positive predictions (e.g. surfactants and

some irritants); false negatives are rare with nickel salts being an example. In the Urbisch et al., 2015 study the LLNA had an accuracy of 82% when compared to human data. In the original validation study, the LLNA (and guinea pig tests) was reported to have an accuracy of 72% when compared to human data (Dean et al., 2001). Variability in the EC3 values of the LLNA has reported depending on vehicle used, etc. Therefore, the results of the LLNA, but also human data, harbour uncertainties. Furthermore, no systematic assessment of the applicability domains for the animal (LLNA or guinea-pig based tests) or human tests has been conducted to date.

11.4 Impact of uncertainty on the DIP's prediction

The different uncertainties can have an effect with the hazard potential for skin sensitisation being either under- or over-estimated. There is no formal integration of an uncertainty assessment into the "2 out of 3 - Sens ITS" testing approach prediction model. The impact of uncertainties in the benchmark data will also lead to uncertainties in the predictions made but as much of the benchmark data is derived from the test methods typically used in the regulatory setting for hazard identification (OECD TG 429: LLNA, OECD TG 406: GPTs), these uncertainties are already accepted for this purpose. When taking the limitations and the high degree of accuracy into account, confidence in the prediction is high.

12. References

Adverse Outcome Pathway (AOP) (2015). Wiki: https://aopkb.org/aopwiki/index.php/Aop:40

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.

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 sensitizing potency, Dermatitis, 25(1):11-21.

Casati S, Aschberger K, Asturiol D, Basketter D, Dimitrov S, Dumont C, Karlberg AT, Lepoittevin JP, Patlewicz G, Roberts DW, Worth A. (2016) Ability of non-animal methods for skin sensitisation to detect pre- and pro-haptens: https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-status-reports/pre-prohapten-workshop-report)

Dean JH, Twerdok LE, Tice RR, Sailstad DM, Hattan DG, Stokes WS. (2001). ICCVAM evaluation of the murine local lymph node assay. Conclusions and recommendations of an independent scientific peer review panel, Regul Toxicol Pharmacol, 34:258-73.

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. http://link.springer.com/article/10.1007%2Fs00204-014-1382-8

The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds: The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and theNational Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICEATM) (1999). https://ntp.niehs.nih.gov/iccvam/docs/immunotox_docs/llna/llnarep.pdf

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

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

OECD Toolbox ver 3.2: http://toolbox.oasis-lmc.org/?section=download&version=latest

OECD test guidelines (TGs): OECD TG 406 (skin sensitization; guinea pig-based sensitization tests); OECD TG 429 (local lymph node assay); OECD TG 442C (In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA)); OECD TG 442D (In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method); OECD draft TG (In Vitro Skin Sensitisation: human Cell Line Activation Test (h-CLAT)

Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods, Regul Toxicol Pharmacol, 71:337-51. http://www.sciencedirect.com/science/article/pii/S0273230014003092

Urbisch D, Becker M, Honarvar N, Kolle SN, Mehling A, Teubner W, Wareing B, Landsiedel R. (2016). Assessment of pre- and pro-haptens using non-animal test methods for skin sensitization. Chem Res Toxicol, [Epub ahead of print] PMID: 27070937.

13. Supporting information

All data (both the publication and test substance information as Excel files) from the Urbisch et al., 2015 paper (open access) can be downloaded at the Regulatory Toxicology and Pharmacology website (http://www.sciencedirect.com/science/article/pii/S0273230014003092). Therefore, no tables are attached.

14 .Abbreviations and definitions

AOP: Adverse outcome pathway

APC: Antigen presenting cell

ARE: Antioxidant response element

DC: Dendritic cell

DIP: Data integration process

DPRA: Direct peptide reactivity assay

GHS: Globally Harmonized System of Classification, Labelling and Packaging of Chemicals

h-CLAT: Human cell-line activation test

ITS: Integrated testing strategy

KE: Key event

LLNA: Local lymph node assay

(m)MUSST: (modified) Myeloid U937 cell-line activation test

REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals

TG: Test guideline

CASE STUDY II

Sequential Testing Strategy (STS) for hazard identification of skin sensitisers

1. Summary

The RIVM Sequential Testing Strategy (STS) is constructed as a tiered approach with a decision point at the end of each tier, allowing stepwise and efficient information gathering. The individual tiers include information sources that address the first four key events (KEs) as defined in the OECD AOP for skin sensitisation. Tier 1 starts with *in silico* methods that predict the probability of a positive or negative LLNA result. In case the results are equivocal peptide reactivity is measured, addressing the molecular initiating event. Tier 2 addresses KE2: events in keratinocytes using two different *in vitro* keratinocyte assays. Tier 3 addresses KE3: events in dendritic cells.

Tools and test methods that are used in this assay are: a Bayesian QSAR approach and the Direct Peptide Reactivity Assay (DPRA) in Tier 1, Keratinosens™ and HaCaT gene signature in Tier 2 and h-CLAT assay in Tier 3. The outcome can be used to predict the skin sensitising potential of a substance, but does not allow potency estimation. In this case study the STS is described in detail and the predictive capacity and a description of limitations and uncertainties of this strategy is based on the training set of 41 substances.

2. General information

2.1 Identifier:

RIVM Sequential Testing Strategy (STS) for hazard identification of skin sensitisers

2.2 Date:

16 March 2016

2.3 Author(s) and contact details:

Janine Ezendam and Emiel Rorije

National Institute for Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, the Netherlands.

Janine.Ezendam@rivm.nl, telephone: +31 30 274 3447

2.4 RF update(s): Version 1.

2.5 Reference to main scientific papers:

Key paper:

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(3):371-9.

Other papers:

Ashikaga, T., Sakaguchi, H., Sono, S., Kosaka, N., Ishikawa, M., Nukada, Y., Miyazawa, M., Ito, Y., Nishiyama, N. & Itagaki, H.(2010). A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA), Altern Lab Anim, 38, 275-84.

Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H. & Toyoda, H. (2006). Development of an in vitro skin sensitization test using human cell lines: the human Cell Line Activation Test (h-CLAT). I. Optimization of the h-CLAT protocol. Toxicol In Vitro, 20, 767-73

Bauch, C., Kolle, S. N., 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.

Gerberick, G. F., Vassallo, J. D., Foertsch, L. M., Price, B. B., Chaney, J. G. & Lepoittevin, J. P. (2007). Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach, Toxicol Sci, 97, 417-27.

Natsch, A., Ryan, C. A., 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.

Rorije, E., Aldenberg, T., Buist, H., Kroese, D. & Schuurmann, G. (2013). The OSIRIS Weight of Evidence approach: ITS for skin sensitisation, Regul Toxicol Pharmacol, 67(2):146-56.

Van der Veen, J. W., Pronk, T. E., van Loveren, H. & Ezendam, J. (2013). Applicability of a keratinocyte gene signature to predict skin sensitizing potential, Toxicol In Vitro, 27, 314-322.

2.6 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™.

The h-CLAT assay is covered by intellectual property rights in Japan.

The HaCaT gene signature is an RIVM in house developed model measuring gene regulation of a specific set of genes in human keratinocytes cells. The test will not be further developed as it was meant as a proof of principle.

The DEREK Nexus software is not-for profit software, but requires a license as it is only available to members of the company Lhasa Ltd. MultiCASE is commercial QSAR software, but the specific MultiCASE skin sensitisation model predictions used here are openly available from a database of QSAR

predictions hosted by the Danish EPA. The other two QSAR software packages (OECD QSAR Toolbox, CAESAR/VEGA) are freeware and can be downloaded for free from the internet.

3. Endpoint addressed

3.1 Endpoint:

This STS predicts the skin sensitisation potential of a substance. This endpoint is currently measured using animal models as described in different OECD test guidelines (TGs 406, TG 429, TG 442A and B).

3.2 Species:

This STS predicts the human hazard.

3.3 Additional information about the endpoint:

Mechanisms involved in skin sensitisation have been captured in the AOP for skin sensitisation, which is described in detail in AOP no. 40 (https://aopwiki.org/wiki/Aop:40) (OECD, 2012).

The skin sensitisation AOP identifies key events (KEs) with subsequently: KE1, covalent binding to skin proteins (termed haptenation) of either the parent substance or of transformation products following abiotic/metabolic activation, which is postulated to be the molecular initiating event (MIE); followed by KE2, activation of epidermal keratinocytes; KE3, activation (maturation) and mobilisation of Langerhans cell and dermal dendritic cells (DC); and KE4, DC-mediated antigen presentation to naïve T-cells and proliferation /activation of allergen specific T-cells.

4. Definition of the purpose and regulatory relevance

The purpose of this defined approach is to provide information on the skin sensitisation hazard of a test substance. This information can be used for regulatory purposes, e.g. for classification and labelling under the Globally Harmonized System of Classification, Labelling and Packaging of Chemicals (UN GHS), for REACH and for the Cosmetics Regulation. This approach does not provide information on the skin sensitising potency and therefore cannot be used tp sub-categorise of substances into CLP/GHS categories 1A and 1B, because it does not.

5. Rationale underlying the construction of the defined approach

This STS is constructed in such a way that it allows stepwise gathering of information on skin sensitisation hazard using a tiered-approach (Figure II.1). The tiers address the MIE (KE1), KE2, KE3 and KE4. Decision criteria are included after each tier in order to enable targeted and efficient information gathering and to avoid redundant testing.

- **Tier 1** starts with a battery of in silico methods: a Bayesian approach to combining QSAR predictions from 4 non-commercial QSAR models (MultiCASE, CAESAR, DEREK and OECD QSAR Toolbox). The first three in silico methods predict the probability that a substance will be tested positive or negative in the LLNA (MultiCASE, CAESAR) or in humans (DEREK). The predictivity of the individual models in the Bayesian decision models was measured against the outcome of the LLNA test. As such, it addresses KE4 (T cell activation) of the AOP. The OECD

QSAR Toolbox has a predictive profile for protein binding and therefore addresses the MIE. The rationale for starting with the in silico approach is to reduce the number of assays that need to be performed. The probability of the prediction is used to decide if additional testing in this tier is needed. If the probability is above the probability threshold (80% probability of being correct for a positive and 90% for negative prediction) the substance proceeds to Tier 2. If it is below the threshold further testing in Tier 1 is needed. The DPRA is used for the equivocal results, since it addresses the MIE protein binding in chemico.

- **Tier 2** addresses KE2 using the *in vitro* keratinocyte assays KeratinoSens and HaCaT gene signature. The results of Tier 1 are used to select the appropriate *in vitro* assay. If in Tier 1 a sensitiser is indicated, the substance is tested in the KeratinoSens assay, which is the assay with the lowest level of false-positive results and thus the highest positive predictive value (PPV). When a non-sensitiser is indicated in Tier 1, the HaCaT gene signature is selected, being the assay with the lowest number of false-negative results and thus the highest negative predictive value (NPV). When Tier 1 and 2 give concordant results, the substance is classified accordingly and no further testing is needed. In case of discordant results, the substance is tested in Tier 3.

HaCaT gene signature is an in-house method. This STS is a conceptual approach and other test methods can be used in this strategy to replace e.g. the HaCaT gene signature. So, a test method that address KE2 and has good negative predictive value (comparable to the HaCaT gene signature assay) would be suitable to replace the HaCaT gene signature in this strategy.

- **Tier 3** addresses KE3, dendritic cell activation. The h-CLAT is only performed for substances with discordant results in Tiers 1 and 2.

- After Tier 3 the final decision is based on the majority voting principle, i.e. two out of three tests need to give concordant results to reach a prediction.

This STS allows inclusion of other information sources once they become available. If these methods are more suitable to test certain substances, they can replace a test method that is currently included. The principles that are applied in this strategy can be applied to those novel methods once they become available. Their place in this strategy is based on the KE that is addressed and in case of novel keratinocyte assays; the predictive performance is used to decide on the optimal place within Tier 2.

6. Description of the individual information sources used

Detailed information on each of the individual information sources can be found in Annexe II.

A. Bayesian QSAR approach

In this approach a battery of four *in silico* models MultiCASE, CAESAR, DEREK and OECD QSAR Toolbox, is used to generate a battery-prediction. This approach is described by Rorije et al. 2013, and the specific battery used here is described in van der Veen et al. 2014. The specificity and sensitivity of each method are used in a Bayesian analysis, taking into account the applicability domain information given by CAESAR and MultiCASE.

- **MultiCASE** generates QSAR models based on substructure fragments linked to biological activity. The Multi-CASE implementation for skin sensitisation from the Danish EPA is used. Predictions with this model are available in a database on the internet and in the

OECD QSAR Toolbox. The outcome of the model is positive (requires classification as skin sensitiser in GHS) or negative (no classification needed).

- **CAESAR** uses atom centred fragments as descriptors in a multivariate statistical model. The model gives a prediction of active or inactive (as skin sensitiser), together with applicability domain information. The model was optimised to reproduce the outcome of the LLNA test.

- **DEREK** knowledgebase from Lhasa Ltd. is a collection of structural alerts linked to skin sensitisation. The model only identifies skin sensitisers and is not meant to identify non-sensitisers. However, in this approach also the probability of being correct in predicting non-sensitisation based on the absence of any DEREK alert is used in the battery approach. DEREK is predicting human skin sensitisation potential.

- **OECD QSAR Toolbox** contains an implementation of a set of protein binding reactivity alerts. Again the model is only meant to identify (potential) skin sensitisers, but the absence of any of the alerts in this profile was used as a prediction of non-sensitisation, with its related (low) probability of being correct.

The Bayesian analysis provides a probability that the QSAR battery prediction is correct in predicting the outcome of the LLNA test. Threshold values are applied to determine whether the analysis should proceed to the next tier. The threshold values are >80% for a positive conclusion and >90% for a negative conclusion. These threshold values are based on the reliability with which the GPMT test predicts the LLNA outcome (or vice versa) in the official LLNA validation study (NICEATM-ICCVAM, 1999).

If there are insufficient or conflicting results from the battery of QSAR models, these thresholds will not be reached and the results are considered equivocal. In that case follow up testing, first in the form of the DPRA, is needed and subsequently further *in vitro* testing will be needed in the next tier(s).

B. DPRA (OECD Test Guideline 442C)

DPRA measures *in chemico* binding to synthetic cysteine and lysine peptides. The read-outs are % peptide depletion of cysteine and lysine. The prediction model described in TG 442C is used to decide if a substance is positive or negative.

C. KeratinoSens™ ARE-Nrf2 Luciferase Test Method (OECD Test Guideline 442D)

KeratinoSens is a reporter gene assay that measures activation of the Nrf2-Keap1 pathway in HaCaT cells. The read-outs are ARE-regulated luciferase induction and cytotoxicity. The prediction model described in TG442D is used to decide if a substance is positive or negative.

D. HaCaT gene signature

The HaCaT gene signature measures gene expression of 10 genes that are able to distinguish skin sensitisers from non-sensitisers. The read-out is gene expression of those 10 genes. For the prediction, a combination of the classification algorithms of Random Forest (RF), Support Vector Machine (SVM) and

Prediction Analysis for Microarrays for R (PAM-R) are used. The changes in gene expression are compared to the gene profiles of the training set which is based on a DNA microarray study. A test chemical was assigned to the most predominant class indicated by these algorithms.

E. h-CLAT

h-CLAT is an assay that measures activation of dendritic-like cells. The read-out is the upregulation of the cell surface markers CD54 and CD86 by measuring relative fluorescence intensitity (RFI) by flow cytometry. The prediction model described in the publications of Ashikaga et al. (2006, 2010) is used to decide if a substance is positive or negative.

7. Data interpretation procedure applied

The process to arrive at a prediction is illustrated the workflow presented in Figure II. 1. The data interpretation is qualitative and results in classification as a skin sensitiser or not.

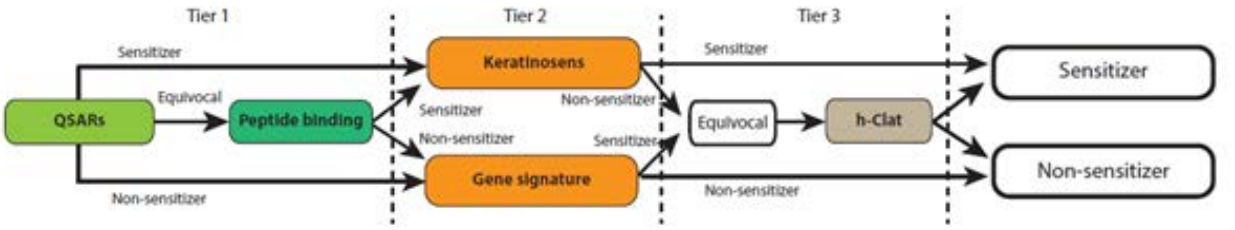


Figure II. 1: Schematic presentation of the RIVM STS.

Short description of the process applied in each tier to reach a decision on a test chemical.

Tier 1:

- All test chemicals are evaluated in the Bayesian QSAR approach. According to the prediction model a chemical is rated positive, negative or equivocal.
- Positive or negative results: test chemical proceeds to Tier 2.
- Equivocal result: the test chemical is tested in the DPRA. According to the prediction model the chemical is rated positive or negative and proceeds to Tier 2.

Tier 2:

- Test chemicals that are rated positive in Tier 1 are tested in the Keratinosens™ assay.
- Test chemical that are rated negative in Tier 1 are tested in the HaCaT gene signature.

- Based on the results of Tier 1 and 2 a decision is made:
    - In case of two positive calls: substance is classified as a skin sensitiser
    - In case of two negative calls: substance is not classified
    - In case of discordant results: test chemical proceeds to Tier 3.

Tier 3:

- Test chemicals that give discordant results in Tiers 1 and 2 are tested in the h-CLAT assay.
- Test chemicals are classified according to any two congruent results obtained in the three tiers (majority voting)
    - If at least two of the three tiers generate a positive result: substance is classified as a skin sensitiser
    - If at least two of the three tiers generate a negative result: substance is not classified.

8. Chemicals used to develop and test the DIP

8.1 Availability of training and test sets:

The full training set of 41 test chemicals is summarized in Appendix II.1 and referenced in Van der Veen et al., 2014. The STS has not yet been evaluated with a test set.

8.2 Selection of the training set and test set used to assess the approach:

The training set consists of 27 sensitising chemicals for which human evidence on skin sensitisation potential is available and 14 non-sensitising chemicals. The set of skin sensitisers represents various potency classes (from weak to strong), pre- and prohaptens, covers a range of physicochemical and structural properties and a some chemicals that are known false-positives or false-negatives in the LLNA is included as well.

8.3 Supporting information on the training and test sets:

Supporting information on chemicals in the training set is provided in Appendix II.1. This includes CAS registry numbers, classifications for each chemical (both human and LLNA classification). In Appendix II.2 the predictions of the individual data sources for each chemical are provided and Cooper statistics are summarized in Appendix II.3. Details on the prediction of the RIVM STS for the training set are reported in Appendix II.4.

8.4 Other information on the training and test sets:

Not applicable.

9. Limitations in the application of the defined approach

The limitations of this approach are not studied in great detail and are confined to the (limited) training set that was tested. In general, the STS is applicable to low-molecular-weight chemicals. Its applicability for polymers, mixtures or engineered nanomaterials has not been tested. Other types of chemicals may fall outside the applicability domain of the STS as well, for example due to technical limitations of the individual data sources (see Annexe 2 for more details).

Technical limitations:

- Technical limitations exist, e.g. if substances or precipitates that are formed interfere with the detection system (bubbles formed by surfactants can interfere with flow cytometric detection in some cytometers, depletion of peptides not due to adduct formation, pigments could interfere with viability readouts).

- Physical state may preclude testing e.g. gases, highly lipophilic substances (cell culture). Substances with a high logP (e.g. exceeding 3.5 in the h-CLAT and 5.0 in the KeratinoSens assay) may pose problems due to the aqueous nature of the cell culture medium and solubility issues.

- Substances must be stable under test conditions e.g. the DPRA uses high alkaline conditions for lysine reactivity.

- Chemicals should be soluble in acetronitrile or water or in a mixture of both for the DPRA.

- In the DPRA, peptide depletion due to adduct formation cannot be differentiated from peptide depletion due to dimerization or oxidation of the peptide.

Substance related limitations:

- Substances with high cytotoxicity cannot always be tested to a sufficiently high concentration.

- Substances that only react with lysine and not with cysteine can lead to false negative predictions as both the DPRA and KeratinoSensTM use cysteine reactivity as a read-out.

- Prohaptens require metabolic conversion. The in vitro assays in this STS have limited metabolic capacity, although most prohaptens tested so far could be detected. Prohaptens cannot be assessed in the DPRA. Due to the structure of this defined approach (majority voting), putative prohaptens in our training set, e.g. eugenol, resorcinol, were correctly identified.

When information from different test methods is integrated, it is important to estimate if the individual limitations can be overcome in the strategy. For example, the Bayesian QSAR only predicts a subset of the substances with sufficient probability, because many substances fall outside the applicability domains. This limitation can be overcome, when these substances can be tested in the DPRA as is true for the substances in this training set. Substances that fall outside the applicability domains of both the in silico methods and the DPRA cannot be tested in Tier 1. More chemicals need to be tested to evaluate if information gathered in Tiers 2 and 3 is sufficient for such substances.

The limitations are mainly based on what is known for the individual data sources and not for the STS itself. The STS has been designed in such a way that misclassifications are minimized in the second tier. More data is needed to see if this also works successfully for a larger set of substances.

The 2 out of 3 approach means that not necessarily all KEs of the AOP are tested and in concordance (all KE positive for a positive prediction, all KE negative for a negative prediction). However, a substance may be misclassified in one of the test methods, due to technical limitations of this specific test, whereas the other 2 test methods are able to identify the sensitisation potential of this substance correctly. Hence, the AOP provides a backbone for the testing strategy, but insight in the individual test methods is needed as well to understand and explain discordant results. For most of these individual test methods some limitations are already well understood, but a complete picture is not yet available.

10. Predictive capacity of the defined approach

The predictive capacity of the defined approach when applied to the training set substances is summarised in Table II. 1. The accuracy of the STS compared to human data is 95.1%, sensitivity and specificity are 96.3% and 92.8%.

Table II. 1. Summary of prediction of the RIVM STS approach compared to human data.

	Percentage	Total number of chemicals
Accuracy	95.1% (100%[1])	39/41 (39/39[1])
Sensitivity	96,3% (100%[1])	26/27 (26/26[1])
Specificity	92,8% (100%[1])	13/14 (13/13[1])

[1]statistics are calculated with the assumption that the two substances showing discordant results in Tier 1 and 2 for which no h-CLAT information was present are predicted incorrectly. If these two substances would be considered as "equivocal", and would not be included in the calculation of the predictivities, the remaining 39 substances are predicted 100% correct by the RIVM STS approach. These numbers are given between brackets.

The STS failed to provide a prediction for two substances in the training set: the skin sensitiser triisobutylphosphate and the non-sensitiser hexaethylene glycol monodecyl ether. The reason for this is that for these two substances no results from the h-CLAT were available in the publically available literature and for this STS we used published h-CLAT data, rather than generating these data ourselves. For these two chemicals, discordant results were obtained in Tiers 1 and 2, hence the h-CLAT assay was required for the final decision-making. Since h-CLAT data were unavailable, we considered that these substances could not be predicted correctly in our STS. Notably, in the LLNA, triisobutylphosphate and hexaethylene glycol monodecyl ether were misclassified as well.

Overall, this defined approach is able to accurately distinguish skin sensitisers from non-sensitisers, because all other substances in the dataset were identified correctly, including the other substances that are false-positive or false-negative in the LLNA. It is important to further test the predictive capacity of this defined with a large test set of chemicals.

11. Known uncertainties associated with the application of the approach

11.1 Sources of uncertainty

11.1.1 Structure

- The STS structure uses (in the end, if all 3 tiers are filled with tests) a 2 out of 3 majority voting to come to a conclusion. This is not in line with the (strict) interpretation of the AOP, as one would expect that if one of the tests is negative, this means a breach of the toxicity pathway, and an adverse effect is not expected. This means that substances for which it can be expected that one of the tests performs inadequate (for example the DPRA test that lacks metabolism is not able to identify pro-haptens) should therefore actually have less weight in the procedure. This is a refinement which can be used to explain the test which is disagreeing, for those substances for which the final, two out of three, decision criterion is needed. A mechanistic explanation for the reason that one test in the STS is discordant actually improves the certainty that the overall decision (based on less than three tests) is correct.

11.1.2 Information sources

- The variability (reproducibility in the individual tests) is not explicitly taken into account. It could serve as an argument to lessen the weight of a test in the case where one of the tests is discordant with the other tests/tiers.

11.1.3 Benchmark data.

- The RIVM STS is used to predict human skin sensitisation. The *in vitro* methods are based on human cell lines. However, the thresholds used for the Bayesian QSAR battery are based on the probability that a substance will be positive in the LLNA test. Overall the RIVM STS will not be able to have less uncertainty that the (regulatory accepted) uncertainty with which the LLNA test is able to predict human skin sensitisation incidence. The statistical performance of the STS has been calculated with regard to the human skin sensitisation potential, and shows that a number of false positives and false negatives in the LLNA (included in the test data set) are predicted correctly. The statistics for the STS in predicting human skin sensitisation potential are better than the LLNA, for this specific data set.

11.2 Impact of uncertainty on the DIP's prediction

Does the STS prediction for a new chemical include an assessment of uncertainty? The RIVM STS uses a 2 out of 3 decision if information for all tiers is present. A qualitative assessment of the uncertainty can be performed as the prediction for substances for which all methods are in agreement will have less uncertainty. Also, if the disagreement of one of the information sources can be explained mechanistically (e.g. by absence of metabolic capacity in one test and not others) this will lower the uncertainty connected to a prediction based on discordant information sources. The uncertainty of the STS is thought to be underestimated by the statistical performance of the STS on the training/validation data set, as this data set is limited in size. It does represent a large variety of structures, including known false negatives and false positives in the LLNA, but it is also a dataset containing well-known sensitisers. Specifically the in silico tier (Bayesian decision model based on QSAR results) will therefore probably show very good predictivity for this specific dataset, which might be lower (more uncertain) for new / unknown chemicals.

12. References

Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H. & Toyoda, H. (2006). Development of an in vitro skin sensitization test using human cell lines: the human Cell Line Activation Test (h-CLAT). I. Optimization of the h-CLAT protocol. Toxicol In Vitro (20, 767-73).

Ashikaga, T., Sakaguchi, H., Sono, S., Kosaka, N., Ishikawa, M., Nukada, Y., Miyazawa, M., Ito, Y., Nishiyama, N. & Itagaki, H. (2010). A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA), Altern Lab Anim, 38, 275-84.

NICEATM-ICCVAM, (1999). The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICEATM), NIH Publication No. 99-4494.

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

Rorije, E., Aldenberg, T., Buist, H., Kroese, D. & Schuurmann, G. (2013). The OSIRIS Weight of Evidence approach: ITS for skin sensitisation, Regul Toxicol Pharmacol, 67(2):146-56.

Van der Veen, J. W., Pronk, T. E., van Loveren, H. & Ezendam, J. (2013). Applicability of a keratinocyte gene signature to predict skin sensitizing potential, Toxicol In Vitro, 27, 314-322.

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(3):371-9.

13. Supporting information

Please see section 8.3.

14. Abbreviations and definitions

AOP	Adverse Outcome Pathway
DPRA	Direct Peptide Reactivity Assay
GHS	Globally Harmonised System for Classification and Labeling
h-CLAT	human Cell Line Activation Test
KE	Key event

MIE	Molecular Initiating Event
NPV	Negative Predictive Value
PPV	Positive Predictive Value
STS	Sequential Testing Strategy
QSAR	Quantitative Structure Activity Relationship

CASE STUDY III

A non-testing Pipeline approach for skin sensitisation

1. Summary

Although this template was designed and constructed to report DIPs per se, as is illustrated by this particular example, it can be shown that the template is flexible enough to be used to document an IATA that is not algorithmic in nature.

This template describes a defined approach to testing and assessment as part of an IATA that exploits non-testing approaches in conjunction with available experimental data to make a determination of likely skin sensitisation potential and, if possible, potency. If the information is insufficient for the decision context needed, directed testing is suggested. The type of testing will be dependent on the alerts and/or reaction domains that are flagged and what physchem information is available that can help inform whether testing limitations may be encountered. The approach relies upon the wealth of structure-activity knowledge of skin sensitisers that has been captured in SARs available in tools such as the OECD QSAR Toolbox as well as commercially based expert systems such as TIMES-SS.

2. General information

2.1 Identifier:

A non-testing pipeline approach for skin sensitisation.

2.2 Date:

22 April 2016.

2.3 Author(s) and contact details:

Grace Patlewicz, email: gpatlewicz@gmail.com or tier.grace@epa.gov

2.4 Template update(s): Version 1.

2.5 Reference to main scientific papers:

Patlewicz G, Kuseva C, Kesova A, Popova I, Zhechev T, Pavlov T, Roberts DW, Mekenyan O. (2014). Towards AOP Application – implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitisation.

2.6 Proprietary aspects:

The OASIS pipeline tool as described specifically in the associated manuscript is proprietary in terms of the software implementation. Aside from TIMES-SS, many of the elements within the pipeline

however rely on published data, what is freely available within the OECD QSAR Toolbox or other QSARs that have been published in the literature.

3. Endpoint addressed

3.1 Endpoint:

Skin sensitisation as principally assessed by the endpoints measured in the Local Lymph Node Assay (LLNA). The LLNA is described by OECD TGs 429, 442A and 442B. The workflow aims to predict the skin sensitisation potential based on a number of different information sources, some of which may map to the LLNA. The Guinea Pig Maximisation Test (GPMT) and Buehler Test are described in OECD TG406.

3.2 Species:

Principally the mouse and secondary the guinea pig.

3.3 Additional information about the endpoint:

Skin sensitisation is an endpoint that has been well studied. An Adverse Outcome Pathway (AOP) was constructed to structure and characterise the information underpinning the induction of skin sensitisation which was published by the OECD in 2012.

4. Definition of the purpose and regulatory relevance

The intended purpose will be context dependent on the substance under evaluation and the breadth and quality of information available. Nominally the DIP has been structured to be used to determine the skin sensitisation potential (i.e. hazard identification) that could be used in a hazard classification and labelling as required by US regulatory authorities or GHS. In certain circumstances, it may be sufficient to permit classification of skin sensitisers in terms of GHS potency categories. The pipeline software implementation provides a preliminary assessment of skin sensitisation potential only which could meet the needs of screening and prioritisation.

5. Rationale underlying the construction of the defined approach

The workflow (pipeline) aims to articulate the different considerations a risk assessment practitioner might take into account when evaluating a given substance for its sensitisation potential and potency. Many of the information sources reflect the spirit of the endpoint guidance reflected in the REACH technical guidance for this endpoint. The workflow considers the physical state of the substance (which can be predicted by reference to melting point, boiling point and vapour pressure in the absence of measured data) and other physical properties (e.g. pKa) that may render any testing either unnecessary or, at minimum, facilitate the interpretation of existing experimental data, especially that generated by in chemico or in vitro means. This provides an indication of whether the dermal route of entry is a relevant route to be considering. Physical properties such as pKa will indicate whether a substance is likely to ionise and behave as a strong acid or base.

The workflow proposes that all available existing experimental data should be gathered together and evaluated in conjunction with what is understood about the substance on the basis of its chemistry, and in concert with where the information may map within the associated AOP. Existing SAR

information as encoded in the OECD Toolbox (in the form of profilers e.g. for protein binding) will be helpful to provide a perspective of the likely reaction chemistry as well as whether any simulated abiotic or metabolic transformations are relevant. This type of information is also essential to evaluate any experimental skin sensitisation data in the appropriate context. E.g. hard electrophiles such as Schiff base formers could give rise to false negative predictions in a glutathione depletion assay.

Existing experimental data could map to various key events (KEs) in the AOP and are weighed based on the strength of the key event relationship as well as the proximity of the specific KE to the adverse outcome (AO). At present the strength of those relationships is only qualitative in nature.

The quality of the information and its relevance to the substance of interest are carefully evaluated. Examples of issues to note might include whether there are technical limitations to the assays that would impact the interpretation of the outcomes (e.g. highly hydrophobic substances), or whether volatility may affect the results (as is the case in the LLNA where test substance could evaporate during the course of the study). If the information collected is sufficient for the decision purpose in mind, then no further action might be merited. The information and associated considerations should be conveniently captured in the modified weight of evidence (WoE) table (see section 7, Table III.1) to enable such a determination to be made. If the available information gathered is insufficient for the decision, then the use of expert systems such as TIMES-SS, mechanistic read-across using analogous substances from the OECD QSAR Toolbox in concert with information from mutagenicity data, where correlations have been identified, might result in sufficient confidence to make a decision. Generating new in chemico or in vitro data for the various KEs can then be considered with animal testing as a last resort. As such the rationale for this defined approach can be likened to a guided WoE assessment exploiting existing information and non-testing approaches each of which characterise various key events within the AOP.

6. Description of the individual information sources used (see Annex II)

More detailed descriptions of the information sources are provided in the Annex II.

The information sources relied upon are as follows:

- Physico-chemical properties

These properties are either collected from public data sources such as the OECD Toolbox or other information sources available online –eChemPortal etc. In absence of measured property information available QSAR tools such as those implemented in the OECD Toolbox or developed elsewhere e.g. Danish QSAR database could be used to generate predictions.

pKa (relevant for substances with ionisable groups) – if pH information is available, this can be used to identify strong bases or acids

Vapour Pressure – vapour pressures in excess of 10-4 mmHg may be indicative of a gaseous substance

Melting Point – values less than 25 deg C would be indicative of a liquid

Boiling Point – values less than 25 deg C would be indicative of a gas

Molecular Weight – substances with values of greater than 500 daltons may be worth closer examination (since at these apparent extremes, there may be technical difficulties with experimental sensitisation non-animal approaches)

Log Kow – substances with values less than -2 or greater than 4 may warrant closer examination (since at these apparent extremes, there may be technical difficulties with experimental sensitisation non-animal approaches)

- Non-testing approaches:

Non-testing approaches to identify relevant transformation products

These simulators are freely available within the OECD Toolbox as well as in TIMES-SS and this OASIS pipeline.

These properties are important to consider to determine whether the substance under evaluation is likely to act directly as a sensitiser or will require some metabolic or chemical activation. For example, does the substance degrade through hydrolysis – e.g. acylating agents compete between forming a protein adduct or hydrolysing with water (a nucleophile) to form the associated degradation product. Some substances are prone to oxidation in the air forming unstable hydroperoxides which can be sensitising or some may become oxidised enzymatically or abiotically to form reactive quinones.

Simulator for Hydrolysis – prediction of likely metabolites form as a result of hydrolysis

Simulator for Autoxidation – prediction of potential metabolites formed as a result of air oxidation

Simulator for Skin metabolism - prediction of potential metabolites formed as a result of metabolism

- Selected Non-testing approaches most closely anchored to the MIE

The MIE represents the covalent binding that occurs between an electrophilic substance and the skin protein. There is a wealth of literature that describe SARs that have been derived from analysing skin sensitisation data. Many of these SARs have been embedded in the OECD Toolbox as profilers but additionally exist in other tools such as expert systems like Derek Nexus. These profilers help to identify electrophilic features known to be correlated with skin sensitisation. Some of the profilers solely describe SARs that are substantiated by skin sensitisation data whereas others also contain theoretical SARs grounded in established organic chemistry principles but not necessarily supported by experimental sensitisation data. The first three profilers below are these SARs. The last profiler listed is a rulebase of SARs that have been extracted from evaluating available experimental DRPA data. Structural features that give rise to activity in the DRPA have been extracted and implemented as a new profiling scheme.

Protein Binding Alerts by OASIS v1.3 – reaction mechanistic domain information which may or may not be substantiated by skin sensitisation data.

Protein Binding Alerts by OECD – reaction mechanistic domain information which may or may not be substantiated by skin sensitisation data.

Protein Binding Alerts for Skin Sensitisation by OASIS v1.1 – reaction mechanistic domain information underpinned by skin sensitisation data.

Profilers for DPRA Cysteine (Lysine) peptide depletion – SARs extracted from substances tested in the DPRA.

- Non-testing approaches capable of predicting the AO or the EC3 in the LLNA (KE4)

TIMES-SS (if available) – provides a semi-quantitative estimate of potency – a QMRF is available within the TIMES software program itself and within the JRC QMRF inventory.

Quantitative Mechanistic Model (QMM) for Schiff Bases – QSAR which relies on Log Kow and reactivity as modelled by sigma* constants (see qsardb.org/repository of the JRC QMRF Inventory for associated QMRF).

- Non-testing approaches characterising KE2 – Gene expression of antioxidant response element (ARE) in keratinocytes

Profiler for Keratinocyte gene expression – SARs that have been extracted from evaluating substances that have been tested in the KeratinoSens™ assay.

- Testing information

Testing approaches characterising MIE – Covalent modification of proteins by skin sensitisers

These are described in more detail in the associated Annex II.

In chemico binding to synthetic peptides expressed as % peptide depletion (DPRA).

In chemico binding – glutathione depletion assay – expressed as RC50 which is inversely related to the kinetic rate constant (Schultz et al., 2005).

- Testing approaches characterising KE2 – Gene expression of antioxidant response element (ARE) in keratinocytes

KeratinoSens™. This is described in the Annex II in more detail.

- Testing approaches characterising KE3 – dendritic cell activation

These are described in the Annex in more detail.

human Cell Line Activation Test (h-CLAT).

Myeloid U937 Skin Sensitization Test (MUSST) (since renamed as U-SENS).

- Testing approaches characterising KE4 – T cell proliferation

Local Lymph Node Assay (LLNA).

- Testing approaches characterising the AO

Guinea Pig Maximisation Test (GPMT).

Buehler Test.

- Other Relevant testing information

These test information have been included as relevant supporting information in the assessment of skin sensitisation. Substances that are corrosive or highly toxic by the dermal route by virtue of an acute study could be excluded from sensitisation testing. Based on the commonality of the MIE in terms of electrophilicity being a key indicator, in certain instances, information from genotoxicity studies may be helpful to consider. A substance that was positive in an Ames test may have the potential to be a skin sensitiser also. Case examples and exclusions were discussed in more detail in Mekenyan et al. 2010.

Skin irritation/corrosion.

Acute dermal test (skin LD50).

Ames mutagenicity.

In vitro chromosomal aberration test (ivt CA).

7. Data interpretation procedure applied

The conceptual diagram outlining the workflow for how to structure the information in an integrated fashion is shown in Figure III.1 with a modified weight of evidence (WoE) matrix (Table III.1). The integration of the information together to make a decision is qualitative in nature.

In the first instance, a consideration is made regarding physical form. If the substance of interest is a gas, then testing is unnecessary. A substance will be predominantly in the gaseous form at vapour pressures of greater than 1x10-4 mmHg. Alternatively other factors may obviate testing – for example if the substance was classified for corrosivity, or if the substance is a strong base or strong acid, testing may also be unrealistic. Strong bases and strong acids will have pH values greater than 11.5 or less than 2. A substance that was inflammable in air at room temperature or possessed explosive or pyrophoric properties would also not warrant testing. Stopping the process for these types of considerations does not mean a substance will not have the potential to sensitise instead it flags certain technical challenges that mean either testing is not feasible or interpretation of experimental outcomes will be difficult e.g a corrosive substance could still be a skin sensitiser. The next step considers what measured skin sensitisation data might already be available for the substance of interest and whether this is sufficient for the decision context in mind. A good quality guinea pig maximisation or LLNA result may be sufficient to conclude on the skin sensitisation potential and/or potency. Any study should be evaluated in concert with what is understood about the substance in terms of its chemistry. A volatile substance may be underpredicted in the LLNA and for in chemico or in vitro data, consideration should be paid to the physicochemical characteristics and reaction domain. Some of the assays are biased towards soft electrophiles or suffer issues with highly hydrophobic substances or simply lack the metabolic capacity to identify substances that need to be transformed. If no measured data on the substance under consideration is available, a QSAR approach can be attempted if a suitable QSAR is available. If not, a read-across can be attempted which will rely on the identification and evaluation of relevant analogues with associated skin sensitisation information – either measured data from any of the in vitro or in chemico assays characterising the respective key events in the AOP or by inference from other in vivo data. If such a non-testing approach is not feasible or requires additional corroborating evidence, new test data in non-animal methods can be generated. If the results are still not sufficiently robust for the decision context then new animal studies may be warranted. Figure III.1 attempts to summarise the steps involved and which of them (in blue) have been encoded and implemented in the OASIS Pipeline software tool and therefore the subject of the associated manuscript.

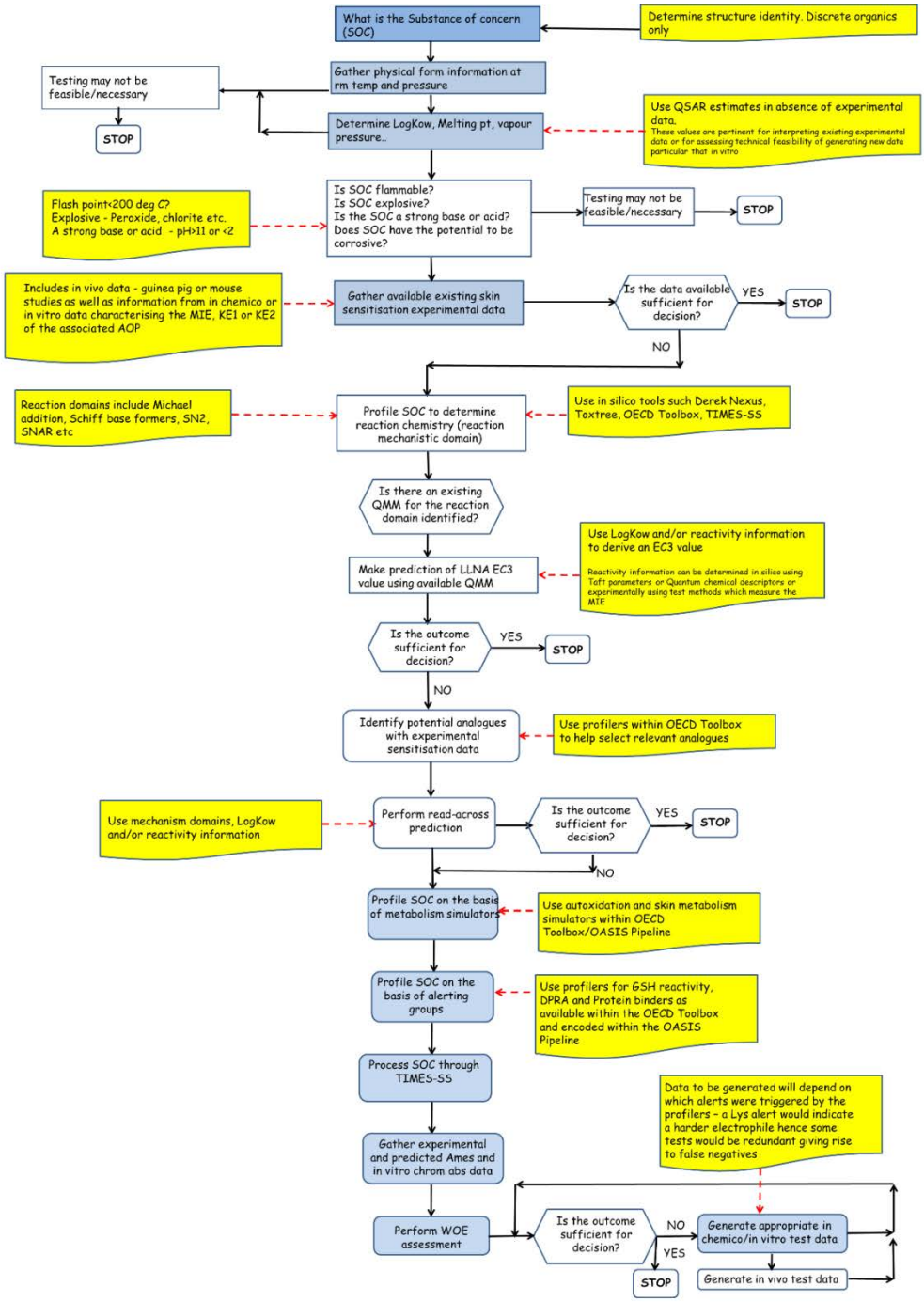


Figure III. 1: A conceptual diagram outlining the workflow for how to structure the information in an integrated fashion. Boxes in yellow provide explanatory notes of what sort of tools, information could be considered to help address each of the steps in the workflow. Boxes in blue represent the components that have been implemented into the OASIS Pipeline software tool.

Notes: ivt CA = in vitro chromosomal aberration data; GSH = Glutathione depletion assay; DPRA = Direct Peptide Reactivity Assay, KE = Key Event; QMM = Quantitative Mechanistic Model

Table III.1: Adapted Matrix for Weight of Evidence Analysis based on Patlewicz et al (2014) IATA-SS.

Component	**Comments**	**Reference (scientific literature, Test Guidelines, Methods etc.)**	**Study result and/or positive (+ve)/negative (-ve) evidence obtained**	**Data reliability e.g. Klimisch**	**Data relevance including coverage/prediction of relevant parameters**	**Consistency with other information**	**Conclusive remarks (adequacy of information for given component)**
Exposure information							
Dermal penetration	Physicochemical properties e.g. pKa, LogKow, vapour pressure (VP) Volatility could be an issue for substances tested in the LLNA, hence VP would be a worthwhile consideration (VP> 1x10-4 mmHg), Log Kow could inform on likely water solubility issues (e.g. Log Kow >4). Low water solubility could be described as <10-2 g/l). Parameters could be estimated through QSAR.						
Degradation/Metabolism information e.g. degradation (including hydrolysis), metabolism, autoxidation	This could be simulated using tools such as the OECD QSAR Toolbox to determine whether chemical under consideration needs to be activated in some manner to exert its effect or is deactivated						

Non-testing approaches							
Protein Binding alerts such as those in the OECD Toolbox, Derek Nexus, Toxtree	Profile the chemical to identify relevant alerts – use these to help interpret existing available in vivo or in vitro information						
Application of available **Quantitative Mechanistic Models** (QMMs)	Based on the outcome of the protein binding alerts and the associated reaction chemistry domain, an available QMMs for Schiff Base, Michael addition, SNAr could be used to make a determination of skin sensitisation potential and potency. Note reactivity information is needed which could be generated experimentally (i.e. MIE test methods) or by computational approaches						
Mechanistic read-across using modified DPRA, GSH or other rate constant information in concert with Log Kow where appropriate for specific reaction domains							
Read-across as performed within the **OECD QSAR Toolbox**							

Components characterising other KE information such as a profiler for DPRA, GSH etc as implemented in the IATA-SS OASIS Pipeline or the OECD Toolbox AOP for SS	SARs to characterise specific key events (KEs)						
Expert systems e.g. TIMES-SS model	Prediction of AO						
Protein binding/reactivity	Assays, approaches to generated data for MIE e.g. DPRA, GSH NB: Considerations might include scope of assay for reaction chemistry domain, water solubility or other technical limitations						
Events in keratinocytes	Assay to generate data for KE2 in vitro data KeratinoSens™ NB: Considerations might include scope of assay for reaction chemistry domain, water solubility or other technical limitations						
Events in dendritic cells	Assays to generate data for KE3 - h-CLAT, MUSST NB: Considerations might include scope of assay for reaction chemistry domain, water solubility or other technical limitations						

Events in lymphocytes	In vivo study (LLNA) Consider outliers to the LLNA i.e. if substance is similar to other substances that have been reported as false positives or false negatives e.g. SLS						
Adverse outcome							
Existing human data e.g. HRIPT							
In vivo (Guinea pig) GPMT, Buehler							
Other relevant information e.g. skin irritation/corrosion data, dermal acute study information (LD50), Ames, in vitro chrom abs	Depending on use considerations, corrositivity could be a consideration when evaluating existing data or deciding on new testing. Correlation between sensitisation and mutagenicity for common reaction chemistry domains (Mekenyan et al., 2010) and SARs exist that can be exploited to substantiate/support a read-across						
Overall conclusions	1. WoE allows a decision on the skin sensitisation potential (and possibly potency) of a substance to be made 2. WoE does not allow a decision of skin sensitisation potential (potency) of a substance to be made. Recommendation of most appropriate additional testing (could be based on other structured ITS) NB: This will also depend on the decision e.g. prioritisation, hazard identification, risk assessment						

8. Chemicals used to develop and test the DIP

8.1 Availability of training and test sets:

No training set is available for the workflow approach itself since it was derived manually by expert judgement. Training sets for the individual elements have been reported in the literature in the primary sources that describe the individual in chemico or in vitro tests. The components referenced in section 6 as non-testing are, with exception to TIMES-SS, publically available in the OECD QSAR Toolbox v3.3. The physicochemical properties such as vapour pressure etc are either measured outcomes or estimated using QSARs such as those within EPI Suite (US EPA, 2012) which is also implemented within the OECD QSAR Toolbox v3.3. The test set of chemicals used to evaluate the feasibility and performance characteristics of the OASIS pipeline that was implemented in a software tool was taken from Teubner et al. 2013.

8.2 Selection of the training set and test set used to assess the DIP:

The test set from Teubner et al. 2013 was chosen because it covered a reasonable diverse set of substances with available in vivo data that were not necessarily the same substances that would have typically been used in the development of many of the in vitro and in chemico assays.

8.3 Supporting information on the training and test sets:

Not provided. The test set is available as Table 1 in Teubner et al. 2013.

8.4 Other information on the training and test sets:

The test set comprised 100 substances which had been extracted from the literature, BASF internal records and from the ECHA REACH dissemination website. Substances were classified in accordance with the Global Harmonised Scheme (GHS) as either Cat 1, 1a or 1b sensitisers or as non-sensitisers using data principally from the GPMT, the Buehler test and the LLNA. Within the set of 100 chemicals, 3 were inorganics. In total 55 of the substances were categorised as sensitisers, the remaining 45 as non-sensitisers.

9. Limitations in the application of the defined approach

The workflow is appropriate for discrete organic substances, preferably ones that are monofunctional in nature that lend themselves to be readily categorised by the available reaction mechanistic domains as proposed by Aptula and Roberts, 2006.

Substances that are pharmaceutical or agrochemical in nature are envisaged to be poorly predicted by this workflow as will polymers, organometallics, nanomaterials or mixtures.

Substances that are likely to be metabolised or transformed abiotically may be incorrectly characterised by the current existing in chemico and in vitro assays. Consideration should be made to the simulated metabolites and whether the metabolite would be the more appropriate target substance to evaluate.

Substances that are highly volatile may be underpredicted in terms of their skin sensitisation potency by the LLNA. Substances that are gases at room temperature are excluded from evaluation due to exposure considerations and technical difficulties in testing. Substances that are highly hydrophobic are likely to be mischaracterised by the current in chemico or in vitro assays to their technical limitations in terms of solubility. Certain assays such as the glutathione depletion assay (Schultz et al., 2005) are only appropriate for soft electrophiles such as direct acting Michael acceptors or SN2 or SNAr activations. Hard electrophiles such as Schiff base formers or Acylating agents can not be tested

in this assay, results from the DPRA will need to be carefully evaluated with an expectation of Lysine depletion. The KeratinoSens™ has a propensity to give rise to false negatives for certain acylating agents since it favours soft electrophiles.

10. Predictive capacity of the defined approach

Whilst a predictive capacity can be provided based on the test set evaluated in the original publication, the approach represents a structured WoE that is to an extent context dependent on the substance being evaluated and the availability of experimental data for that substance. In Patlewicz et al. 2014, performance characteristics were evaluated for the Teubner et al. 2013 dataset using the OASIS Pipeline software implementation which mirrors many of the components represented in Figure III.1 (shown in blue) but does not capture the entire workflow. Processing the original 100 substances through the first two components of the pipeline filtered out 35 substances either on the basis that those substances had experimental in vivo data in the OECD Toolbox or triggered the extreme values in the physchem filter. For the remaining 35 substances, an assessment of skin sensitisation hazard could be made using the remaining components. The performance characteristics were as follows: sensitivity was 74%, specificity was 73.7% and the overall accuracy was 73.85%. After evaluation of the incorrect predictions, refinements were made to the alert descriptions in TIMES which reduced the number of false negatives and several of the false positives were re-evaluated with respect to the in vivo data available. After the refinements/corrections proposed, the overall accuracy increased to 87.6%. The sensitivity and specificity values were 89% and 86.4% respectively.

11. Consideration of uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

The approach has aimed to address as many of the KEs as possible though the confidence will be driven to an extent on the applicability domains of the assays themselves. The approach is biased towards information sources that inform KE1 – this is explicit by the non-testing (in silico components) that are relied upon. KEs 2 and 3 are included particularly if new information needs to be generated.

The approach is also biased to chemicals that act as direct acting soft electrophiles based on the capacity of the available in chemico and in vitro assays. There is residual uncertainty in estimating the skin sensitisation potential of hard electrophiles.

1. The information sources used within the defined approach,

This has not been evaluated beyond the original test set of 100 chemicals. The workflow does use some in silico components that will be reproducible provided the same software version of TIMES-SS and the OECD Toolbox are used.

2. Benchmark data used,

Not evaluated specifically. The workflow aims to infer the sensitising outcome of the LLNA. The variability of the LLNA will of course affect the accuracy and confidence in a sensitisation prediction for humans. The correlation of the LLNA with human NOELs and LOELs is reasonable but far from perfect. The WoE table lists human data as an input that can be considered but the intention was more to present these as a source of comparison.

3. Others sources

Phase I metabolic pathways are not fully represented - Uncertainty likely to cause under-estimation for some phenolic pro-haptens

11.2 Impact of uncertainty on the DIP's prediction

The approach presented does not provide an explicit assessment of prediction uncertainty. The modified WoE table aims to make explicit all the assumptions associated with the different considerations to facilitate an assessment of uncertainty by the expert undertaking the analysis. The exact makeup of the inputs is context dependent on the chemical under evaluation which will dictate the nature of the information that may be generated by in chemico or in vitro assays or is already available.

The user will have higher confidence in predictions by those chemicals:

- that lie within the applicability domain of TIMES-SS
- are Schiff base formers or SNArs for which QMMs are available
- that lie within the optimal ranges of LogKow that minimise any technical limitations in the running of the available in chemico or in vitro tests
- that are direct acting in nature and do not require metabolic or chemical activation
- for which corresponding alerts for genotoxicity are valid and where the supporting experimental data helps to build the weight of evidence.

On the other hand, predictions with lower confidence are those where the substances fall outside of the TIMES applicability domain, multifunctional chemicals for which assignment of a reaction domain is challenging and assigned as a special case, acylating agents for which assays such as the KeratinoSens™ tend to give rise to false negative results and substances which are associated with data showing them to be corrosive or highly irritating.

12. References

Aptula A.O. and Roberts D.W. (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.

Mekenyan O, Patlewicz G, Dimitrova G, Kuseva C, Todorov M, Stoeva S, Kotov S, Donner EM. (2010). Use of genotoxicity information in the development of integrated testing strategies (ITS) for skin sensitization, Chem. Res. Toxicol, 23(10):1519-1540.

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

Patlewicz G, Kuseva C, Mehmed A, Popova Y, Dimitrova G, Ellis G, Hunziker R, Kern P, Low L, Ringeissen S, Roberts DW, Mekenyan O. (2014). TIMES-SS--recent refinements resulting from an industrial skin sensitisation consortium, SAR QSAR Environ Res, 25(5):367-391.

Roberts DW, Aptula AO. (2008). Determinants of skin sensitisation potential, J. Appl. Toxicol, 28(3): 377-387.

Schultz TW, Yarbrough JW, Johnson EL. (2005). Structure-activity relationships for reactivity of carbonyl compounds with glutathione, SAR QSAR Environ. Res, 16: 313–322

Teubner W, Mehling A, Schuster PX, Guth K, Worth A, Burton J, van Ravenzwaay B, Landsiedel R. (2013) Computer models versus reality: how well do in silico models currently predict the skin sensitisation potential of a substance, Regul. Toxicol. Pharmacol, 67: 468-485.

US EPA (2012). EPI Suite. http://www.epa.gov/oppt/exposure/pubs/episuite.htm

13. Supporting information

None.

14. Abbreviations and definitions

QMM – Quantitative Mechanistic Modelling

TIMES-SS Tissue Metabolism Simulator for Skin Sensitisation

CASE STUDY IV

"Stacking" meta-model for skin sensitisation hazard identification

1. Summary

In this case study, hazard identification is based on the combination of multiple in vitro and in silico parameters covering the Adverse Outcome Pathway's (AOP's) key events 1 to 3 leading to skin sensitisation.

This defined approach comprises in silico predictions (TIMES-SS, Toxtree), Direct Peptide Reactivity Assay (DPRA), U-SENS™ and KeratinoSens™ as well as physico-chemical parameters (pH, volatility) built on 165 chemicals having a LLNA-based Sensitisers/Non Sensitisers (S/NS) classification.

A meta-model stacking five different statistical methods (Boosting, Naïve Bayes, Support Vector Machine (SVM), Sparse PLS-DA and Expert Scoring) was established to determine a probability of belonging to the group of interest ("sensitiser"). Based on defined thresholds of this probability, a prediction model is proposed to classify a chemical as S or NS. The outcome can be used to predict the skin sensitising potential of a chemical, but does not allow potency estimation.

2. General information

2.1 Identifier:

Stacking meta-model for Skin Sensitiser hazard identification.

2.2 Date:

29 April2016

2.3 Author(s) and contact details:

Alépée N, Del Bufalo A.; Detroyer A.; Eilstein; Gomes C.; J.;Nocairi H.; Pauloin T.; Piroird C.; Rousset F.; Teissier S.; Tourneix F.

L'Oreal Research & Innovation, 1 Eugène Schueller, 93600 Aulnay-sous-bois, France.

Corresponding author: Alépée Nathalie; nalepee@rd.loreal.com; Tel +33 1 48 68 91 02

2.4 RF update(s): Version 1

2.5 Reference to main scientific papers:

Piroird, C., Ovigne, JM., Rousset, F., Martinozzi Teissier, S., Gomes, C., Cotovio, J., Alépée, N. (2015). The myeloid U937 skin sensitization test (U-SENS) addresses the Activation of Dendritic Cell Event in the Adverse Outcome Pathway for Skin Sensitization, Toxicol In Vitro, 29: 901-916.

Alépée, N., Piroird, C., Aujoulat, M., Dreyfuss, S., Hoffmann, S., Hohenstein, A., Meloni, M., Nardelli, L., Gerbeix, C., Cotovio, J. (2015). Prospective multicentre study of the U-SENS test method for skin sensitization testing, Toxicol in Vitro, 30: 373-82.

Gomes, C., Noçairi, H., Thomas, M., Collin, J.F., Ibanez, F., Saporta, G. (2012). Stacking prediction for a binary outcome. COMPSTAT, 20th International Conference on Computational Statistics, Limassol, 271-282.

Gomes, C., H. Nocairi, H., Thomas, M., Collin J.F., Saporta, G. (2014). A simple and robust scoring technique for binary classification, Artificial Intelligence Research, vol. 3(1): 52-58.

2.6 Proprietary aspects:

A license agreement is needed for TIMES-SS, software commercially available from OASIS LMC. The Myeloid U937 Skin Sensitization Test (U-SENS™) is using the U937 cell line that can be freely used for research purposes only. Otherwise, Dr. K. Nilsson (kenneth.nilsson@genpat.uu.se) is restricting the commercial use of this cell line to the payment of a license fee. The KeratinoSens™ including the Luciferase reporter gene is a proprietary method for which a license agreement with Givaudan is needed. The two test methods are now widely offered by CRO's.

3. Endpoint addressed

3.1 Endpoint:

The present Data Interpretation Procedure (DIP) was developed to predict skin sensitisation hazard identification (sensitisers *vs* non sensitisers), addressing the endpoint of existing test guidelines e.g. Local Lymph Node Assay (LLNA) (OECD TG429), Guinea Pig Maximisation Test (GPMT) and Buehler Test (OECD TG406).

3.2 Species:

Mouse (primary target), Human (species of interest).

3.3. Additional information about the endpoint:

The skin sensitisation endpoint is described by the OECD AOP, which captures the impact of skin exposure and describes key events starting from the molecular initiating event i.e. covalent binding of a chemical to skin protein (protein haptenation, Key Event 1), via intermediate key events like keratinocytes and dendritic cells activation (Key Event 2 and 3 respectively), to the final determining event which is the induction of hapten-specific T cells (Key Event 4) that after being challenged by the substance, will be the effector cells in the clinical manifestation of skin sensitisation: Allergic Contact Dermatitis (ACD). The present DIP is based on the physicochemical properties, intrinsic chemical reactivity and capacity to induce innate immune events. In its current version, it does not integrate the specific T cells induction key event.

4. Definition of the purpose and regulatory relevance

The defined approach aim at the identification of skin sensitisation hazard with the classification of chemicals for regulatory purposes, e.g. for classification and labelling under the Globally Harmonized System of Classification, Labelling and Packaging of Chemicals (UN GHS) and for Cosmetics Regulation / REACH legislation. The defined approach does not provide skin sensitisation potency information and therefore cannot be used to sub-categorise skin sensitisers into UN GHS subcategories 1A and 1B.

5. Rationale underlying the construction of the defined approach

The Stacking meta-model is a probabilistic hazard identification approach which encompasses the most relevant events of the skin sensitisation AOP key events. As such, the combined approach is based on intrinsic physico-chemical properties of the chemical and addressed the early innate immune cell responses key events, as described below.

1) Based on intrinsic physico-chemical properties of the chemical

- Gives some descriptors allowing to integrate stability and/or bioavailability characteristics that could potentially impact *in vitro/in vivo* correlation due to different testing conditions (aqueous culture media versus topical application in AOO, DMSO or other rather lipophilic vehicles used in the LLNA). As such, the measured pH and the calculated volatility were considered as relevant variables to combine with *in silico*, *in chemico* and *in vitro* methods, as defined in a splitting statistical analysis (Gomes et al., 2012). See individual information sources for rationale description ("Mechanistic basis including AOP coverage" section).

- Informs on chemical (protein) reactivity (which is directly linked to the initial key event: haptenation of skin proteins): this is the case for the Toxtree skin sensitisation alerts (Aptula and Roberts, 2006). The Times-SS predictions also mainly take into account electrophilic binding to skin proteins either directly or following metabolism but it is not the only mechanism that is integrated (Patlewicz et al., 2007). Finally, the *in chemico* DPRA (Gerberick et al., 2004; OECD Test Guidelines 442C), related to AOP key event 1, is a method giving a measurement of molecular initiating events (MIEs) as cysteine and lysine peptides modifications by the chemical.

2) Addressing early innate immune cell responses, the defined approach relies on:

- Key event 2: i.e. keratinocytes activation, with the KeratinoSens™ assay assessing the induction of the Nrf-2 pathway (Emter et al., 2010; OECD Test Guidelines 442D).

- Key event 3: i.e. dendritic cells activation, with the existing DC-surrogates based CD86 activation U-SENS™ assay (Piroird et al., 2015; Alépée et al., 2015).

From the large number of supervised classification models proposed in the literature, five different methods: Boosting, Naïve Bayes, Support Vector Machine (SVM), Sparse PLS-DA and Expert Scoring were selected (Gomes et al., 2014). These methods have strong differences, but they all produce posterior probability of belonging to the group of interest ("sensitiser"). Instead of trying to choose a specific method, we combine them by the stacking methodology of Wolpert (1992) and Breiman (1996), in order to obtain a specific "stacking" meta-model. This stacking meta-model has shown better performances than each of the five models taken separately (Gomes et al., 2012).

6. Description of the individual information sources used (see Annex II)

The 7 input data (described below) were combined for the construction of the prediction defined approach.

- Qualitative volatility class

- Quantitative measured pH

- Qualitative class prediction from TIMES-SS

- Qualitative reactivity prediction from ToxTree

- Qualitative reactivity class prediction for DPRA

- Qualitative S/NS class prediction for KeratinoSens™

- Qualitative S/NS class prediction for U-SENS™

1. Volatily: Expressed through the vapor pressure calculated by the MPBPVP model in Episuite software (from US EPA / Opensource). Based on the structure of a given chemical, the MPBPVP model estimates vapor pressure from various physico-chemical equations. In turn these equations all use as input data, measured or calculated boiling points derived from group contribution QSAR methods. A final "suggested" vapor pressure estimation is chosen depending on the fact whether the chemical is a solid, liquid or gas (On-line MPBPVPWin™ user's guide). These values were converted into volatility classes according to Spicer (Spicer et al., 2002): VP<10-7 mmHg = non-volatile; VP between 10-7 and 10-1 mmHg = semi volatile; VP between 10-1 and 380 mmHg = volatile; VP>380 mmHg = very volatile (These last two groups are for the stacking meta-model purpose grouped together into a "very volatile" class).

2. pH: Measured quantitative value (between 1 and 14) obtained with a method adapted from OECD Guideline for the Testing of Chemicals No. 122. (OECD TG 122).

3. TIMES-SS: In silico TIssue MEtabolism Simulator integrating Skin Sensitisation prediction model (from OASIS-LMC).

The TIMES-SS model analyses the parent molecule structure as well as its simulated metabolites to predict skin sensitisation potency based on structural alerts and 3D-QSARs. The "in domain" prediction that is retained is that of the most potent structure among parent and metabolites, and is expressed as one of 3 classes: non sensitiser, weak or strong sensitiser (Patlewicz et al., 2007). Weak and strong sensitisers are grouped into "sensitisers" in the stacking meta-model.

4. Toxtree: In silico prediction software containing Skin Sensitisation Alerts based on the Reaction Mechanistic Domains classification (from Ideaconsult Ltd / Opensource). Its "Skin Sensitisation Alerts" decision tree which relies on a Reaction Mechanistic Domains classification, will output alerts for a parent chemical structure. With SNAr, SN2, Acyl transfer agent, Michel acceptor and Shiff base formation alerts the chemical is classified as reactive (sensitiser), with no skin sensitisation alert as non-reactive (non-sensitiser) (Aptula and Roberts, 2006).

5. **Key event 1:** DPRA: Direct peptide reactivity assay (OECD TG 442C)

Sensitisers are electrophilic chemicals which can react with the side chain of nucleophilic amino acids of proteins. The reactivity assay DPRA is based on the depletion measurement of two synthetic peptides (Cysteine and Lysine peptides) after reaction with the test chemical. Chemicals that induce mean peptide depletion of cysteine- and lysine-containing peptide above 6.38% are considered to be reactive (Gerberick et al., 2004).

6. **Key event 2:** KeratinoSens™ (OECD TG 442D)

KeratinoSens™ assay measures activation of the Nrf2-Keap-1 pathway, the well described cellular sensor of electrophilic/redox stress which is induced in skin cells in response to sensitisers (Ade et al., 2009; Natsch et al., 2013) and in particular in keratinocytes (Key event 2: keratinocytes activation). The KeratinoSens™ is performed using HaCaT Nrf2-luciferase reporter cell line. Cells are exposed to a concentration range of the test chemical for 48 hours. 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).

7. **Key event 3:** U-SENS™: Myeloid U937 Skin Sensitisation Test (OECD TG draft)

DC activation tests are representative for the maturation of DCs as only mature DC can activate naïve T cells. The cell surface marker CD86 is one of the classical markers indicating DC maturation. CD86 binds to molecules on T-cell surface and is a key molecule in the T-cell priming process. U-SENS™ is considered in this approach: activation of cell surface marker CD86 measurement in U937 cells, which serve as surrogates for dendritic cells. Co-stimulatory molecule CD86 as a marker of cell activation as well as cell viability assessed using propidium iodide exclusion are measured by flow cytometry. Chemicals that induce the expression of CD86 higher than 1.5 fold, at cell viabilities above 70%, compared to the controls are predicted to have a DC activating potential and therefore a sensitisation potential (Ade et al., 2006; Piroird et al., 2015; Alépée et al., 2015). The test method is under peer-review at EURL-ECVAM and is integrated in OECD TG programme. The S/NS prediction of U-SENS™ was used in this Stacking meta-model for Skin Sensitiser hazard identification

7. Data interpretation procedure applied

The global stacking model (built with the complete learning set) was done using the variables that have been selected across all the previous meta-models.

The 7 input variables (see section 6) are entered into the model where they are run in 5 different supervised classification models (Boosting, Naïve Bayes, SVM, Sparse PLS-DA and Expert Scoring) (Figure IV.1, Appendix IV.1).

• Chemicals with probability to be sensitiser ≥ 70% are predicted “Sensitiser”;

• Chemicals with probability to be sensitiser ≤ 30% are predicted “Non Sensitiser”;

• Chemicals with probability between those two thresholds are predicted “Equivocal”.

The thresholds of 70% and 30% were set to optimize the balance between predictivity and rate of conclusive calls using the current 7 input variables.

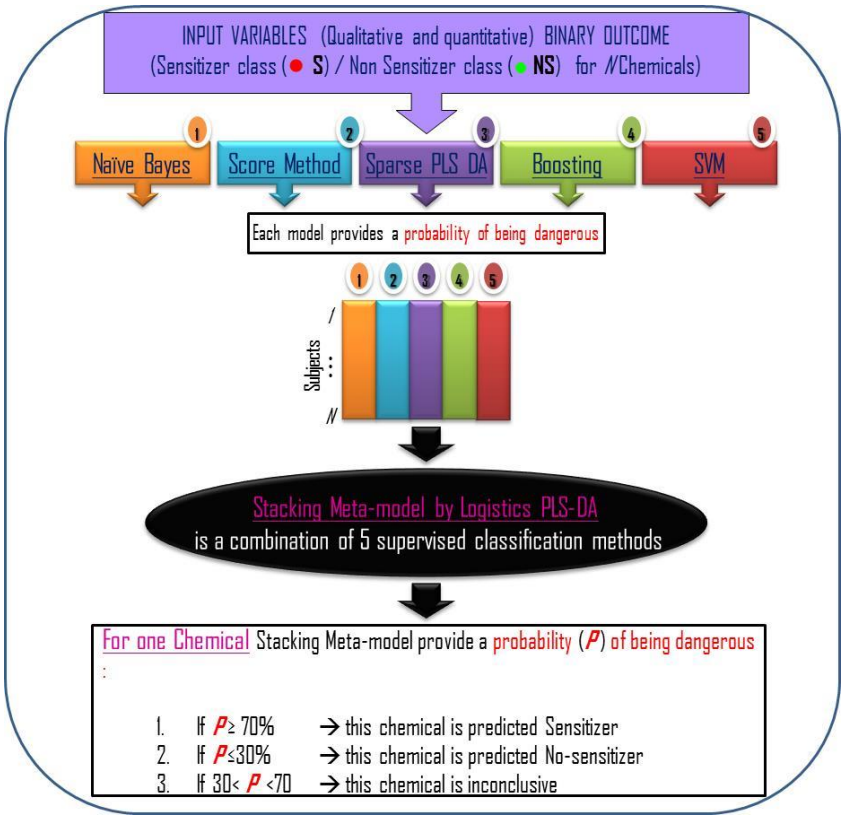


Figure IV.1: Visualization of the methodology of Stacking

8. Chemicals used to develop and test the DIP

8.1 Availability of training and test sets:

The global set comprises 165 chemicals (Appendix IV.2).

The choice of chemicals selected in the global set is summarized in Figure IV.2 and was guided by:

- Presence of good quality animal data (and/or human data)
- Balanced set for S/NS in vivo classes
- To cover a wide diversity of physico-chemical properties: MW, pH
- Good representation of the diversity of cosmetic classes

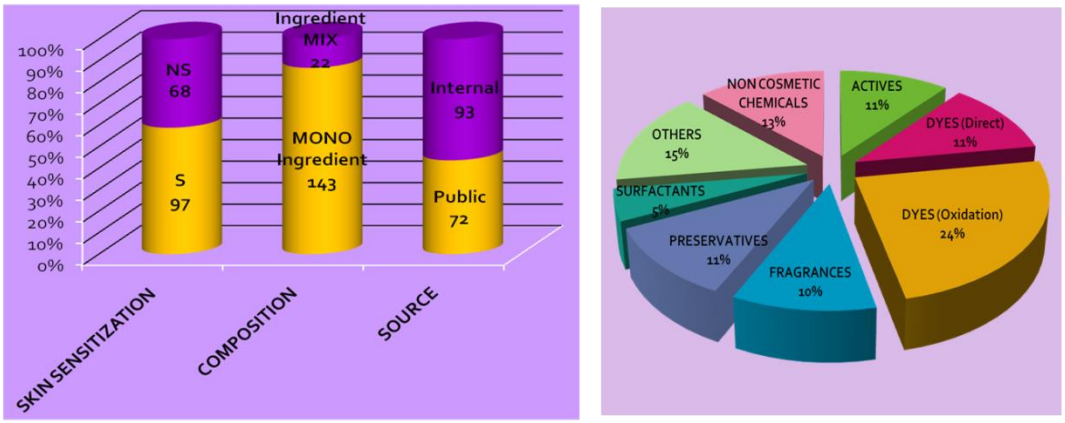


Figure IV.2: Information related to the training and test sets of chemiclas used for DIP development and testing.

For the 165 chemicals, the 3 in chemico / in vitro assays (DPRA, KeratinoSens™ and U-SENS™) qualitative parameters (R/NR or S/NS classes) were used in the DIP. The ranges of quantitative parameters associated to these assays for the global set are given below:

- DPRA: the mean values of cysteine and lysine percent depletion (or cysteine only when lysine depletion was not measurable) were included between 0 to 97% %;

- KeratinoSens™: the EC1.5 values (i.e. the concentration needed for a significant induction of luciferase activity above the positive threshold) were comprised between 1 to > 2000 µM (for chemicals with defined MW) and between 3 to > 400 µg (for chemicals without defined MW);

- U-SENS™: the EC150 values (i.e. the concentration needed for a significant induction of CD86 above the positive threshold) were comprised between 0.1 to > 200 µg/mL.

For the measured pH, the large range of values (from 1.5 to 11.5) indicates that the set contains a wide variety of acids, bases and neutral chemicals.

For the in silico determined volatility, in order to obtain an equilibrated set for the stacking meta-model purpose the 2 Spicer classes “volatile” and “very volatile” were grouped together into one “very volatile” class.

In the set, all reactive classes determined in silico with Toxtree, (SNAr, SN2, Acyl transfer agent, Michel acceptor and Shiff base formation) are represented, be not in the same number.

In the set all 3 classes predicted in silico with Times are represented, with the weak and strong sensitisers grouped into “sensitisers” in the stacking meta-model.

8.2 Selection of the training and test set used to assess the DIP:

The chemicals used in the training set, representative of the global set (165), were selected among all chemicals in order to avoid any bias (not pertinent decision rules) that may be induced by a particular choice of subsets.

In the case of a small data set, sampling into a learning and test subsets may lead to some empty categories of both the response (rarely) and (frequently) of the predictors. In this case it is impossible to estimate some parameters. In any case, it is suitable to have minimum number of observations in each category of each variable.

The solution relies on a specific stratification technique:

A balanced stratification according to the categories of the outcome *y* is necessary to keep constant their proportions. As such, the following heuristics are then used:

- Perform a random split into training and test data sets with a stratified sampling upon both categories of the outcome (NS /S);
- Reject a sample if a category of a predictor has not enough representatives in both the training and test samples;
- Repeat until acceptance;
- Repeat until getting 6 balanced samples.

Applying this methodology:

- 113 chemicals constituted the training set (66 S / 47 NS);
- 52 chemicals constituted the test set (31 S / 21 NS).

8.3 Supporting information on the training and test sets:

Supporting information on the training and test sets is provided in Appendix IV.2. For each chemical, this includes CAS registry numbers, SMILES, ingredient type, chemical class, training or test set designation, MW, Log P, volatility, pH, qualitative class prediction from each in silico/in chemico/in vitro individual data sources (TIMES-SS, ToxTree, DPRA, KeratinoSens™, U-SENS™), LLNA classification, and the defined approach hazard classification with its probability of being a sensitiser determined.

In Appendix IV.1, the description of the individual statistical data sources and the stacking meta-model are summarised.

8.4 Other information on the training and test sets:

Not applicable

9. Limitations in the application of the defined approach

The strengths and limitations on the seven individual test inputs are detailed in the respective individual data sources. Potentially interferences for volatiles, color, highly cytotoxicity, low solubility, pre- or pro-haptens, membrane disrupting chemicals might occur depending of the individual sources (see Annex II). By integrating the different individual data sources, the satcking meta model minimises individual limitations and allows a correct classification of pre- pro-haptens, dyes and low soluble chemicals, as examples. It is also important to notice that each individual source is “disengageable”. As such, the DIP could be applied even when the outcome of in silico tools could be not applicable (for chemical entities with no defined structure) or an inconclusive for in chemico / in vitro (e.g. depletion of peptides not due to adduct formation, pigments could interfere with viability readouts) is obtained (Appendix IV.2).

The limitation of the defined approach is for chemicals with a probability given between the two thresholds of 70% and 30% (uncertainty probability to be a sensitiser). Then, the chemical are predicted "Equivocal".

Overall applicability domain of the defined approach comprises several classes of cosmetic chemicals (fragrances, dyes, preservatives, actives, surfactants and UV filters) and non-cosmetic organic chemicals. The classification model is applicable to chemical entities with a defined molecular structure. It can also be applied to polymers, multi-constituent substances, substances of unknown or variable composition, complex reaction products or biological materials (UVCB) and mixtures.

Results should be interpreted with care for agrochemicals, metals, nanomaterials, or mixtures since the representation of these categories in the learning set is low or absent. The DIP is not applicable for gases.

10. Predictive capacity of the defined approach

The predictive capacity of the DIP against LLNA classifications is shown in the following figure. The probability to belong to the class of "Sensitisers" (Y-axis) is represented for each chemical (Red dots = chemicals classified as sensitisers by the LLNA and Green dots = chemicals classified as non-sensitisers by the LLNA). The performances of the DIP were calculated based on the prediction approach with optimized thresholds of classification at 70% and 30% (i.e.: if p>70%: Sensitiser / if p<30%: Non Sensitiser). A category of 30 % < p < 70 % (equivocal) was established on the training set to optimize the kappa value leading to a high confidence and a good prediction on the 2 other S/NS categories (and validated on the test set).

The chemicals falling in the category 30 % < p < 70 % (equivocal) were not included in the calculations of predictive capacity. Thirteen mono ingredients et 2 complex fell in the category "equivocal" (subsets:11 for training and 4 for test). There is not a dominant functional group, nor physico chemical properties (MW, cLogP, volatile class) or reactivity binding class (SN2, Schiff base formation, Acyl Transfer agent, Michael Acceptor) for equivocals as a subset group with higher frequencies also corresponds to the most populated global set. There are no dominant in vivo drivers of classification for equivocals (7 S and 8 NS in LLNA).

Please note that the predictive capacity is given for the training set and for the test set (Figure IV.3).

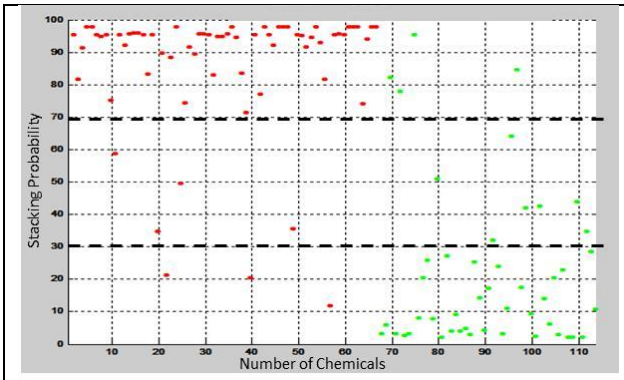
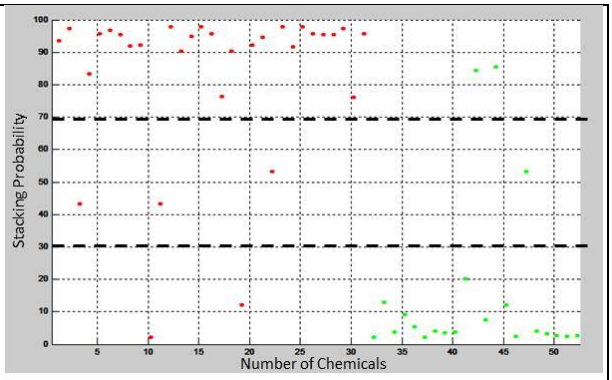
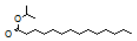
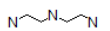
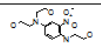
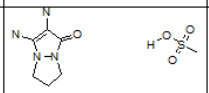
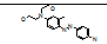
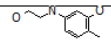
	
Training Set (n = 113)	**Test Set (n = 52)**
Accuracy = 93.14%	Accuracy =91.67%
Sensitivity = 95.16%	Sensitivity=92.86%
Specificity = 90%	Specificity=90%
Kappa = 0.85	Kappa = 0.82
Equivocal = 10%	Equivocal =8 %

Figure IV.3: Predictive capacity of the DIP for the training and the test set.

A detailed analysis of the 11 false positives (FP) and false negative (FN) with regard to LLNA is given below (Table IV.1):

Ingredient n°	INCI name or chemical name	CAS	Structure	Chemical Class	MW	cLog P v5.2	Water Sol (mol/L) Epiwin v4.0	pH	Volatility	Toxtree v2.5.0	TIMES	DPRA	U-SENS	Keratino Sens	ITS (p)	ITS result	LLNA classification	LLNA EC3 (%)	Set
39	ISOPROPYL MYRISTATE	110-27-0		FATTY COMPOUND	270.46	7.37	5.01E-08	8.4	Semi volatil	No binding	NS	NR	NS	NS	2.2	NS	Weak sensitizer	44	Test
42	diethylenetriamine	111-40-0		NON COSMETIC	103.17	-2.28	9.69E+00	11.2	Very volatil	Schiff base formation	INC	NR	NS	NS	12.0	NS	Moderate sensitizer	5.8	Training
85	COUMARIN	91-64-5		FRAGRANCE INGREDIENT	146.14	1.41	3.51E-02	8.9	Semi volatil	Michael Acceptor	NS	NR	S	S	20.7	NS	Weak Sensitizer	30	Training
103	HC BLUE No.2	33229-34-4		COLOR ADDITIVE	285.30	0.21	9.79E-02	6.4	Non volatil	Michael Acceptor	S	R	NS	NS	21.4	NS	Moderate sensitizer	4.1	Training
151	OA57	OA57	OA57	ACTIVE COMPOUND	233.36	2.66	2.04E-03	7.2	Semi volatil	Acyl Transfer agent	INC	INC	NS	NS	12.1	NS	Weak sensitizer	17	Test
137	OA44	OA44	OA44	DYE/PIGMENT	378.49	-0.81	(approx.) 1,62E-03	8.1	Semi volatil	Michael Acceptor	INC	INC	S	S	85.5	S	Non sensitizer		Test
152	2,3-DIAMINODIHYDROPYRAZOLO PYRAZOLONE DIMETHOSULFONATE	857035-95-1		DYE/PIGMENT	346.39	-1.53	6.49E+00	1.8	Semi volatil	SN2	INC	R	NS	S	84.5	S	Non sensitizer		Test
111	HC YELLOW No.7	104226-21-3		DYE/PIGMENT	314.39	2.67	1.11E-04	7.9	Non volatil	Michael Acceptor	S	R	S	NS	82.4	S	Non sensitizer		Training
123	2-METHYL-5-HYDROXYETHYLAMINOPHENOL	55302-96-0		DYE/PIGMENT	167.21	0.71	1.85E-01	7.3	Semi volatil	Michael Acceptor	S	R	S	S	95.6	S	Non sensitizer		Training
131	OA38	OA38	OA38	ACTIVE COMPOUND	181.19	1.47	1.87E-02	7.4	Semi volatil	Michael Acceptor Acyl Transfert agent	INC	NR	S	S	78.0	S	Non sensitizer		Training
156	OA61	OA61	OA61	DYE/PIGMENT	196.25	-0.05	3.19E-01	1.6	Semi volatil	SN2	INC	R	NS	S	84.9	S	Non sensitizer		Training

NS: Non-sensitser; S: Sensitiser; INC: Inconclusive; R: Reactive; NR: Non-reactive

Table IV.1: Detailed analysis of 11 false positives (FP) and false negative (FN) predictions based on DIP with regard to LLNA.

No strong or extreme *in vivo* LLNA chemicals were misclassified as a non-sensitiser. This defined approach misclassified as false negatives few chemicals (i.e. oily, tension active, antimicrobial, polymer chemicals) in the test set. The DIP also misclassified few dyes as false positives. Knowing that for the categories classes above, the majority of the chemicals were well classified, no specific classes' limitation could be defined.

When compared to LLNA, 11/165 (6.7%) of the investigated chemicals were falsely predicted by DIP, with 5 false positives and 6 false negatives. Analysis of LogP, pH or reactivity (ToxTree) did not reveal specificities when compared to the 154 correct predictions. Even if LLNA results constitute the most consistent *in viv*o database, this method presents some variability that has to be kept in mind, and human data, when available, are the gold standard.

False Negative (FN) results of the defined approach with respect to LLNA:

Chemical name (code id.)	Rational
Isopropyl myristate (#39)	Predicted weak sensitiser (EC3=44%) by LLNA. No evidence in human (class 5 Basketter et al., 2014). Reported by Urbisch et al., (2015), as probably LLNA False Positive. ⇨ True Negative in DIP
Diethylenediamine (#42)	Known to be a frequent contact sensitiser in Human, with positive HRIPT (Kligman, 1966). FN also negative in other approach Urbisch et al. (2015). ⇨ FN in DIP
Coumarin (#85)	Weak sensitiser in LLNA (EC3=30%). Conflicting observations reported in Human due to possible presence of impurities (Vocanson et al., 2006). No alert for reactivity (TimesSS, DPRA). Positive in cellular tests (U-SENS™, Keratinosens™). Probable impurities issues. ⇨ FN in DIP
HC Blue No.2 (#103)	Equivocal LLNA data reported. In 2013, SCCS Memorandum on Safety Assessment revised the classification of this chemical as non-sensitiser (COLIPA B037). No human sensitisation case reported. ⇨ True Negative in DIP
OA57 (#151)	No human data available. Classified as weak sensitiser (EC3= 17%) by LLNA. ⇨ FN in DIP

False Positive (FP) results of the defined approach with respect to LLNA:

Chemical name (code id.)	Rational
Yellow No. 7 (#111)	Non-sensitiser by LLNA. According to the SCCS memorandum (2013), this testing was insufficient. For Sanchez-Perez et al. 2004 hydrolysis of the dye leads to PPD release, a feature that may explain allergic reactions. Considered also as a contact allergen (Lepoittevin et al. 2009). ⇨ True Positive in DIP
2-methyl-5-hydroxyethylaminophenol (#123)	Tested up to only 50% in LLNA. No report of sensitisation cases in Human. Positive in a GPMT (SCCNFP 1999). ⇨ True Positive in DIP
OA38 (#131)	Tested up to only 25% in LLNA. Negative also with Buehler test or GPMT test. Positive HRIPT (internal data). Interspecies differences (mouse/guinea pig versus Human) regarding metabolic capacities can explain this misclassified between in vivo and clinical results. ⇨ True Positive in DIP
OA 44 (#137)	Tested up to only 1%. Another form (different counter ion) of this dye was predicted at least moderate sensitiser in LLNA when tested up to 5%. It appears necessary to moderate LLNA conclusion raised in suboptimal conditions. ⇨ No sufficient elements to conclude
2,3-Diaminodihydroxy pyrazolopyrazolone dimethosulfonate (#152)	Tested up to only 18% in LLNA. The SCCS opinion in 2012 was "No firm conclusion regarding the sensitising potential can be drawn". No reported human data ⇨ FP in DIP
OA61 (#156)	Tested up to only 10% in LLNA due to solubility limitation. ⇨ No sufficient elements to conclude

Altogether, from the 11 misclassified chemicals in respect to LLNA, the analysis of LLNA data and consideration of existing human data revealed that 5 out of 11 chemicals were correctly predicted by the DIP. Finally, only 6 out of 165 chemicals were misclassified by the DIP.

11. Consideration of uncertainties associated with the application of the approach

11. 1 Sources of uncertainty

1. The DIP's structure

This defined approach aims the identification of skin sensitisation hazard. The use of this approach for potency assessments is not possible.

The defined approach is mostly based on parameters addressing key events 1, 2 and 3 of the AOP. Key event 4 is not included due to lack of available tests. Therefore confidence in the integrated strategy to predict Skin Sensitisation Hazard is lower for substances that act by a molecular initiating event other than the covalent binding to proteins like for example for metals (e.g. nickel).

The most comprehensive DIP structures are probably those that are built mechanistically on a strict (linear) interpretation of the AOP, according to which one would expect that if the assay of key event 1 is positive (i.e. Reactive), the assays of key events 2 and 3 should be positive as well and therefore lead to the adverse effect. Nevertheless, despite the widely accepted key events described in the AOP, the dynamics linking these key events leading to skin sensitisation remain largely unknown impacting the certainty of such approaches. Faced with these knowledge gaps, we chose to take advantage of the large set of reference data (n=165 chemicals) to rely on a robust statistical approach to develop an integrated prediction model. In the present DIP structure all data (in silico / phys chem. properties / in chemico / in vitro) are integrated at once, at the same level into 5 different supervised classification models (Boosting, Naïve Bayes, SVM, Sparse PLS-DA and Expert Scoring), each of it providing a probability of being a sensitiser. These 5 outputs are then integrated into a stacking meta-model, providing a final probability to be a sensitiser. Confidence (kappa) in the stacking meta-model to integrate the data sources was proved to be very high, and its overall accuracy was better than the ones from the five individual machine learning approaches used, showing that the stacking meta model minimize the potential biais of each statistical model and therefore the uncertainty linked to the structure of the prediction model.

2. The information sources within the defined approach

In the defined approach, all *in chemico/in vitro* methods used have been shown to be reliable (intra- and inter-laboratories) and relevant (S/NS) through multicentre studies evaluations. DPRA and KeratinoSens™ assays have been regulatory accepted by OECD (TGs 442C and 442D, respectively) and U-SENS™ assay has been submitted to EURL-ECVAM for peer review assessment and considered in OECD TG program (Alépée et al., 2016).

The uncertainties for the defined approach that are related to the DIP information sources include the following:

- The variability (reproducibility in the individual data sources was not explicitly taken into account. Inconsistent results in the source data for a given chemical would reduce the confidence in the hazard predictions.

- Volatility was predicted rather than measured.

- The uncertainty linked to in silico parameters (predictions versus experimental data) is even higher when the learning set of the in silico tools is overlapping whith the training set and test set of the DIP.

- Results from Times 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 is unknown).

3. Benchmark data used.

The defined approach is used to predict hazard skin sensitisation.

The benchmark data used to develop the test methods was primarily based on data obtained from the murine LLNA. The variability of the reference in vivo data inevitably affects the accuracy of prediction. This variability originates from the intrinsic variability of the biological model and from the testing variability (between- and within-laboratory variability). The LLNA between-laboratory concordance for sensitiser/non-sensitiser classifications is around 80% (NICEATM-ICCVAM, 1999). In the original validation study, the LLNA (and guinea pig tests) was reported to have an accuracy of 72% when compared to human data (Dean et al., 2001). Variability in the EC3 values of the LLNA has reported depending on vehicle used (Dimitrov et al., 2016).

Around those uncertainties, the defined approach was developed using the most prevalent reference result for LLNA hazard classification. The thresholds used for the stacking meta-model (Boosting, Naïve Bayes, SVM, Sparse PLS-DA and Expert Scoring) was based on the probability that a chemical will be positive in the LLNA test.

11.2 Impact of uncertainty on DIP's prediction

The defined approach does not provide an explicit assessment of prediction uncertainty, although the end-user can give higher confidence to those chemicals that fall within the applicability domain of the defined approach. Uncertainty on DIP's prediction has been limited considering measured quantitative value of pH and in *chemico*/*in vitro* validated test methods for which the quantitative variability of the methods was reduced by defining only a categorisation class (eg. NR, NS, R; S) as the input. When taking the final probabilty and the high degree of accuracy into account, confidence in the prediction is high.

12. 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(4): 189-197.

Ade N, Leon F, Pallardy M, Peiffer JL, Kerdine-Romer S, Tissier MH, Bonnet PA, Fabre I, Ourlin JC. (2009). HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: role of the Keap1/Nrf2 pathway. Toxicol Sci, 107(2): 451-460.

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

Aptula A.O. and Roberts D.W. (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.

Dean JH, Twerdok LE, Tice RR, Sailstad DM, Hattan DG, Stokes WS. (2001). ICCVAM evaluation of the murine local lymph node assay. Conclusions and recommendations of an independent scientific peer review panel. Regul Toxicol Pharmacol, 34:258-73.

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). Dermatitis. 25(1): 11-21.

Breiman, L. (1996). Stacked Regressions. Machine Learning. 24, 49-64.

Dimitrov, S., Detroyer, A., Piroird, C., Gomes, C., Eilstein, J., Pauloin, T., Kuseva, C., Ivanova, H., Popova, I., Karakolev, Y., Ringeissen, S., Mekenyan, O. (2016). Accounting for data variability, a key factor in *in vivo/in vitro* relationships: application to the Skin Sensitization Potency (*in vivo* LLNA versus *in vitro* DPRA) example, Journal of Applied Tox, available online, DOI 10.1002/jat.3318.

Emter, R., Ellis G., and Natsch, A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. Toxicol. Appl. Pharmacol, 245: 281-290.

Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP. (2004). Development of a peptide reactivity assay for screening contact allergens. Toxicol Sci.,81: 332-343.

Gomes, C., Noçairi, H., Thomas, M., Collin, J.F., Ibanez, F., Saporta, G. (2012). Stacking prediction for a binary outcome, 20th International Conference on Computational Statistics, 271-282.

Gomes, C., H. Nocairi, H., Thomas, M., Collin J.F., Saporta, G. (2014). A simple and robust scoring technique for binary classification. Artificial Intelligence Research, vol. 3(1): 52-58.

Kligman A.M. (1966). J. Invest. Dermatol. 47: 375-392.

Lepoittevin J.P. et al, (2009). Dictionary of Contact Allergens, Springer Eds, p 136.

Natsch, A., et al., (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation, Journal of Applied Toxicology 33(11), 1337-1352.

NICEATM-ICCVAM, (1999). The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICEATM), NIH Publication No. 99-4494.

OECD, (1992). Skin Sensitisation, Test Guideline No. 406, Guidelines for Testing of Chemicals, OECD, Paris. http://www.oecd.org/env/testguidelines

OECD, (2010). Skin Sensitization: Local Lymph Node Assay, Test Guideline No. 429, Guidelines for the Testing of Chemicals, OECD, Paris. http://www.oecd.org/env/testguidelines

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

OECD, (2013). OECD Guidelines for the Testing of Chemicals, Section 1 Physical-Chemical properties, Test No. 122: Determination of pH, Acidity and Alkalinity; OECD publishings ISSN: 2074-5753.

OECD, (2015). Skin Sensitization: In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA), Test Guideline No. 442C, Guidelines for the Testing of Chemicals, OECD, Paris. http://www.oecd.org/env/testguidelines

OECD, (2015). Skin Sensitization: In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method, Test Guideline No. 442D, Guidelines for the Testing of Chemicals, OECD, Paris. http://www.oecd.org/env/testguidelines

On-line MPBPVPWinTM user's guide, paragraph 7.3: Vapor Pressure. EPI Suite™ software, version 4.1, US EPA.

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(2):225-239.

Piroird, C., Ovigne, JM., Rousset, F., Martinozzo-Teissier, S., Gomes, C., Cotovio, J., Alépée, N. (2014). The Myeloid U937 Skin Sensitization Test (U-SENS) to address the Activation of Dendritic Cell Event in the Adverse Outcome Pathway for Skin Sensitization. Toxicology in Vitro, 29: 901-916.

Sanchez-Perez J, García del Río I, Alvarez Ruiz S, García Diez A. (2004). Allergic contact dermatitis from direct dyes for hair colouration in hairdressers' clients. Contact Dermatitis. 50(4): 261-272.

Spicer, C.W. et al. (2002). Hazardous poluant handbook, measurements, properties and fate in ambiant air. CRC Press LLC, FL, USA.

Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul Toxicol Pharmacol, 71(2): 337-351.

Vocanson M, Goujon C, Chabeau G, Castelain M, Valeyrie M, Floc'h F, Maliverney C, Gard A, Nicolas JF. (2006). Int Arch Allergy Immunol, 140(3): 231-238.

Wolpert, D. (1992). Stacked Generalization. Neural Networks 5, 241-259.

13. Supporting information

Details on the global chemical set (165 chemicals), *in chemico* / *in vitro* / *in silico* data as well as the calculation of the stacking meta-model are given in the Appendixes which form an integral part of this case study.

Appendix IV.1: Description of the statistical tools and the stacking meta-model

Appendix IV.2: Description of the global chemical set, the informative source data and the DIP outcomes.

14. Abbreviations and definitions

ACD : Allergic Contact Dermatitis

AOP : Adverse Outcome Pathway

DC : Dendritic Cell

DIP : Data Interpretation Procedure

DPRA : Direct Peptide Reactivity Assay

HRIPT : Human Repeat Insult Patch Test

IATA : Integrated Approach to Testing and Assessment

GPMT : Guinea Pig Maximalisation Test

LLNA : Local lymph node assay

(Q)SAR : (Quantitative) Structure Activity Relationship

U-SENS: Myeloid U937 cell-line activation test

CASE STUDY V

Integrated Decision Strategy for Skin Sensitisation Hazard

1. Summary

The ICCVAM Integrated Decision Strategy for Skin Sensitisation Hazard is a defined approach that uses a machine learning method, support vector machine, as the data interpretation procedure (DIP). Support vector machine is used to integrate the data from eight non-animal information sources. The information sources cover elements of the OECD AOP for skin sensitisation. The information sources include: (1) h-CLAT, an in vitro assay that addresses Key Event 3 of the AOP, dendritic cell activation; (2) an in silico read-across prediction based on the in vivo data for analogs, which aligns with all events of the AOP up to Key Event 4, T-cell proliferation, or the adverse outcome; and (3) six physicochemical properties (octanol:water partition coefficient, water solubility, vapor pressure, melting point, boiling point, molecular weight) that can affect skin penetration. The integrated decision strategy predicts skin sensitisation hazard based on LLNA reference data; however, it does not predict potency categorisation.

2. General information

2.1 Identifier:

Integrated Decision Strategy for Skin Sensitisation Hazard

2.2 Date:

28 April 2016

2.3 Author(s) and contact details:

Judy Strickland, Integrated Laboratory Systems, Inc., Contractor Supporting the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, Research Triangle Park, NC, USA, strickl2@niehs.nih.gov, Tel.: 1-919-281-1110 ext. 245 (primary contact).

Qingda Zang, Integrated Laboratory Systems, Inc., Contractor Supporting the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, Research Triangle Park, NC, USA, dan.zang@nih.gov, Tel.: 1-919-281-1110 ext. 425

*2.4 Template update(s):*Version 1.

2.5 Reference to main scientific papers:

Key paper:

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. http://dx.doi.org/10.1002/jat.3281

2.6 Proprietary aspects:

h-CLAT, one of the inputs included in the integrated decision strategy has intellectual property rights protected by Patent N. 4270702 only in Japan.

3. Endpoint addressed

3.1 Endpoint:

The Integrated Decision Strategy for Skin Sensitisation Hazard predicts the skin sensitisation hazard classification of a test substance as assessed by the murine local lymph node assay (LLNA), OECD TG 429 (OECD, 2010).

3.2 Species: Mouse

3.3 Additional information about the endpoint:

The endpoint addressed by the Integrated Decision Strategy for Skin Sensitisation Hazard applies to the AOP for skin sensitisation initiated by covalent binding to proteins (OECD, 2012). To make a prediction of skin sensitisation hazard for each target substance, the strategy uses the following data:

- Physicochemical property data, which influence skin absorption that occurs before the molecular initiating event (MIE).
- In vitro data from h-CLAT for Key Event 3, which quantifies changes in the expression of cell surface markers associated with dendritic cell activation (i.e., CD86 and CD54)
- In silico read-across information, which uses in vivo data from (1) LLNA, which aligns to Key Event 4, activation and proliferation of T-cells, and (2) guinea pig tests and human data, which align to the adverse outcome.

4. Definition of the purpose and regulatory relevance

The purpose of this approach is to predict the skin sensitisation hazard of a test substance, which can be used in hazard classification and labelling as required by U.S. regulatory authorities or the GHS (UN 2015). This integrated decision strategy does not permit the classification of sensitisers into GHS subcategories, 1A and 1B.

5. Rationale underlying the construction of the defined approach

The Integrated Decision Strategy for Skin Sensitisation Hazard predicts the skin sensitisation hazard classification of a test substance using three major information sources: 1) h-CLAT, an in vitro assay that addresses Key Event 3 of the AOP, 2) an in silico read-across prediction based on the in vivo data of analogs, and 3) six physicochemical properties that affect skin permeability. The information sources are integrated by a support vector machine model to predict LLNA outcomes. The rationale for using these particular information sources with this particular machine learning approach is that it had the highest performance among the 54 models evaluated (see Strickland et al. (2016) for further details).

Briefly, ICCVAM evaluated multiple information sources and machine learning approaches for performance in predicting LLNA sensitiser/nonsensitiser outcomes. The following information sources were evaluated:

- DPRA
- KeratinoSens
- h-CLAT
- In silico read-across using OECD QSAR ToolboxV3.2
- Six physicochemical properties: molecular weight, log octanol:water coefficient, log water solubility, log vapour pressure, melting point, and boiling point

Along with two test battery approaches, the following six machine learning methods were evaluated for integrating the data:

- Artificial neural network
- Naïve Bayes algorithm
- Classification and regression tree
- Linear discriminant analysis
- Logistic regression
- Support vector machine

Thirty-six models were evaluated using the six machine learning methods with six different combinations of the information sources. The best performing models for each machine learning method had accuracy of 77-92%, sensitivity of 74-90%, and specificity of 86-100% for predicting LLNA outcomes of the test set substances. Of the six machine learning methods, support vector machine had the highest performance with two different groups of information sources. Thus, support vector machine was evaluated with 18 additional combinations of information sources. The additional support vector machine models had accuracy of 73-96%, sensitivity of 73-95%, and specificity of 71-100% for the test set.

The rationale for the model selected for the Integrated Decision Strategy for Skin Sensitisation Hazard is that it had the highest performance of the 54 models evaluated for predicting LLNA outcomes. The model with the best performance used support vector machine to integrate data from (1) h-CLAT (positive or

negative), (2) read-across using QSAR ToolboxV3.2 (positive or negative), and (3) the six physicochemical properties (numerical values). For the test set, accuracy = 96%, sensitivity = 95% and specificity = 100%.

Rationale for the Information Sources Evaluated

The *in vitro* and *in chemico* assays were selected for evaluation because they have undergone international validation and have been peer reviewed by EURL ECVAM, and because relevant test data were publicly available. In addition, OECD test guidelines have been adopted for DPRA (OECD 2015a) and KeratinoSens (OECD 2015b). The test guideline for h-CLAT (OECD 2016) has been adopted in 2016. The OECD test guidelines recommend that these methods should be used in an integrated approach and offer the following performance statistics against the LLNA:

- DPRA: accuracy = 80% (125/157), sensitivity = 80% (88/109), specificity = 77% (37/48)
- KeratinoSens: accuracy = 77% (155/201), sensitivity = 78% (71/91), specificity = 76% (84/110)
- h-CLAT: accuracy = 85% (121/142), sensitivity = 93% (94/101), specificity = 66% (27/41)

These assays are relevant to the AOP for skin sensitisation (OECD 2012). The DPRA assesses the extent to which a test substance produces the MIE of the AOP, Key Event 1. KeratinoSens assesses whether a test substance activates keratinocytes to produce inflammatory cytokines and induce cytoprotective genes, Key Event 2. h-CLAT assesses whether a substance activates and mobilizes dendritic cells in the skin by the induction of inflammatory markers and surface molecules, Key Event 3.

An *in silico* read-across method, accomplished using QSAR ToolboxV3.2, was selected for evaluation because it could incorporate the assessment of auto-oxidation products and skin metabolites. Also, QSAR Toolbox is freely available software that is supported by OECD. The in silico read-across method uses in vivo skin sensitisation data (i.e., LLNA, guinea pig tests, and human outcomes) for analogs, and thereby includes absorption and skin metabolism and the key events up to and including Key Event 4 (from LLNA results), activation and proliferation of T-cells, and the adverse outcome (from guinea pig and human tests) in predictions of skin sensitisation hazard.

The six physicochemical properties were evaluated as information sources because they could influence chemical skin absorption, and thus, bioavailability. Such properties have been important for other models or weight-of-evidence assessments for skin sensitisation potential (Jaworska et al., 2013, 2015; Patlewicz et al., 2014). Although skin penetration is not a driving factor for skin sensitisation (Roberts and Aptula, 2008), model performance was higher when physicochemical properties were included as information sources.

6. Description of the individual information sources used (see Annex II)

Three types of information sources were used in the Integrated Decision Strategy for Skin Sensitisation Hazard to predict LLNA outcomes.

1. h-CLAT is an *in vitro* assay that addresses Key Event 3 of the AOP. The assay measures cell surface markers of activation in THP-1 cells, which serve as surrogates for dendritic cells. When a hapten is applied to the skin, the expression of CD54 and CD86 surface markers on skin dendritic cells are increased. Since CD54 is involved in DC migration to draining lymph nodes and CD86 stimulates T-

cell activation during antigen-presentation by dendritic cells, both surface markers are essential in the induction of skin sensitisation. The calculated relative fluorescence intensity is used as indicator of CD86 and CD54 expression. Substances producing ≥2-fold induction for CD54 and/or ≥1.5-fold induction for CD86 at cell viabilities ≥50% in two of three separate experiments are classified as sensitisers in this assay (Ashikaga et al., 2006). The positive or negative outcome of the h-CLAT was used in this integrated decision strategy.

2. The *in silico* read-across prediction using OECD QSAR ToolboxV3.2 aligns to two areas of the AOP based on the *in vivo* data used for analogs. It aligns to Key Event 4, because it uses LLNA data, and the adverse outcome, because it uses guinea pig tests and human data. A group of analogs for the test substance is formed based on structural and mechanistic categories. The available *in vivo* data (i.e., LLNA, guinea pig, and/or human) for the analogs are then used to make a prediction about the hazard classification of the test substance. The protocol, which is attached as supporting information, incorporates the evaluation of auto-oxidation products and skin metabolites when the test substance has no structural alerts for protein reactivity. The positive or negative read-across prediction was used as an input in the integrated decision strategy.

3. The six physicochemical properties required for the model are relevant to skin penetration, which occurs prior to the molecular initiating event of the AOP. The properties are: log octanol:water partition coefficient, log water solubility (moles/L), log vapor pressure (mm Hg), molecular weight (g/mol), melting point (^{o}C), and boiling point (^{o}C). Experimental values were preferred but predicted values were used if experimental values were unavailable. Because we could not find one or more values for 10 substances, we predicted them using quantitative structure-property relationship models that were built using binary molecular fingerprints and machine learning approaches such as support vector regression (Zang et al., 2016).

The following databases, which were selected because they are publically available, were used to identify experimental and predicted physicochemical property values:

- EPI Suite™ from SRC, Inc. (http://esc.syrres.com/interkow/EPiSuiteData.htm)
- ChemID*plus* from U.S. National Library of Medicine (NLM) Toxicology Data Network (TOXNET) (http://chem.sis.nlm.nih.gov/chemidplus)
- ChemSpider from Royal Society of Chemistry (http://www.chemspider.com/)
- Hazardous Substances Databank (http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB) from NLM Toxicology Data Network (TOXNET)
- Registration, Evaluation, Authorisation and Restriction of Chemicals *(*REACH*)* database from European Chemicals Agency - (http://echa.europa.eu/information-on-chemicals)

7. Data interpretation procedure applied

The Integrated Decision Strategy for Skin Sensitisation Hazard uses eight information sources for an individual chemical:

- h-CLAT binary result – 0 for negative or 1 for positive

- In silico read-across binary result – 0 for negative or 1 for positive
- Six numerical physicochemical properties:
    - Molecular weight (MW; g/mol)
    - Log octanol:water coefficient (LogP)
    - Log water solubility (LogS; mol/L)
    - Log vapour pressure (LogVP; mmHg)
    - Melting point (MP; ^{o}C)
    - Boiling point (BP; ^{o}C)

The data are integrated using a support vector machine, which produces a qualitative outcome of sensitiser (1, POS) or nonsensitiser (O, NEG). Support vector machine performs the sensitiser/nonsensitiser classification by finding the hyperplane that maximises the margin between the closest data points of each class (Shen et al., 2011). The process for deriving a prediction involves installing the open source software, R, and then using the software and the eight independent variables listed above to predict a qualitative sensitiser or nonsensitiser outcome. In the instructions that follow, commands for R are written in italics. Responses from R are written in italics preceded by ">."

1. Download the statistical software R from http://www.r-project.org/.

2. Install R on local machine.

3. Create a folder named *"C:/MachineLearning"*.

4. Save the training data (*"SkinData-Training-updated.txt"*), the test data (*"SkinData-Test-updated.txt"*), and example new chemical data (*"SkinData-New.txt"*) to the folder *"C:/MachineLearning"*.

5. Open R and set the working environment to *"C:/MachineLearning"*.

setwd("C:/MachineLearning")

6. Read in the training data with 94 chemicals and 17 columns.

SkinTraining <- read.table("SkinData-Training-updated.txt", header=T, sep="\t", as.is=T)

dim(SkinTraining)

> 94 17

Read in the test data with 26 chemicals and 17 columns.

SkinTest <- read.table("SkinData-Test-updated.txt", header=T, sep="\t", as.is=T)

dim(SkinTest)

> 26 17

7. The first two columns are chemical name and CASRN, the last column indicates if the chemical is for training or test set, and the other columns are variable values.

names(SkinTraining)

> *[1] "ChemicalName" "CASRN" "MW" "LogP" "LogS"*

> *[6] "LogVP" "MP" "BP" "Lys" "Cys"*

> *[11] "avg.Lys.Cys" "hCLAT" "DPRA" "Keratino" "OECD"*

> *[16] "LLNA" "Training.Test"*

8. There are 26 LLNA non-sensitisers (negatives) and 68 LLNA sensitisers (positives) in the training set.

TrainingClass <- factor(SkinTraining[,"LLNA"])

There are 7 LLNA non-sensitisers (negatives) and 19 LLNA sensitisers (positives) in the test set.

TestClass <- factor(SkinTest[,"LLNA"])

9. Set a weight for balancing the negative and positive classes, where *"7"* is an optimal value from a series of trials from 1 to 10.

wts <- 7/ table(TrainingClass)

10. Install and load package *e1071,* which contains the *svm* function.

install.packages("e1071") [Note: after this command, wait for a message from R to confirm that it has been downloaded.]

library(e1071)

11. The independent variables consist of h-CLAT data (hCLAT) as positive or negative, QSAR Toolbox read-across prediction of skin sensitisation hazard (OECD) as positive or negative, and six physicochemical properties: molecular weight (MW), octanol:water partition coefficient (LogP), water solubility in M (as LogS), vapor pressure in mmHg (as LogVP), melting point in ^{o}C (MP), and boiling point in ^{o}C (BP). The support vector machine model, which provided the best performance for predicting skin sensitisation hazard, applies the following weights to the variables: hCLAT (18.48), OECD (11.15), MP (8.93), Log S (8.87), BP (6.85), log VP (2.20), log P (0.59), and MW (0.28).

traindata<-SkinTraining[, c("hCLAT", "OECD", "MW", "LogP", "LogS", "LogVP", "MP", "BP")]

testdata<-SkinTest[, c("hCLAT", "OECD", "MW", "LogP", "LogS", "LogVP", "MP", "BP")]

12. Build the support vector machine model using the svm function. Two parameters *cost* and *gamma* are set to 400 and 0.081, respectively via an optimization procedure.

SVMmodel <- svm(traindata, TrainingClass, cost = 400, gamma = 0.081, class.weights = wts)

13. Predict the training set and produce a confusion matrix. 25/26 non-sensitisers and 66/68 sensitisers are correctly predicted.

PredTrain<-predict(SVMmodel, traindata)

table(PredTrain, TrainingClass)

> PredTrain NEG POS

> NEG 25 2

> POS 1 66

14. Predict the test set and produce a confusion matrix. 7/7 non-sensitisers and 18/19 sensitisers are correctly predicted.

PredTest<-predict(SVMmodel, testdata)

table(PredTest, TestClass)

> PredTest NEG POS

> NEG 7 1

> POS 0 18

15. Prediction of new chemicals. A fictional example "New Chemical" is in *SkinNew* is predicted as positive.

SkinNew <- read.table("SkinData-New.txt", header=T, sep="\t", as.is=T)

Newdata<-SkinNew[, c("MW", "LogP", "LogS", "LogVP", "MP", "BP", "hCLAT", "OECD")]

PredNew<-predict(SVMmodel, Newdata)

PredNew

>1

> POS

> Levels: NEG POS

8. Chemicals used to develop and test the DIP

8.1 Availability of training and test sets:

The training and test set substances are available in Appendix V.3.

8.2 Selection of the training set and test set used to assess the DIP:

The selection of the 120 substances in the database was based on the public availability of DPRA, KeratinoSens, h-CLAT, and LLNA data for each substance. The database was divided into training and test sets in the approximate proportions of 80% to 20%. Substances were first characterised by sensitiser or non-sensitiser classification by the reference LLNA result and then the substances in each class were parsed into structurally similar groups that were determined by the expert judgment of a chemist who examined the structures (Figure V.1). For the 87 sensitisers, 51 were organised into 19 structurally similar groups. To allocate approximately 20% of the chemicals into the test set and 80% into the training set, one substance from each structural group was allocated to the test set, and the remaining 32 substances were allocated to the training set. A subset of 17 of the 33 non-sensitisers was divided into seven structurally similar groups. One substance in each group was assigned to the test set and the remaining 10 substances were allocated to the training set. This procedure placed 68 substances (51 sensitisers and 17 nonsensitisers) into the training (42 substances) and test sets (26 substances). The remaining 52 substances (36 sensitisers and 16 nonsensitisers), which represented a structurally diverse set, were allocated to the training set. This process yielded a training set with 94 substances (78% of the 120), which consisted of 68 (72%) LLNA sensitisers and 26 (28%) LLNA non-sensitisers. The test set consisted of 26 substances (22% of the 120), with 19 (73%) LLNA sensitisers and 7 (27%) LLNA non-sensitisers. The training and test sets were selected so that they would be similar with respect to the distributions of LLNA outcome, product use categories, diversity of chemical structures, prehaptens and prohaptens, and protein binding mechanisms (Strickland et al., 2016).

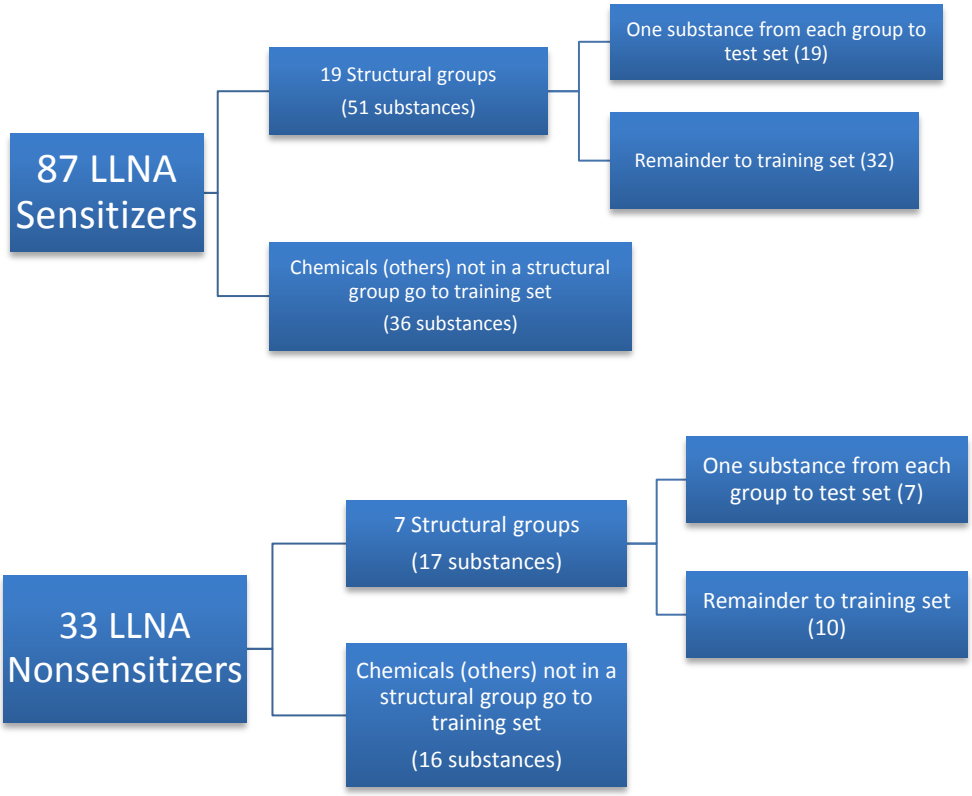


Figure V.1: Procedure for Distribution of 120 Substances into Training and Test Sets. Abbreviations: LLNA = murine local lymph node assay.

Structural groups were determined manually by expert judgment based upon examination of chemical structures.

8.3 Supporting information on the training and test sets:

The training and test sets can be found in Appendix V.3.

8.4 Other information on the training and test sets:

Not applicable.

9. Limitations in the application of the defined approach

There are limitations for the individual methods that serve as information sources. Regarding technical limitations, substances tested in the h-CLAT must be soluble or form a stable dispersion in saline or dimethyl sulfoxide and must be stable in cell culture conditions (OECD 2016). Test substances with log K_{ow} > 3.5 may produce false negative results; however, positive results for these substances are acceptable. Substances that fluoresce may interfere with the flow cytometry detection. Volatile substances may produce unreliable results due to evaporation or contamination of nearby wells. The h-CLAT also has limitations related to other characteristics of the test substances. Due to the limited metabolic capability of the cell line used and because of the experimental conditions, pro-haptens (which need metabolic activation to produce sensitisation) and pre-haptens (which need oxidation to produce sensitisation) may produce false negative results in the h-CLAT (OECD 2016).

QSAR Toolbox may be used to evaluate a broad range of chemicals, including metal compounds and the full range of organic functional groups and protein binding mechanisms relevant to skin sensitisation. The technical limitations include the size of the database of substances with *in vivo* skin sensitisation data, as the read-across predictions depend heavily on the availability of similar group members with such data. The algorithms used by QSAR Toolbox are determined by the category formed by the user; no training set of substances was used to determine the skin sensitisation algorithm. The read-across algorithm uses the skin sensitisation outcome that appears most often for the five nearest neighbors, based on log K_{ow}, to predict the skin sensitisation hazard of the target substance. If the log K_{ow} of the target substance is outside the log K_{ow} range for the analogs, the prediction is classified as unreliable because it is outside the applicability domain.

10. Predictive capacity of the defined approach

The performance of the Integrated Decision Strategy for Skin Sensitisation Hazard was assessed with respect to LLNA hazard classifications. Performance statistics for the test set of 26 substances (19 LLNA positive and 7 LLNA negative) and the training set of 94 substances (68 LLNA positive and 26 LLNA negative) are provided in Table V.1.

Table V.1. Performance for the Integrated Decision Strategy for Skin Sensitisation Hazard

Dataset	**TP**	**TN**	**FP**	**FN**	**Accuracy**	**Sensitivity**	**Specificity**	**Balanced Accuracy**	**Kappa Statistic**
Training (n=94)	66	25	1	2	97%	97%	96%	97%	0.921
Test (n=26)	18	7	0	1	96%	95%	100%	97%	0.906
Leave-one-out Cross-validation									
Dataset	**TP**	**TN**	**FP**	**FN**	**Accuracy**	**Sensitivity**	**Specificity**	**Balanced Accuracy**	**Kappa Statistic**
All (120)	74	31	2	13	88%	85%	94%	90%	0.716

The model misclassified only coumarin, as a false negative, in the test set. Coumarin had equivocal LLNA results (i.e., an equal number of positive (2) and negative (2) LLNA tests). It was a weak sensitiser in the positive tests (EC3 = 29.6%). In the reference for three of the LLNA results (1 negative and 2 positive), the sensitisation properties of coumarin were attributed to contaminants in commercial coumarin products (Vocanson et al. 2006). Both h-CLAT and the QSAR ToolboxV3.2 read-across had negative results for coumarin.

The model misclassified three substances in the training set: 3-phenoxypropiononitrile was false positive; and nonanoic acid and benzylidene acetone were false negatives. 3-phenoxypropiononitrile was positive in both h-CLAT and the QSAR ToolboxV3.2 read-across. Nonanoic acid and benzylidene acetone were positive in the h-CLAT and negative in the QSAR ToolboxV3.2 read-across. Nonanoic acid is a very weak LLNA sensitiser (EC3 = 35%) and a strong irritant (Anderson et al. 2011). The LLNA has been known to misclassify irritants that are not sensitisers (Anderson et al. 2011). Benzylidene acetone is a moderate sensitiser (EC3 = 3.7%). The physicochemical properties of the misclassified substances are well within the ranges of the remaining substances in the training and test sets.

An additional evaluation was also performed using a leave-one-out cross-validation procedure (LOOCV). While the original training and test sets were chosen to reflect the overall sensitiser/non-sensitiser distribution, and to adequately cover the range of structural diversity in the data, it could introduce bias. The LOOCV avoids any bias introduced during the selection of test and training sets. For LOOCV, the training and test set substances were combined, and the performance of the model was evaluated against every substance in the dataset when it appears in an external test set on its own. Thus, the procedure was performed 120 times with each substance used exactly once as the external validation set. The performance statistics are calculated by averaging individual values over the 120 runs.

The LOOCV resulted in lower performance, but due to the removal of selection (of training and test sets) bias, the LOOCV is more likely to reflect the performance when the model is applied to additional external datasets (Table V.1).

11. Consideration of uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

1. DIP structure

The uncertainties for the Integrated Decision Strategy for Skin Sensitisation Hazard that are related to the DIP's structure are associated with the AOP for skin sensitisation initiated by covalent binding to proteins (OECD 2012) and the machine learning approach used.

- Confidence in the Integrated Decision Strategy for Skin Sensitisation Hazard is high based on full coverage of the AOP. It covers skin penetration via the physicochemical properties, Key Event 3 via the h-CLAT input, and all events including Key Event 4 and/or the adverse outcome via the in silico input from read-across outcomes developed with in vivo data. The in vivo data endpoints used for read-across reflect all events from skin absorption and metabolism to the most terminal events in the AOP, T-cell proliferation and the adverse outcome.

- Confidence in the Integrated Decision Strategy for Skin Sensitisation Hazard is lower for substances that act by a molecular initiating event other than Key Event 1, covalent binding to proteins.

- Although the use of other machine learning approaches may also provide good performance for integrating the data sources, confidence in the support vector machine to integrate the data sources is high because it performed better than the five other machine learning approaches tested. It is a frequently used data mining algorithm (Wu et al. 2008).

This case study is based on a dataset of 120 substances. Confidence in the predictions of similar chemicals is high. Its applicability to predict the skin sensitisation hazard of substances that deviate significantly from the structures and physicochemical properties represented in the training and test datasets is lower.

2. Information sources used

The uncertainties for the Integrated Decision Strategy for Skin Sensitisation Hazard that are related to the DIP information sources include the following:

- h-CLAT results are less reliable for substances with log K_{ow} > 3.5 and for substances that require auto-oxidation or metabolism to cause sensitisation

- Read-across results from QSAR Toolbox 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 is unknown)

- Read-across results may come from substances that are not sufficiently similar to target substances

- Read-across results may vary by the assessor, who selects the analogs for the test chemical

- Some physicochemical properties are predicted rather than measured (i.e., those for which measured values could not be located)

The Integrated Decision Strategy for Skin Sensitisation Hazard was developed using the most prevalent result for source data when multiple outcomes were available for individual substances. This practice increases our confidence in the prediction when multiple results reflect the same outcome. Inconsistent results in the source data for a given chemical would reduce the confidence in the hazard predictions. Our confidence is also decreased when the multiple results reflect an equal number of positive and negative outcomes or when there is only one measurement per substance. For substances that had an equal number of positive and negative results, we used a positive result as the final outcome. This may overestimate the actual skin sensitisation hazard.

- Of the 120 substances, 69 had multiple h-CLAT results. For the h-CLAT data, only two substances had an equal number of positive and negative results. Multiple results were not obtained for the read-across predictions or the physicochemical properties.

3. Benchmark data

The Integrated Decision Strategy for Skin Sensitisation Hazard was developed using the most prevalent result for LLNA hazard when multiple tests were available for individual substances. This increases our confidence in the model prediction because it increases the confidence in our reference data when multiple results reflect the same outcome. Our confidence is decreased when the multiple results reflect an equal number of positive and negative outcomes or when there is only one measurement per substance. For substances that had an equal number of positive and negative results, we used a positive result as the final outcome. This may overestimate skin sensitisation hazard.

- Of the 120 substances, 45 had multiple LLNA results and four of these had an equal number of positive and negative results. Confidence in the prediction is higher when most tests agree than when there are an equal number of positive and negative tests.

4. No other sources of uncertainty are known.

11.2 Impact of uncertainty on the DIP's prediction

The impact of the individual sources of uncertainty on the overall uncertainty of the sensitiser/non-sensitiser predictions of the Integrated Decision Strategy for Skin Sensitisation Hazard are unknown. The individual sources of uncertainty could under- or over-estimate skin sensitisation potential. An assessment of uncertainty (such as a probability or confidence interval) is not included in the model prediction of skin sensitisation hazard.

12. References

Anderson SE, Siegel PD, Meade BJ. (2011). The LLNA: a brief review of recent advances and limitations. J Allergy, 2011: 424203. http://www.ncbi.nlm.nih.gov/pubmed/21747867

Ashikaga T, Yoshida Y, Hirota M, Yoneyama K, Itagaki H, Sakaguchi H, Miyazawa M, Ito Y, Suzuki H, Toyoda H. (2006). Development of an in vitro skin sensitization test using human cell lines: The human Cell Line Activation Test (h-CLAT) I. Optimization of the h-CLAT protocol. Toxicol In Vitro, 20: 767–773.

Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y, Miyazawa M, Ito Y, Nishiyama N, Itagaki H. (2010). A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). Altern Lab Anim, 38: 275-284.

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

Bauch C, Kolle SN, Fabian E, Pachel C, Ramirez T, Wiench B, Wruck CJ, van Ravenzwaay B, Landsiedel R. (2011). Intralaboratory validation of four in vitro assays for the prediction of the skin sensitizing potential of chemicals. Toxicol In Vitro, 25: 1162-1168.

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.

ECHA (2014). Illustrative examples with the OECD QSAR Toolbox workflow. Part 1: Introductory note. European Chemical Agency. http://echa.europa.eu/documents/10162/21655633/illustrative_example_qsar_part1_en.pdf

Emter R, Ellis G, Natsch A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. Toxicol Appl Pharmacol, 245: 281-290.

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: 1683-1696.

Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP. (2004). Development of a peptide reactivity assay for screening contact allergens. Toxicol Sci, 81: 332-343.

Gerberick GF, Troutman JA, Foertsch LM, Vassallo JD, Quijano M, Dobson RL, Goebel C, Lepoittevin JP. (2009). Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. Toxicol Sci, 112: 164-174.

Jaworska J, Harol A, Kern PS, Gerberick GF. (2011). Integrating non-animal test information into an adaptive testing strategy - skin sensitization proof of concept case. ALTEX, 28: 211-225.

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

Jaworska J, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M. (2015). Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. Arch Toxicol, 89: 2355-2383.

Kern PS, Gerberick GF, Ryan CA, Kimber I, Aptula A, Basketter DA. (2010). Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. Dermatitis, 21: 8-32.

Lalko JF, Dearman RJ, Gerberick GF, Troutman JA, Api AM, Kimber I. (2013). Reactivity of chemical respiratory allergens in the Peroxidase Peptide Reactivity Assay. Toxicol In Vitro, 27: 651-661.

Matheson JM, Zang Q, Strickland J, Kleinstreuer N, Allen DG, Lowit AB, Jacobs A, Casey W. (2015). ICCVAM integrated decision strategy for skin sensitization. The Toxicologist, 144 (1): 90.

Natsch A, Haupt T. (2013). Utility of rat liver S9 fractions to study skin-sensitizing prohaptens in a modified KeratinoSens assay. Toxicol Sci, 135: 356-368.

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: 1337-1352.

Nishijima T, Tokura Y, Imokawa G, Takigawa M. (1999). Photohapten TCSA painting plus UVA irradiation of murine skin augments the expression of MHC class II molecules and CD86 on Langerhans cells. J Dermatol Sci, 19: 202-207.

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, (2016). The OECD QSAR Toolbox website, OECD Publishing. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm

OECD, (2010). Test No. 429. Skin Sensitisation: Local Lymph Node Assay, OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects, OECD Publishing.

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

OECD, (2015). Test No. 442C. In Chemico Skin Sensitization: Direct Peptide Reactivity Assay (DPRA). In OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects, OECD Publsihing.

OECD, (2015). Test No. 442D. In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. In OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects, OECD Publishing.

OECD, (2016). Test No. 442E. In Vitro Skin Sensitisation: human Cell Line Activation Test (h-CLAT). In OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects, OECD Publsihing.

Patlewicz G, Kuseva C, Kesova A, Popova I, Zhechev T, Pavlov T, Roberts DW, Mekenyan O. (2014). Towards AOP application--implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. Regul Toxicol Pharmacol, 69: 529-545.

Roberts DW, Aptula AO. (2008). Determinants of skin sensitization potential. J Appl Toxicol, 28: 377-387.

The R Core Team, (2013). R: A Language and Environment for Statistical Computing. Reference Index.R, Foundation for Statistical Computing. http://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf

Shen MY, Su BH, Esposito EX, Hopfinger AJ, Tseng YJ. (2011). A comprehensive support vector machine binary hERG classification model based on extensive but biased end point hERG data sets. Chem Res Toxicol, 24(6): 934-949.

Strickland J, Choksi NY, Allen DG, Casey W. (2015). In silico predictions of skin sensitization using OECD QSAR Toolbox. The Toxicologist, 144 (1): 90. Abstract 422.

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: 1150-1162.

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.

Troutman JA, Foertsch LM, Kern PS, Dai HJ, Quijano M, Dobson RL, Lalko JF, Lepoittevin JP, Gerberick GF. (2011). The incorporation of lysine into the peroxidase peptide reactivity assay for skin sensitization assessments. Toxicol Sci, 122: 422-436.

United Nations, (2015). Globally Harmonized System of Classification and Labeling of Chemicals. 6th revised edition. ST/SG/AC.10/30/Rev.6, United Nations.

Vocanson M, Goujon C, Chabeau G, Castelain M, Valeyrie M, Floc'h F, Maliverney C, Gard A, Nicolas JF. (2006). The skin allergenic properties of chemicals may depend on contaminants - evidence from studies on coumarin. Int Arch Allergy Immunol, 140: 231-238.

Wu X, Kumar V, Quinlan JR, Ghosh J, Yang, Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D. (2008). Top 10 algorithms in data mining. Knowl Inf Syst, 14:1-37.

Zang Q, Mansouri K, Allen D, Kleinstreuer N, Casey W, Judson R. (2016). Prediction of physicochemical properties of environmental chemicals. J Chem Inf Model. Submission pending.

13. Supporting information

The following files are attached as supporting information:

- Appendix V.2: Protocol for Generating Read-across Prediction for Skin Sensitization Hazard Using QSAR Toolboxv3.2, which contains the protocol for using QSAR ToolboxV3.2 to make skin sensitisation hazard predictions

- Files to build and test the model with R software (Appendix V.3: Skin Data Training Updated, Skin Data Test Updated, and Skin Data New)

14. Abbreviations and definitions

AOP = adverse outcome pathway

BP = boiling point

DIP = data interpretation procedure

DPRA = direct peptide reactivity assay

GHS = Globally Harmonized System of Classification and Labeling of Chemicals

h-CLAT = human cell line activation test

ICCVAM = Interagency Coordinating Committee on the Validation of Alternative Methods

IDS = integrated decision strategy

LLNA = murine local lymph node assay

MIE = molecular initiating event

MP = melting point

MW = molecular weight

OECD = Organisation for Economic Co-operation and Development

P = octanol:water partition coefficient

QSAR = quantitative structure-activity relationship

S = water solubility

VP = vapor pressure

CASE STUDY VI

Classification consensus model of decision trees based on in silico descriptors to predict skin sensitisation hazard

1. Summary

The defined approach presented in this document consists of a classification consensus model of two decision trees based on in silico descriptors (referred to hereafter as "consensus model") to predict skin sensitisation hazard according to the UN Global Harmonised System (UN GHS, 2013), i.e. to distinguish sensitisers (UN GHS Category 1) from non-sensitisers. The consensus model does not provide potency information and, therefore, it cannot be used to sub-categorise skin sensitisers in GHS potency sub-categories 1A and 1B.

The descriptor with the highest discriminating power to distinguish sensitisers from non-sensitisers is used as first node of the classification trees and corresponds to an in silico prediction of protein reactivity, which addresses the Molecular Initiating Event (MIE) of the skin sensitisation adverse outcome pathway (AOP) and accounts for metabolic transformation. Within each tree of the consensus model, the protein binding predictions are refined by in silico descriptors mainly related to structural features. Skin sensitisation classifications obtained from the local lymph node assay (LLNA) (Basketter et al., 2002) were used as reference data for building the consensus model.

The consensus model was developed to maximise its sensitivity, therefore the number of false negatives predictions compared to LLNA classifications is limited. In addition, the consensus model shows a superior performance than the one of the LLNA in predicting responses in humans (Basketter et al., 2014).

Since the final prediction is based entirely on in silico descriptors the consensus model can be applied to large numbers of compounds for which chemical structures are available, furthermore given the in silico nature of the model, there is no variability associated to the final prediction.

2. General information

2.1 Identifier:

Consensus of classification trees for skin sensitisation hazard prediction**.**

2.2 Date:

29 April2016

2.3 Author(s) and contact details:

David Asturiol, Silvia Casati, Andrew Worth. Joint Research Centre – ECVAM, Via E. Fermi 2749, I-21027 (Italy). JRC-IHCP-computox@ec.europa.eu; andrew.worth@ec.europa.eu, Tel: +39 0332789566

2.4 RF update(s): Version 1.

2.5 Reference to main scientific papers:

Talete Srl, DRAGON (2010). http://www.talete.mi.it/ (Software for Molecular Descriptor Calculation)

Dimitrov, S. D. *et al*. (20015). Skin sensitisation: Modeling based on skin metabolism simulation and formation of protein conjugates. Int. J. Toxicol. 24**,** 189–204.

Asturiol D., Casati S., Worth A. (2016). Consensus of classification trees for skin sensitisation hazard prediction. Submitted to Toxicology in vitro.

2.6 Proprietary aspects:

TIMES and DRAGON are commercial software from LMC Bourgas and Talete Srl., respectively.

3. Endpoint addressed

3.1 Endpoint:

Skin sensitisation.

3.2 Species:

Mouse (primary target), Human (secondary target)

3.3 Additional information about the endpoint:

Skin sensitisation is a common information requirement in chemicals legislation. The mechanistic understanding of the endpoint is captured by an OECD-defined AOP. The AOP contains four Key events:

1- Protein binding reactions, reactivity and metabolism

2- Events in keratinocytes or biochemical pathways related to skin sensitisation

3- Events in dendritic cells

4- Events in lymphocytes

4. Definition of the purpose and regulatory relevance

The model generates hazard information for regulatory purposes, e.g. for classification and labelling under the GHS scheme and for contributing to satisfy information requirements under the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) legislation. The model does not provide skin sensitisation potency information and therefore cannot be used to sub-categorise skin sensitisers into GHS subcategories 1A and 1B.

5. Rationale underlying the construction of the defined approach

In vitro and in chemico methods for skin sensitisation testing address and measure mechanisms under specific key events (KE) of the AOP. In the ideal case, in vitro/in chemico methods would be 100% concordant with in vivo events, perfectly matching in vivo results; but this is probably an impossible task, given that in vitro/in chemico methods are not an exact reproduction of in vivo systems but in vivo models/approximations. Due to the surrogate character of these methods, it is possible that they provide (partially) overlapping information or generate discordant outcomes. In such a scenario, other descriptors might complement/fix those cases in which in vitro and in chemico methods are not concordant with in vivo results. Thus, it is expected that combinations of in vitro/in chemico and/or in silico descriptors will provide better skin sensitisation predictions than individual surrogate methods.

An objective and easily interpretable way to determine the best combination of in chemico, in vitro, and in silico descriptors for the prediction of skin sensitisation is to use a mathematical algorithm that automatically selects the best combination(s) of descriptors to predict given target data. The use of a mathematical algorithm for the selection of the descriptors instead of using a pre-selection of descriptors or expert judgement may lead to a combination of descriptors that were unrelated to the AOP or very difficult to interpret, but it also assures that the selected descriptors correspond to the combination that is the most discriminating one for skin sensitisation hazard prediction. The ideal case, of course, is the one in which the algorithm selects a combination of descriptors that is in perfect agreement with the AOP. See next section for the description of the descriptors selected in the current DIP.

We built a dataset of organic compounds with in vivo skin sensitisation (LLNA) data and data generated with in chemico and in vitro validated methods, i.e. Direct Peptide Reactivity Assay (DPRA; OECD TG 442C), KeratinoSens™ (OECD TG 442D), and the human Cell Line Activation Test (h-CLAT; OECD TG 442E). These methods address KE-1, KE-2, and KE-3 of the AOP, respectively, and their corresponding readouts account for the ability of a chemical to bind to proteins (DPRA), ability of a test chemical to activate the Keap-1-Nrf2-ARE pathway in keratinocytes (KeratinoSens^{TM}), and capacity of a test chemical to induce the activation of membrane markers in a human monocytic cell line (h-CLAT). Further details on the validated methods and the corresponding readouts can be found in the respective OECD TGs. The readouts provided by the validated methods were supplemented with in silico predictions (descriptors) calculated with various licensed and free software packages, i.e. OECD QSAR ToolBox, Derek Nexus, Toxtree, Dragon, Vega, TIMES, and ADMET Predictor. The descriptors obtained from the predictions vary from simple structural features of the test chemicals, e.g. presence of OH groups, to more complex predictions accounting for protein binding, formation of metabolites, or skin sensitisation hazard predictions.

The model to predict skin sensitisation hazard (LLNA) was developed using Weka, a free software package that includes a mathematical algorithm to build decision trees (C4.5) (Salzberg, 1994). In short, the algorithm uses a subset of the compounds present in the dataset with their corresponding descriptors and classification (sensitiser/non-sensitiser), the so-called training set, to "train" a decision tree to classify the compounds into sensitisers and non-sensitisers. The tree consists of a sequence of nodes, each of which corresponds to a descriptor with a defined threshold that splits the data into two subsets. The algorithm

selects the descriptors and corresponding thresholds for each node so that the final splitting yields the best match to the real classes, i.e. the splitting is based on the normalized information gain (difference in entropy). The compounds not included in the training set form the so-called "test set" and are used to "validate" the decision tree generated with the training set. This "validation" shows whether the tree generated with the training set is valid for use with another set of compounds for which it has not been optimised, i.e. the test set.

The two "best" decision trees generated for the different subsets of data were used in a conservative consensus approach to predict skin sensitisation hazard, *i.e.* a negative prediction (non-sensitiser) is only considered if both trees yield concordant negative predictions, in any other case the prediction is considered positive (sensitiser).

Decision trees were used as prediction models because of their transparent, reproducible and interpretable nature.

6. Description of the individual information sources used (see Annex II)

TIMES Software (v 2.27.13)

Chemical/biological mechanism addressed by the information source: MIE / Key event 1, Covalent modification of proteins by reactive chemicals.

Biological relevance of the test system used in the information source: Reactivity is key for the MIE, thus intrinsic or metabolically triggered reactivity has a key biological relevance.

Readout of the information source: TIMES (LMC-Bourgas) is a licensed software package with prediction modules for several endpoints including skin sensitisation. The skin sensitisation module includes a skin metabolism and autoxidation simulator and it outputs multiple readouts including skin sensitisation hazard and potency prediction, transformation active alert driving sensitisation, transformation reliability, vapour pressure, structural domain, and amount of protein-hapten adduct formation. Only the two last readouts are used in our approach:

- Total Structural domain: It determines whether the test compound (percentage of two-atom fragments) falls in the applicability domain of TIMES

- Amount of protein-hapten adduct formation (TIMES-ProtBind): It is a prediction of the amount of test chemical – either parent compound or any of the predicted metabolites or autoxidation products – that will bind covalently to a mole of skin protein

TIMES-SS uses 2D structural information of the test compound to match it with 236 possible transformations. The parent compound and resultant metabolites are subsequently matched against 47 alerting groups associated with skin protein reactivity. 3D based QSARs, steric effects around the active site, molecular size, shape, solubility, lipophilicity and electronic properties are also taken into account for the prediction.

DRAGON (v 6.0.7)

Chemical/biological mechanism addressed by the information source: not applicable

Biological relevance of the test system used in the information source: not applicable

Readout of the information source: DRAGON (Talete Srl.) is a licensed software package that generates a large number of descriptors (~4885) ranging from simple structural features like presence of alcohol groups in a molecule to 3D and 4D descriptors derived from graph theory (see http://www.talete.mi.it/help/dragon_help/index.html?new_molecular_descriptors.htm and links therein).

Of all the descriptors generated by DRAGON, the following were selected for inclusion in the consensus model:

Decision Tree 1:

Dragon-Mor32s: 3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) are derived from Infrared spectra simulation using a generalised scattering function. This descriptor corresponds to signal 32 weighted by l-state.

Dragon-SpDiam_EA(bo): Spectral diameter from edge adjacency matrix weighted by bond order

Dragon-O-056: Presence of alcohol (-OH) groups

Dragon-Eig08_AEA(bo): Eigenvalue n. 8 from augmented edge adjacency matrix weighted by bond order

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).

Decision Tree 2:

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 (Todeschini et al., 1994, 1997). 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 weighting 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.

Dragon-H-052: H attached to C(sp3) with 1 heteroatom attached to the next C.

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).

Dragon-Mor24u: 3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) are derived from Infrared spectra simulation using a generalised scattering function. This descriptor corresponds to signal 24 un-weighted.

7. Data interpretation procedure applied

a) The predictions of the consensus model are qualitative (sensitiser or non-sensitiser) and are the result of combining the outputs of the following decision trees (see Figure VI.1 and Figure VI.2) in a conservative consensus way (see Table VI.2).

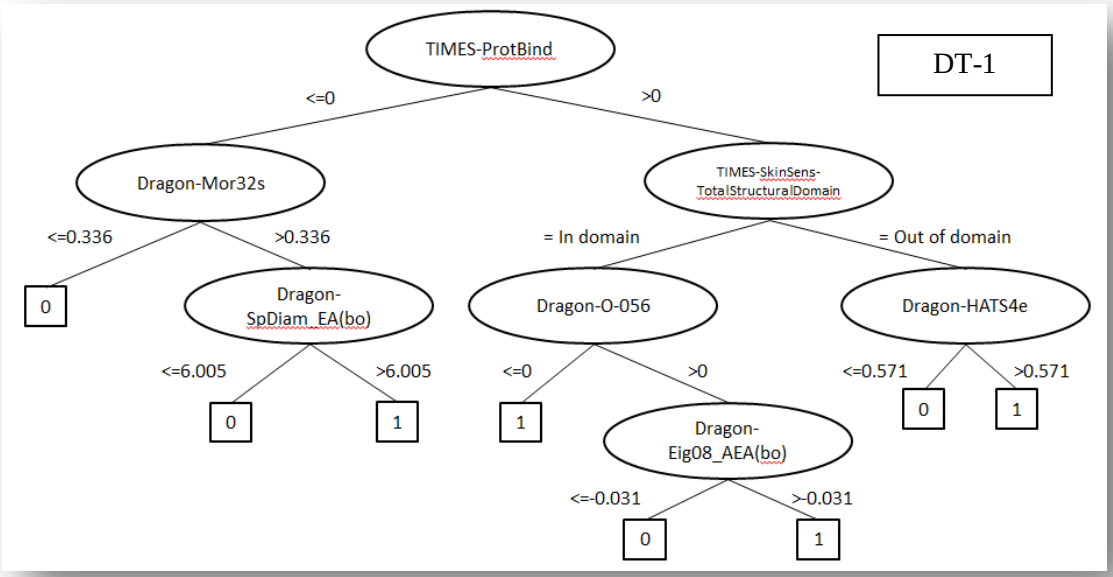


Figure VI.1. First decision tree of the consensus model (DT-1). Negative predictions (non-sensitiser) are indicated with a 0 and positive predictions (sensitiser) are indicated with a 1.

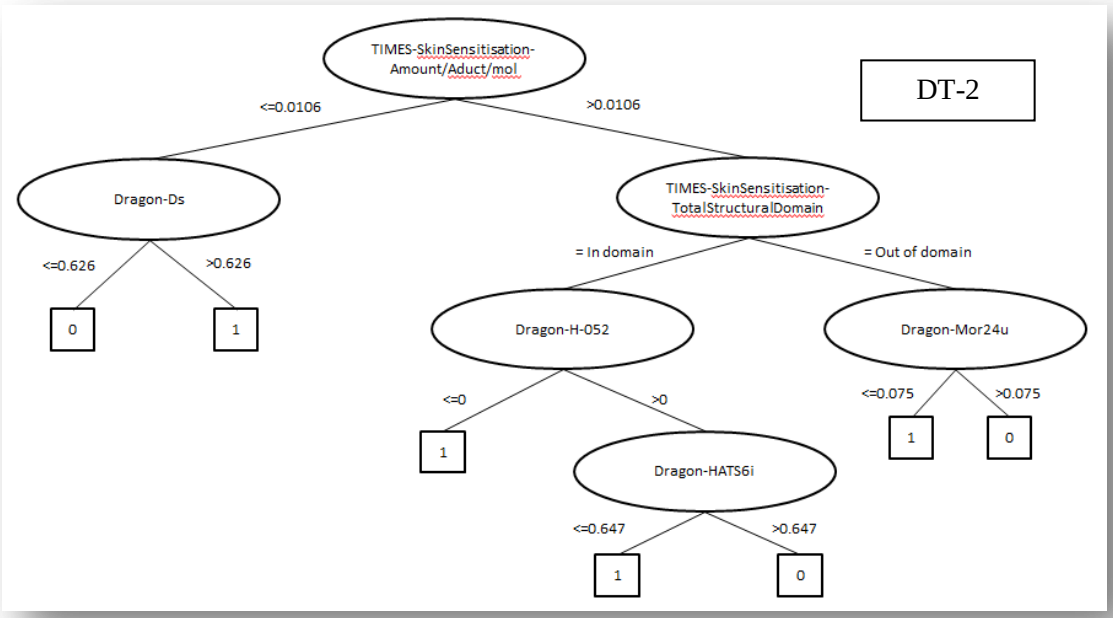


Figure VI.2. Second decision tree of the consensus model (DT-2). Negative predictions (non-sensitiser) are indicated with a 0 and positive predictions (sensitiser) are indicated with a 1.

In order to obtain the output of the decision trees several chemical descriptors need to be generated with different software packages, *i.e.* TIMES-SS (LMC-Bourgas) and DRAGON (Talete Srl.). The descriptors that are needed for each decision tree and the software packages from which they can be obtained are listed

in Table VI.1. The chemical structure of the compounds, which can be provided as SMILES codes, is needed in order to run the predictions.

Table VI.1. List of the descriptors that are needed for each decision tree and software packages from which they can be obtained.

Descriptor	Software package	Necessary for
Skin Sensitisation Amount/Aduct/mol	TIMES-SS	DT-1 & DT-2
Skin Sensitisation Total Structural Domain	TIMES-SS	DT-1 & DT-2
Mor32s	DRAGON	DT-1
SpDiam_EA(bo)	DRAGON	DT-1
O-056	DRAGON	DT-1
Eig08_AEA(bo)	DRAGON	DT-1
HATS4e	DRAGON	DT-1
Ds	DRAGON	DT-2
H-052	DRAGON	DT-2
HATS6i	DRAGON	DT-2
Mor24u	DRAGON	DT-2

The conservative consensus approach that needs to be applied to obtain the consensus model prediction is shown in Table VI.2.

Table VI.2. Derivation of the predictions of the conservative consensus model from DT-1 and DT-2 outputs. Negative predictions (non-sensitiser) are indicated with a 0 and positive predictions (sensitiser) are indicated with a 1.

Output from DT-1	Output from DT-2	Consensus Prediction
1	1	1
1	0	1
0	1	1
0	0	0

8. Chemicals used to develop and test the DIP

8.1 Availability of training and test sets:

The list of chemicals used to develop the consensus model is attached as supporting information (S1). The role of each chemical in the process to build each of the 2 individual decision trees is also indicated.

8.2 Selection of the training set and test set used to assess the DIP:

The training and test sets were selected following a structural diversity split for LLNA positives and LLNA negatives, respectively. The following procedure was used to define the training and test sets (see Figure VI.3 for a schematic representation of the procedure):

a. Structural descriptors were calculated for all the compounds

b. The compounds were then divided between positive compounds (P) and negative compounds (N) according to their LLNA-derived classification.

c. A principal component analysis (PCA) on the structural descriptors was performed keeping the principal components that accounted for 90% of the variance. This step was carried separately for Ps and Ns.

d. The PCA vectors derived in step c were used to cluster the compounds in structurally similar bins. Ps and Ns were clustered in as many bins as the number of compounds divided by 10 and 5, respectively.

e. 80% of the compounds in each cluster were randomly picked in a stratified way to form the training set. The remaining 20% were assigned to the test set. This procedure was applied to the cluster of Ps and Ns independently.

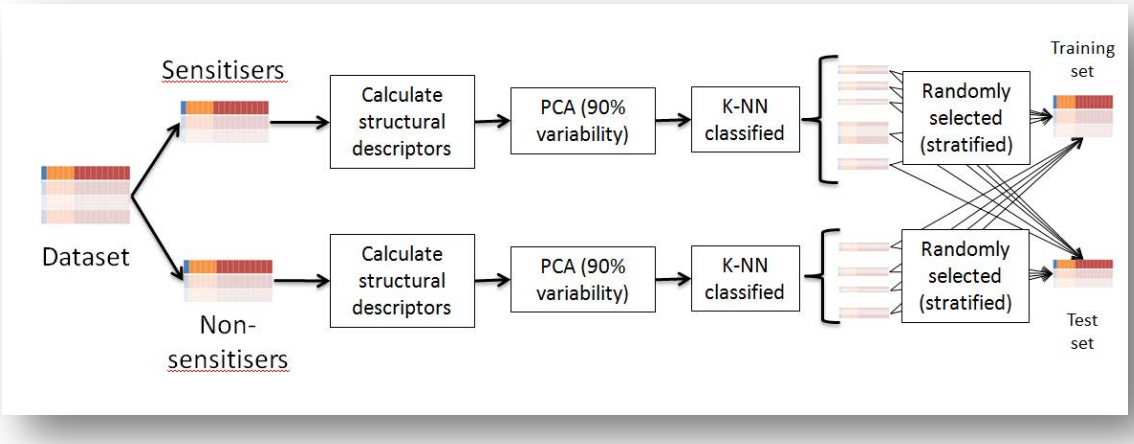


Figure VI.3. Data splitting procedure

8.3 Supporting information on the training and test sets:

The training and test sets are included as supporting information with the chemical name, SMILES, DT-1 output, DT-2 output, consensus model prediction, and LLNA hazard classification. The database including LLNA, human, in chemico, and in vitro data, and the in silico descriptors used in the consensus model can be found on the ECVAM website https://eurl-ecvam.jrc.ec.europa.eu/.

8.4 Other information on the training and test sets:

Chemicals used in the training and test sets were obtained from publicly available sources (EURL-ECVAM, 2012, 2014, 2015; Bauch et al., 2012; Emter et al., 2010; Gerberick et al., 2007, 2004; Natsch and Emter, 2008; Natsch et al., 2013; Nukada et al., 2013; Takenouchi et al., 2013).

9. Limitations in the application of the defined approach

The classification model is applicable to all organic chemical entities with a defined molecular structure. Thus, it cannot be applied to polymers, metals, multi-constituent substance, substances of unknown or variable composition, complex reaction products or biological materials (UVCB) and mixtures.

The applicability domain (AD) of the model is defined by the chemicals of the training sets that are correctly predicted by their corresponding trees. The ranges of the values of the descriptors used in the training set with ±30% confidence are shown below (Table VI.3).

Table VI.3. Applicability domain of the consensus model

Descriptor	Range	Necessary for
Dragon-SpDiam_EA(bo)	[-1.32 - 10.15]	CT-1
Dragon-Eig08_AEA(bo)	[-1.47 - 3.85]	CT-1
Dragon-Mor32s	[-2.91 - 3.35]	CT-1
Dragon-HATS4e	[-0.25 - 1.93]	CT-1
Dragon-O-056	[0 - 8]	CT-1
Dragon-Mor24u	[-2.23 - 1.39]	CT-2
Dragon-Ds	[0.11 - 1.16]	CT-2
Dragon-HATS6i	[-0.35 - 2.73]	CT-2
Dragon-H-052	[0 - 13]	CT-2
TIMES-SkinSensi-AmountAduct/mol/	[0 - 2.3]	CT-1&CT-2

TIMES-SkinSensi-Total Structural domain	[Out of domain, In domain]	CT-1&CT-2

10. Predictive capacity of the defined approach

The performances of the individual decision trees against LLNA classifications and that of the consensus model against LLNA and human classifications are shown below.

10.1 Statistics of the consensus approach

The number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) for the training and test sets of the individual decision trees and the consensus model with respect to LLNA classifications are provided next. Common Cooper statistics (accuracy, sensitivity, specificity) as well as the corresponding balanced statistics (prevalence independent) for accuracy (Bal-Acc), positive predictive value (Bal-PPV), and negative predictive value (Bal-NPV) are also provided (Tables VI.4, VI.5 and VI.6).

Table VI.4. Decision Tree 1 statistics (from dataset A[1], 223 compounds) against LLNA

Subset	TP	TN	FP	FN	Accuracy	Sensitivity (TP/TP+FN)	Specificity (TN/TN+FP)	Bal-Acc	Bal-PPV	Bal-NPV
Training (80%)	99	71	4	2	0.97	0.98	0.95	0.96	0.95	0.98
Test (20%)	23	16	3	4	0.85	0.85	0.84	0.85	0.84	0.85

Table VI.5. Decision Tree 2 statistics against LLNA

Subset	TP	TN	FP	FN	Accuracy	Sensitivity (TP/TP+FN)	Specificity (TN/TN+FP)	Bal-Acc	Bal-PPV	Bal-NPV
Training (80%)	125	72	6	9	0.93	0.93	0.92	0.93	0.92	0.93
Test (20%)	29	18	3	7	0.82	0.81	0.86	0.83	0.85	0.82

[1] DT-1 was derived from a subset of the dataset that consists of all compounds for which KeratinoSens™ data were available (223 compounds)

Table VI.6. Consensus model statistics against LLNA and against human data (Basketter et al., 2014)

Subset	TP	TN	FP	FN	Accuracy	Sensitivity (TP/TP+FN)	Specificity (TN/TN+FP)	Bal-Acc	Bal-PPV	Bal-NPV
vs LLNA	166	84	15	4	0.93	0.98	0.85	0.91	0.87	0.97
vs Human	57	23	13	6	0.81	0.90	0.64	0.77	0.71	0.87

The compounds that resulted false negatives in the consensus model are presented in Table VI.7.

Table VI.7. FN results of the consensus approach with respect to LLNA

– **Chemical Name**	– **Protein-hapten adduct formation/mol (TIMES)**	– **Rational**
– **Hexyl salicylate**	– 0	– No alerts for reactivity. It is an irritant. It is probably a FP in LLNA (Urbisch et al., 2015). – Data on h-CLAT shows it is positive for CD54 activation, only.
– **Pyridine**	– 0	– No alerts for reactivity. It is probably a FP in LLNA (Urbisch et al., 2015). – Data on h-CLAT shows it is positive for CD86 activation only, with very high CV75 and EC150(CD86)
– **Xylene**	– 0	– No alerts for reactivity. It is probably a FP in LLNA (Urbisch et al., 2015). – It is also negative in all in vitro methods.
– **Tocopherol**	– 0.162	– Predicted by TIMES to have metabolites that react with proteins. It falls outside Total Structural Domain of TIMES. It is a NS in humans. It is probably a FP in LLNA (Urbisch et al., 2015)

Table VI. 6 shows that the consensus model predicts 88 chemicals as negative (non-sensitisers) and that 84 of these correspond to chemicals classified as non-sensitisers by the LLNA. Thus, the other 4 predictions are FNs and correspond to hexyl salicylate, pyridine, xylene, and tocopherol. These chemicals are listed in Table VI.7 together with the possible reasons for the misprediction and some additional considerations. TIMES-ProtBind values are also listed.

i) Hexyl salicylate, pyridine, and xyline do not have chemical features that indicate they are reactive and they are predicted to be non-reactive to proteins by TIMES. This is supported by the fact that the three chemicals are negative in the DPRA and KeratinoSens™, and xyline is also negative in the h-CLAT assay. Therefore, most certainly neither these compounds nor their metabolites/products are reactive to proteins. In addition, xyline is a well-known FP in the LLNA, hexyl salicylate is a known irritant (which is a confounding factor in the LLNA and may give FPs results), and pyridine is positive for only one of the markers measured in h-CLAT (*i.e.* CD54) but at very high cytotoxic concentrations. Thus, evidence suggests that the three chemicals might be FPs in the LLNA. This interpretation of the mispredictions is aligned with what was reported in other works (Basketter et al., 2014; Urbisch et al., 2015). The other compound, tocopherol, is also a FP in the LLNA and is negative in h-CLAT. However, it is predicted reactive to proteins by TIMES and falls outside the total structural domain of TIMES. This chemical falls in the "weak part" of the model.

Due to the structure of the consensus approach, the number of FPs is rather high. Table VI.8 shows that when DT-1 and DT-2 predictions are not concordant, the possibility of having FPs is higher.

Table VI.8. Summary of TP and FP with respect to DT-1 and DT-2 predictions

Substances predicted to be sensitisers by the consensus model			
Result	**DT-1 & DT-2 concordant**	**DT-1 and DT-2 discordant**	**All**
TP	144	22	166
FP	1	14	15

The consensus model predicts 181 substances to be sensitisers, 166 of which are TP. DT-1 and DT-2 have concordant positive predictions on 145 compounds and only 1 of these is a FP. In other words, 14 out of 15 FPs of the consensus model correspond to discrepancies between DT-1 and DT-2. Thus, those predictions with concordant outputs from DT-1 and DT-2 have higher confidence than those that are not concordant. The list of FP compounds with TIMES protein binding predictions and the individual predictions of DT-1 and DT-2 are shown in Table VI.9.

Table VI.9. FP results of the consensus model with respect to LLNA

– Chemical Name	– Protein-hapten adduct formation/mol (TIMES)	– DT-1 Prediction	– DT-2 Prediction
ii) 1-Chloro-2-methyl-3-nitrobenzene	iii) 0	iv) 0	v) 1
vi) 1-Iodohexane	vii) 0.93	viii) 1	ix) 0
x) 1-Methoxy-4-methyl-2-nitrobenzene	xi) 0.29	xii) 0	xiii) 1
xiv) 1-Octen-3-yl acetate	xv) 1.44	xvi) 1	xvii) 0
xviii) 2-Mercaptobenzoxazole	xix) 0.75	xx) 0	xxi) 1
xxii) 2-Nitro-3-pyridinol	xxiii) 0	xxiv) 0	xxv) 1
xxvi) 3-Hydroxy-4-nitrobenzoic acid	xxvii) 0	xxviii) 0	xxix) 1
xxx) Dihydromyrcenol	xxxi) 0.31	xxxii) 1	xxxiii) 0
xxxiv) Ethyl benzoylacetate	xxxv) 0.46	xxxvi) 1	xxxvii) 0
xxxviii) Geranyl nitrile	xxxix) 0.74	xl) 1	xli) 1
xlii) Hydrocortisone	xliii) 0.24	xliv) 0	xlv) 1
xlvi) N-p-benzonitrile menthanecarbosamide	xlvii) 0.93	xlviii) 0	xlix) 1
l) p-Nitro-benzaldehyde	li) 0.07	lii) 1	liii) 0
liv) Saccharin	lv) 0.58	lvi) 1	lvii) 0
lviii) Sodium 1-nonanesulfonate	lix) 0	lx) 0	lxi) 1

11. Known uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

1. The DIP's structure

- Our approach (consensus model) is mostly based on key event 1. This could be a source of uncertainty for those compounds that are positive for key event 1 but are not sensitisers, or for those that are negatives for key event 1 but are sensitisers. The different nodes of the decision trees take care of these possible mispredictions and modulate the outcome of the prediction based on key event 1. However, our results show that these corrections are minimal as the descriptor that accounts for key event 1 has a rather high predictive power, most certainly because it addresses metabolism.

- Chemicals that are predicted to be reactive to proteins by TIMES and to fall outside the skin sensitisation total structural domain of TIMES have higher uncertainty than others. In total, in our database there are 31 compounds that are predicted to be reactive to proteins by TIMES and to fall outside the structural domain. Out of these 31 compounds, there are 7 TN, 15 TP, 8 FP, and 1 FN. These results clearly show that this is the weakest point of the consensus model as it accounts for half of the mispredictions of the model.

- Key events 2 and 3 are not included. Models using KE-2 and KE-3 were also generated, but their performance was significantly lower than the current ones. Our study also showed that adding *in vitro* results from KE-2 or KE-3 to our approach added no predictive value. In fact, one of the descriptors related to KE-2 that was included in these models corresponded to IC50 (concentration of test compound yielding 50% of cell viability) of KeratinoSens^{TM}, which is not directly related to KE-2.

- Key event 4 is not included due to lack of available tests. However, it is assumed that a given amount of modified protein and a given amount of induction of early cellular events will lead to a predictable lymphocyte proliferation. Lymphocyte proliferation depends directly on these events and only indirectly on the test chemical nature. However, the nature of the test chemical appears to be sufficient information for hazard prediction.

- The consensus model was built to boost sensitivity to maximise human health protection and, therefore, was built to reduce the number of FNs at the expense of increasing the number of FPs. In fact, most FPs of the consensus model correspond to discrepancies between DT-1 and DT-2. Thus, one can evaluate a positive result by considering the individual results of DT-1 and DT-2.

2. The information sources used within the defined approach

- The defined approach uses as descriptor the output of a software package (TIMES-SS) that contains many compounds of our database in its training set. In principle, this could be a source of artificially high performance of the proposed defined approach. However, we have analysed independently the performances of those compounds that belong to the training set of TIMES-SS and those that do not belong, and they are identical, suggesting that the model is not overfitted to data in the TIMES-SS training set.

- Reproducibility. Given that the descriptors used in the defined approach are all *in silico* predictions, the reproducibility of the results is 100% (provided that the same software version and procedures to derive the descriptors are used).

- The use of 3D descriptors like Mor32s and Mor24u involves the calculation of 3D structures for the test chemicals. These descriptors are structure-dependent and can be sensitive to small variations in the 3D structure. Thus, their values may vary depending on the optimization process of the chemical structure that is carried out. Long linear molecules with rotatable bonds are more prone to lead to different structures than planar aromatic chemicals.

4. **Benchmark data used**

- The variability of the reference in vivo data inevitably affects the accuracy of prediction. This variability originates from the intrinsic variability of the biological model and from the testing variability (between- and within-laboratory variability). The LLNA between-laboratory concordance for sensitiser/non-sensitiser classifications is around 80% (ICCVAM database, see case study V). Probably, the concordance of our subset of LLNA reference data is higher than 80% because it was obtained from databases that were previously used to evaluate the performance of in chemico and in vivo methods which underwent some level of expert scrutiny.

- The uncertainty is probably higher for the human data since this was obtained with different test protocols and from different individuals. Correlation of the LLNA with the human lowest observed effect level (LOEL) is far from perfect (Natsch et al., 2015) which can partly be attributed to these limitations in the human dataset. However, we do not use the human data in the model development, but simply show the results against the human data reported in Basketter et al. 2014.

11.2 Impact of uncertainty in the DIP's prediction

The defined approach does not provide an explicit assessment of prediction uncertainty, although the user can give higher confidence to those chemicals that:

- fall within the applicability domain of the defined approach

- have concordant individual predictions generated by DT-1 and DT-2

- are predicted to be non-sensitisers by the consensus model (confidence is even higher for those chemicals that are also predicted to be non-reactive to proteins by TIMES)

On the other hand, the predictions with lower confidence are those that include both: TIMES predictions of positive reactivity with proteins and to fall outside the Total Structural Domain of TIMES.

12. References

Basketter, D. A. et al. (2002). Local lymph node assay - validation, conduct and use in practice. Food Chem. Toxicol, 40, 593–598.

Basketter, D. A. et al. (2004). Categorization of chemicals according to their relative human skin sensitizing potency. Dermatitis, 25, 11–21.

Bauch, C., Kolle, S.N., 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.

Emter, R., Ellis, G. & Natsch, A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. Toxicol. Appl. Pharmacol, 245, 281–290.

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

EURL-ECVAM, (2014). KeratinoSens™ relevant validation study reports https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/recommendation-keratinosens-skin-sensitisation

EURL-ECVAM. (2015). h-CLAT validation study reports. https://eurl-ecvam.jrc.ec.europa.eu/news/news_docs/h-clat-validation-study-report

Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., Lepoittevin, J.P. (2004). Development of a peptide reactivity assay for screening contact allergens. Toxicol. Sci, 81, 332–343.

Gerberick, G.F., Vassallo, J.D., Foertsch, L.M., Price, B.B., Chaney, J.G., Lepoittevin, J.P. (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. Toxicol. Sci, 97, 417–427.

Natsch, A., Emter, R. (2008). Skin Sensitizers Induce Antioxidant Response Element Dependent Genes: Application to the In Vitro Testing of the Sensitization Potential of Chemicals. Toxicol. Sci, 102, 110–119.

Natsch, A. et al. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. J. Appl. Toxicol, 33, 1337–1352.

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., Miyazawa, M., Kazutoshi, S., Sakaguchi, H., Nishiyama, N. (2013). Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicol. In Vitro, 27, 609–18.

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 ENV/JM/MONO(2012)10/PART2.http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part2&doclanguage=en

Salzberg, S. L. (1994). C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., Mach. Learn. 16, 235–240.

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.

Todeschini, R., Lasagni, M. & Marengo, E. (1994). New molecular descriptors for 2D and 3D structures. Theory. J. Chemom, 8, 263–272.

Todeschini, R. & Gramatica, P. (1997). 3D-modelling and prediction by WHIM descriptors. Part 5. Theory development and chemical meaning of WHIM descriptors. Quant. Struct. Relationships 16, 113–119.

United Nations, (2013). The Globally Harmonized System of Classification and Labelling of Chemicals (GHS) Hazard communication. Fifth revised edition.

Urbisch, D. et al. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul. Toxicol. Pharmacol, 71, 337–351.

13. Supporting information

All relevant information can be easily obtained as open access Excel file (SI_Dataset.xls.) at the publisher's website http://www.sciencedirect.com/science/article/pii/S0887233316301436.

14. Abbreviations and definitions

Acc Accuracy

AOP Adverse Outcome Pathway

Bal-Acc Balanced accuracy

Bal-NPV Balanced negative predictive value

Bal-PPV Balanced positive predictive value

DPRA Direct Peptide reactivity Assay

DT-1 Decision Tree 1

DT-2 Decision tree 2

ECVAM European centre for the validation of alternative methods

FN False negative

FP False positive

GHS	Globally harmonized system
h-CLAT	Human cell line activation test
KE	Key event
LLNA	Local lymph node assay
LOEL	Lowest observed effect level
MIE	Molecular initiating event
N	Negatives (non-skin sensitisers)
NPV	Negative predictive value
P	Positives (skin sensitisers)
PPV	Positive predictive value
TG	Test guideline
TN	True negative
TP	True positive
UN	United Nations

CASE STUDY VII

Sensitiser potency prediction based on Key event 1 + 2: Combination of kinetic peptide reactivity data and KeratinoSens® data

1. Summary

Quantitative dose-response data from KeratinoSens® (concentrations inducing luciferase and cytotoxicity) do correlate to sensitiser potency in mice (expressed as EC3 values) and in humans (expressed as DSA_{05} values). The same holds true for the kinetic reaction rate between sensitisers and peptides in a modification of the DPRA.

These quantitative data are combined in a simple multiple regression model to yield estimated an EC3 or human DSA_{05} values.

Global regression models based on all available data were created for both the human and mouse prediction.

Chemicals which fit into defined mechanistic domains can be further predicted with domain-specific regression models. These predictions are preferred over the global predictions and more accurate.

However, for chemicals for which no such domain models exist (due to scarcity of in vivo data), the potency is predicted with the global regression model.

In this case study, the derivation of the relevant input parameters from the in vitro and in chemico tests is described, and the regression models are presented. Thus with the information provided here, these estimations can be directly made based on data from these in vitro, in chemico and in silico information sources.

2. General information

2.1 Identifier:

Sensitiser potency prediction based on Key event 1 + 2.

2.2 Date:

29 April2016

2.3 Author(s) and contact details:

Andreas Natsch, Bioscience, Givaudan Schweiz AG, Ueberlandstrasse 138, CH-8600 Duebendorf, Switzerland, andreas.natsch@givaudan.com, Tel.: ++41 44 824 21 05, Fax: ++41 44 824 29 26

2.4 Template update(s): Version 1.

2.5 Reference to main scientific papers [1-4]:

Emter R, Ellis G, and Natsch A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitiz-ers in vitro. Toxicol. Appl. Pharmacol, 245: p. 281-290.

Natsch A. and 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(2): p. 464-478.

Roberts D.W. and Natsch A. (2009). High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of michael acceptor electrophiles. Chem Res Toxicol, 22(3): p. 592-603.

Key paper:

Natsch A, Emter R, Gfeller H., Haupt T. and Ellis G. (2015). Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: Global vs. domain-based assessment. Toxicological Science, 143:319-332.

2.6 Proprietary aspects:

TIMES is a commercial software from LMC Bourgas. 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 Corp., but a licence for use in sensitisation assessment is included in the MTA of KeratinoSens®.

3. Endpoint addressed

3.1 Endpoint:

In its main embodiment, the current DIP predicts a '*most likely LLNA EC3 value*' based on readout from assays addressing the AOP Key events 1 + 2. It is thus addressing the endpoint addressed by OECD TG 429 on the local lymph node assay in mice.

In a second embodiment, 'likely human DSA_{05}' values in HRIPT/HMT are being predicted. No OECD guideline for human data is available but human method is described by McNamee et al., 2008**.**

3.2 Species:

Mouse (primary target), Human (secondary target)

3.3 Additional information about the endpoint:

The DIP uses data from in vitro and in chemico assays addressing Key event 1 (initial event, Covalent binding at cysteine and/or lysine in proteins in the skin) and Key Event 2 (Keratinocyte activation, activation of the Nrf2-pathway in keratinocytes) as defined in the OECD AOP (OECD, 2012).

4. Definition of the purpose and regulatory relevance

This defined approach consists of two tiers:

i) First hazard prediction is made based on peptide adduct formation and binary read-out from KeratinoSens®

ii) Secondly, prediction of potency of a chemical based on Key events 1 +2 is made as a discrete value, which can be used as point of departure for risk assessment, in particular to

a) eventually (at least when combined with other evidence) to be used as a starting point in quantitative risk assessment (QRA), applying adequate safety factors

b) The resulting estimated EC 3 values may also be used for classification (e.g. into GHS classes 1A and 1B)

For regulatory purposes, outcome from tier 1 may allow the classification of substances as sensitisers / non-sensitisers, such as in the Globally Harmonized System of Classification, Labelling and Packaging of Chemicals (GHS; UN GHS rev 1, 2005) and/or for REACH without the need to use animal-based test methods. Outcome from tier 2 may be used for GHS subclassification into subcategory 1A and 1B.

5. Rationale underlying the construction of the defined approach

The amount of modified protein in the skin appears as a key driver for the strength of the immunological reaction. The amount of modified protein is strongly influenced by the reaction rate of a chemical with the nucleophilic residues in proteins: The higher the reaction rate the less chemical and time is needed to reach a certain threshold amount of modified proteins. Therefore the reaction rate in peptide assays partly explains potency of test chemicals (Roberts et al., 2009; Natsch et al., 2011) and is used here as key input (MIE, Key event 1). Reaction rates of sensitisers of different potency vary over several orders of magnitude. A kinetic reaction rate including multiple test concentrations and sampling times offers a much larger dynamic range as compared to the guideline DPRA with only one concentration and one time point. Therefore, for potency prediction, the kinetic readings offer more information as compared to DPRA depletion values. Therefore this assay and readout was preferred over the guideline DPRA.

In addition, electrophilicity of chemicals will also influence the magnitude of activation of the Nrf2-pathway in keratinocytes. Finally, cytotoxicity of chemicals is known to partly correlate to sensitisation potency (dos Santos et al., 2011), due to (i) correlation to reactivity (high cytotoxicity of reactive chemicals) and (ii) danger signal formation, with cytotoxicity leading to local trauma triggering dendritic cell activation (Both these latter endpoints are measured in the KeratinoSens® assay covering Key event 2). h-CLAT offers complementary information on this endpoint. Since we found, by statistical analysis, that the two information sources provide partly redundant data and since the KeratinoSens® database available at the time is bigger, focus of this case study is on KeratinoSens® data. However, the same approach can be simply broadened by adding h-CLAT data and this was shown in the underlying publication (Natsch et al., 2015).

Finally, the physiochemical parameters vapour pressure and cLogP were included where statistically relevant to account for the (minor) contribution of epidermal disposition.

Thus, the readouts (i) reaction rate with peptide, (ii) ARE activation and (iii) cytotoxicity all contribute to potency, although these endpoints are also partly redundant as they are all influenced by intrinsic reactivity of chemicals. These early and potentially rate-determining effects will partly explain potency and can be measured in vitro. Added information from h-CLAT (i.e. dendritic cell activation) could be used, but it was found that it only slightly improves predictivity (Natsch et al., 2015) since these information sources are correlated.

Physicochemical parameters can be used to additionally refine prediction. Most important is volatility, as highly volatile chemicals rapidly evaporate from mouse ears after application and thus their sensitisation potential in the animal test is reduced. Hydrophobicity may affect disposition in the skin, although in general the effect on sensitiser potency is marginal with the exception of some chemical domains.

All this information relating to potency is combined in the DIP described here to arrive at a 'most likely' potency value (expressed as dose in LLNA or human test) based on information from Key event 1 +2.

All data are logarithmically transformed, and the Log-transformed values on their own have a significant correlation to potency. A multivariate regression model is used for prediction. This model is based on a number of assumptions, assuming a continuous linear scale for both the dependent and the independent variable. Multivariate regression was selected based on the underlying assumption that the measured factors work in an at least partially additive fashion. The advantage of the model is prediction of a discrete LLNA EC3 value, the disadvantage is that constant coefficients apply over the dynamic range. The DIP might be improved with more sophisticated statistical modelling using the same input variables to also predict a continuous scale of the target variable (LLLNA EC3 or human DSA_{05}).

6. Description of the individual information sources used

A) LC-MS and fluorescence-based kinetic peptide reactivity assay (Cor1C420-assay)

Chemical/biological mechanism addressed by the information source:

MIE / Key event 1, Covalent modification of proteins by skin sensitisers

Biological relevance of the test system used in the information source:

The actual proteins most often modified by skin sensitisers *in vivo* are unknown. The test peptide in the Cor1C420-assay is based on an experimentally identified hotspot in the human proteome (Cysteine C420 in the human protein Coronin1, found to be particular reactive (Dennehy et al., 2006). With two Lysine residues in proximity to a cysteine residue it has a particular high nucleophilicity allowing testing at relatively low concentrations and it picks up cysteine and lysine-reactive chemicals.

Readout of the information source:

Several readouts are obtained:

lxii) Presence and putative molecular weight of covalent peptide modifications

lxiii) Extend of chemical-catalysed peptide oxidation/dimerization

lxiv) Peptide depletion at 24h (as in DPRA) and at earlier time points for reactive chemicals

lxv) Maximal rate constant of the reaction (based on depletion) between chemical and peptide

Adduct formation is used for hazard assessment and domain attribution in the first step of the DIP and for mechanistic understanding, while rate constant is used for quantitative predictions. The Cor1C420 assay thus gives added information as compared to the validated DPRA: (i) kinetic rate constant with higher dynamic range for potency prediction, (ii) information on adduct formation (the actual MIE) and discrimination of adduct formation from oxidation. For these reasons it is used in this DIP.

B) TIMES SS Software

Chemical/biological mechanism addressed by the information source:

MIE / Key event 1, covalent modification of proteins by skin sensitisers.

Biological relevance of the test system used in the information source:

Reactivity is key for the MIE, thus intrinsic or metabolically triggered reactivity has a key biological relevance.

Readout of the information source:

Several readouts are obtained:

i) Key mechanistic alert for direct reactivity

ii) Key theoretical metabolites and their mechanistic alerts for direct reactivity

iii) Relevant calculated physicochemical parameters (cLogP and vapour pressure used here)

iv) Prediction of 3 classes (Non-sensitiser, weak or moderate/strong)

The former two readouts are used here for domain attribution in the first step of the DIP. Vapour pressure is used in the potency models.

The in silico prediction (iv) of the three classes is not directly used here, but it can be further combined in a weight of evidence approach to challenge or further support the potency predicted by this DIP based on experimental data.

C) KeratinoSens® ARE-Nrf2 Luciferase Test Method (Test Guideline 442D)

Chemical/biological mechanism addressed by the information source:

Key event 2, Gene expression of Nrf2-pathway in keratinocytes.

Biological relevance of the test system used in the information source:

Based on recent data (El Ali et al., 2013; van der Veen et al., 2013), the Nrf2 pathway is a key pathway of defence (PoD) (Hartung et al., 2012) triggered by sensitisers *in vivo*. To our current knowledge it is the key common molecular pathway which triggers gene expression in response to electrophilic chemicals at sub-toxic concentrations (Dinkova-Kostova et al., 2005; Neves et al., 2011).

Readout of the information source:

i) $EC1.5_{KS}$ and $EC3_{KS}$ values are giving the dose-response of ARE-regulated luciferase induction (concentration for 1.5-and 3-fold induction over solvent control)

ii) $IC50_{KS}$ gives the cytotoxicity (concentration for 50% reduction of viability)

iii) Positive/negative rating according the prediction model.

7. Data interpretation procedure applied

1) In a first step, a hazard prediction model is applied. Based on the two tests used here, a chemical is rated positive if it is either positive in the KeratinoSens® assay or if it does form a covalent adduct with the test peptide. This yields a qualitative yes/no rating.

2) For a quantitative potency assessment, chemicals rated positive are attributed to mechanistic domains based on structural alerts from in silico modelling (TIMES, optional OECD toolbox) and experimental data (LC-MS based peptide reactivity) on adduct formation according to Scheme in Figure VII.1. In principle, chemicals reacting with target proteins / peptides with the same reaction mechanism are attributed to the same domain. Domain attribution can be made based on chemical structural features in the test chemical (e.g. aldehyde functionality, α,β-unsaturated ketone) or based on directly observed and characterised peptide adducts. Details on how chemicals were attributed to domains can be found inNatsch et al. 2015.

3) $EC1.5_{KS}$, $EC3_{KS}$ and $IC50_{KS}$ from KeratinoSens®; vapour pressure from TIMES SS, cLogP and rate constant for peptide depletion from the Cor1C420-assay are then used for quantitative predictions of potency (pEC3):

   a) Chemicals which can be attributed to domains with sufficient comparative animal data and an established domain-based regression equation are predicted with the domain-based model. Currently models are available for (i) Michael acceptors, (ii) chemicals reacting by addition elimination, (iii) epoxides, (iv) quinone methides and (v) aldehydes.

   b) Chemicals which cannot be attributed to such domains are predicted with a global regression equation.

   c) A single regression equation is available to predict human potency – however it is based on a lower training set as less data are available.

This general procedure is depicted in Figure VII.1.

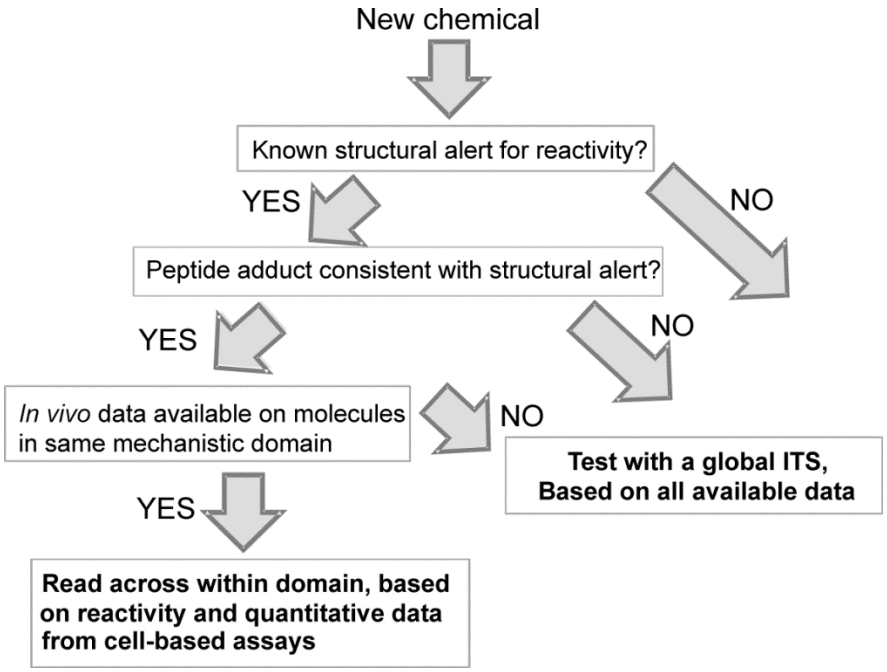


Figure VII.1: General Scheme for combining domain-based and global models for potency prediction within the DIP.

Process applied to derive the final prediction/assessment:

Data from the *in chemico* and *in vitro* assays are normalized and log transformed as described in detail (Natsch et al., 2015).

Case A - Chemical can be attributed to domain: Prediction of LLNA EC3

If chemicals are attributed to one of the following domains, a domain based regression equation is used:

i) Chemicals predicted to react with proteins by and **addition-elimination reaction** are predicted by Equation 1:

(Eq1)

$$pEC3 = 0.557 + 0.518 \times \text{Log } K_{norm} + 0.304 \times \text{Log } IC50_{norm} - 0.666 \times \text{Log } VP_{norm}$$

$S = 0.56$ $R^2 = 88.2\%$ R^2 (adj) = 85.9%, $p < 0.0005$

ii) Chemicals predicted to react with proteins by a direct **Michael addition** are predicted by Equation 2:

(Eq2)

$$pEC3 = 0.540 + 0.230 \times \text{Log } K_{norm} + 0.292 \times \text{Log } EC3_{norm} - 0.290 \times \text{Log } VP_{norm}$$

$S = 0.39$ $R^2 = 61.3\%$ R^2 (adj) = 58.4%, $p < 0.0005$

iii) Chemicals predicted to react with proteins by **epoxide-opening** are predicted by Equation 3:

(Eq3)

$$pEC3 = 1.57 + 0.475 \times \text{Log } IC50_{norm} - 0.662 \times \text{Log } VP_{norm}$$

$S = 0.31$ $R^2 = 79.5\%$ R^2 (adj) = 76.3%, $p < 0.0005$

iv) Aldehydes predicted to react with proteins by **Schiff Base formation** are predicted by Equation 4:

(Eq4)

$$pEC3 = 0.545 + 0.298 \times \text{Log } EC3_{norm} + 0.535 \times \text{Log } IC50_{norm} - 0.221 \text{ cLogP}$$

$S = 0.54$ $R^2 = 49.3\%$ R^2 (adj) = 43.0%, $p = 0.001$

v) Chemicals predicted to be **quinone/quinone methide percursors** according to the TIMES SS model are predicted by Equation 5:

(Eq5)

$pEC3 = 1.80 - 0.417\ cLogP + 0.832 \times Log\ IC50_{norm}$

$S = 0.78 \quad R^2 = 52.3\% \quad R^2\ (adj) = 48.9\%,\ p < 0.0005$

Case B - Chemical cannot be attributed to a domain: Prediction of LLNA EC3

If chemicals are not attributed to a domain, their most likely pEC3 value is then calculated from these normalized data based on regression equation 6:

(Eq6)

$pEC3 = 0.04 + 0.38 \times Log\ K_{norm} + 0.25 \times Log\ EC1.5_{norm} + 0.25 \times Log\ IC50_{norm} - 0.19 \times Log\ VP_{norm}$

$S = 0.74; \quad R^2 = 63.0\%; \quad R^2\ (adj) = 62.3\%,\ p < 0.0005$

Case C - Estimation of likely human sensitisation level

In order to arrive at an estimated human DSA05 value, Equation 7 is used:

(Eq7)

$pEC_{human} = -1.74 + 0.074\ Log\ K_{norm} + 0.35\ Log\ EC3_{norm} + 0.37\ Log\ IC50_{norm}$

$S = 0.61 \quad R^2 = 52.9\%; \quad R^2\ (adj) = 49.5\%;\ p < 0.0005$

8. Chemicals used to develop and test the DIP

8.1 Availability of training and test sets:

The full training set for the global equation (Eq 6, n = 244) and the dataset including additional data used to derive Eq1 – Eq5 (n = 312) is attached as supporting information to Natsch et al., 2015. Database including all chemical- and toxicological information is available in this document and can be downloaded from http://toxsci.oxfordjournals.org/content/143/2/319/suppl/DC1.

This published database contains all the LLNA and human potency data and *in vitro*, *in chemico* and *in silico* data.

8.2 Selection of the training set and test set used to assess the DIP:

The training set for LLNA prediction is composed of two datasets:

a) 144 of the 145 chemicals used in a recent integrated testing strategy (Jaworska et al., 2013). This set of chemicals was selected by toxicological experts from P&G, mainly based on curated LLNA data compilations published by the same team (Gerberick et al., 2005; Kern et al., 2010) This set was used in the validation of the DPRA.

b) The set of chemicals used in the ECVAM validation of KeratinoSens® assay. This latter dataset consists of (i) a set of 114 chemicals selected mainly from the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) validation of the LLNA (Haneke et al., 2001) and the ICCVAM database (ICCVAM et al., 2008) with congruent data in LLNA and guinea pig and/or humans; this selection and the selection criteria are described in the KeratinoSens® submission to ECVAM [see attachment 12b in (ECVAM, 2014) for this list]. (ii) studies on specific groups of chemicals which had been synthesized to evaluate structure-activity relationships in the LLNA (Ball et al., 2011; Delaine et al., 2011). (iii) a set of additional chemicals for which only negative sensitisation evidence from the LLNA was available, which was used to balance the number of positives and negative chemicals in the validation data set, as the number of available non-sensitisers with congruent human/guinea pig and LLNA data was too small. Addition of such additional chemicals to the database was requested by ECVAM (ECVAM, 2014). These chemicals were mainly selected from (Gerberick et al., 2005; Kern et al., 2010).

These two sets are partly overlapping, but (b) contains 100 chemicals in addition to the 144 chemicals from the dataset of 145 (Delaine et al., 2011). This published set (n=244) of chemicals is considered the most robust database as it is based on most high quality *in vivo* data and it was used for a global assessment of potency predictions in the LLNA. The global model was also developed separately for set a) and set b) to check how strongly Equation 6 it is dependent on the dataset (see Natsch et al., 2015), and very similar results with the two independent sets were achieved, indicating that using either subset as a training set would yield similar results. For the local models, chemical potency predictions were made with a leave-one out analysis to avoid overfit in absence of a test set.

A further set of chemicals (n= 68) with less solid in *vivo* evidence (only one LLNA value, or LLNA and human data are contradictory) was used to enrich the datasets for individual chemical domains.

The dataset used to compare with human data is largely congruent with the dataset used by ICCVAM to validate potency predictions by using LLNA (ICCVAM, 2011) and it is a subset of the full database referred to above.

8.3 Supporting information on the training and test sets:

Since all the *in vivo* and *in vitro* data and the data on chemical specifications including chemical structure are publicly available at http://toxsci.oxfordjournals.org/content/143/2/319/suppl/DC1, no additional data need to be provided here.

8.4 Other information on the training and test sets

Since all data are available, no further information needed.

9. Limitations in the application of the defined approach

a) **Technical constraints for testing**: Chemicals with a cLogP < 5 in general can be tested with KeratinoSens® and the reactivity assay, even if in some cases at the top concentration

supersaturated solutions are formed. Higher cLog-P materials are difficult to test. To investigate adduct formation in LC-MS assay and in order to attribute chemicals to domains, a defined chemical structure is a prerequisite and should be available to make this DIP actionable. Chemicals should be soluble in acetronitrile or water or in a mixture of both for the reactivity assay. For KeratinoSens®, chemicals should be soluble in DMSO or directly in cell culture medium containing 1% serum and 1% DSMO. Chemicals soluble only in ethanol can also be tested in both systems by applying appropriate controls.

b) **Limitations based on applicability domain/ limited predictive capacity**: The test battery has a limited sensitivity for some prohaptens typically needing P450-mediated activation such as some phenolic compounds. Assays to specifically test such compounds have been published (Natsch and Haupt, 2013). Phenolic compounds form plants may not be reliably assessed with this DIP as they may be false-positive. Specifically amine-reactive compounds are detected as sensitisers, but underpredicted for potency (Natsch et al., 2015), as the DIP is slightly more sensitive for chemicals reacting with cysteine as MIE. Limited predictivity was also noted for silicone materials.

10. Predictive capacity of the defined approach

a) Hazard classification

Hazard classification is not the main purpose of this defined approach, yet it is the first step based on the decision criteria laid out in Section 7. Table VII.1 gives the Cooper statistics for this analysis:

Table VII.1. Cooper statistics on the set of 244 chemicals. Results for KeratinoSens®, LC-MS (adduct forming = positive) and the combination of both measures.

	KeratinoSens®	**Peptide adduct formation**	**KeratinoSens® positive and/or adduct forming (=" *in vitro* hazard positive")**
Sensitivity	82.0%	63.2%	88.7%
Specificity	77.5%	91.9%	75.7%
Accuracy	79.9%	76.2%	82.8%

Note: For these Cooper statistics the LLNA prediction is taken, unless in cases where there is very clear human and/or guinea data overruling LLNA (e.g. surfactants).

b) Potency assessment - continuous

The overall R^2 for equation 6 to predict the (logarithmic) pEC3 is 62%. Table VII.2 lists the results for chemicals in different potency classes. Accuracy of prediction is described as fold-misprediction on either side of the measured *in vivo* LLNA value. Chemicals in the weak and moderate class are predicted with a mean error of 3 – 3.4-fold. There is a tendency to underpredict the actual LLNA value of strong and especially extreme sensitisers.

Table VII.2. Summary of predictions with Equation 6 in the different LLNA potency classes.

Classification[1)]	**N**	**Geomean LLNA[5)]**	**Geomean Predicted LLNA (Eq. 1) [6)]**	**Geomean fold-misprediction[7)]**
Correct negatives [1)]	83	n.a.	*93.9*	n.a.
False-negatives [2)]	15	7.29	*41.7*	n.a.
False-positives [3)]	27	n.a.	27.9	n.a.
Correct positives [4)]-Weak	32	20.39	9.6	**3.0**
Correct positives-Moderate	46	3.25	3.34	**3.4**
Correct positives-Strong	26	0.32	1.20	**5.8**
Correct positives-Extreme	14	0.02	0.33	**17.81**

[1)] Chemicals not forming a peptide adduct AND negative in KeratinoSens®

[2)] Chemicals not forming a peptide adduct AND negative in KeratinoSens®, but with positive LLNA

[3)] Chemicals forming a peptide adduct AND/OR positive in KeratinoSens®, but with negative LLNA

[4)] Chemicals forming a peptide adduct AND/OR positive in KeratinoSens® AND positive in LLNA

[5)] Geometric mean of the LLNA EC3 values of the chemicals evaluated in the particular class

[6)] Geometric mean of the **predicted** LLNA EC3 values (Equation 6) of the chemicals evaluated in the particular class. Comparison to column 3 indicates whether there is a general over or underprediction.

[7)] The LLNA EC3 and the EC3 predicted by Equation 6 are compared by calculating the ratio between the two, and fold-misprediction on either side was calculated for each molecule in class. Given is the geometric mean of this misprediction. E.g. for the weak sensitisers the predicted LLNA EC3 value is typically 3-fold higher or lower as compared to measured value.

For the domain models, the accuracy of prediction is in general higher, as shown in Table VII.3.

Table VII.3. Different mechanistic domains and key results of correlation analysis vs. LLNA EC3.

Domain [1)]	**N**	**R^2-adj. of best model (*p*-value)**	**Fold-misprediction domain model** [2)]	**Fold- misprediction global model**
Michael acceptors	44	58.4% [3)] (< 0.0005)	2.26	3.22
Addition-elimination (S_N2/S_NAr)	19	85.9% (< 0.0005)	2.60	3.43
Epoxides	16	81.2% (< 0.0005)	1.97	2.88
Aldehydes	28	43% (0.001)	3.16	3.26
pre-quinone-domain	32	48.2% (< 0.0005)	4.54	6.45

[1)] Domains were formed based on TIMES SS prediction and/or observed adduct formation.

[2)] Best subset analysis / regression equation is calculated based on n-1 chemicals, and then used to predict potency of the chemical left-out. Each chemical is thus individually treated as the test set with the remaining chemicals acting as the training set.The LLNA EC3 and the predicted EC3 are compared by calculating the ratio between the two, and fold-misprediction on either side was calculated for each molecule in class. Given is the geometric mean of this misprediction. E.g. for the Michael acceptors the predicted LLNA EC3 value is typically 2.2-fold higher or lower as compared to measured value.

[3)] Pentaerythritol triacrylate has a strong weight – without this chemical, regression coefficient is 69%. This chemical is a strong human sensitiser and strongly positive *in vitro*, but negative in the LLNA.

From this analysis it is clear that for chemicals within the established data-rich mechanistic domains a more reliable prediction is possible as compared to chemicals with no attribution to data-rich domains. Thus for the chemicals out of these domains either more data for a weight-of-evidence is needed or a more conservative safety assessment factor is needed in QRA.

c) Potency assessment – GHS classes

The main goal of this analysis is prediction of a continuous likely LLNA or human potency value as point of departure for risk assessment. A secondary goal is potency class prediction according to GHS. Chemicals were thus classified into the GHS classifications according to the predicted LLNA EC3 or human DSA thresholds. This analysis is shown in Tables VII.4 and VII.5.

Table VII.4. Prediction of human GHS classes by *in vitro* data and the LLNA.

In vivo class (threshold DSA_{05} of 500 µg/cm^2)		Prediction by this DIP [2)]		
	n[1)]	Non-sensitiser	GHS 1B	GHS 1A
Human class 1B	44	16	**24**	4
Human class 1A	27	3	5	**19**
		Prediction by LLNA[3)]		
	n	Non-sensitiser	GHS 1B	GHS 1A
Human class 1B	44	9	**32**	3
Human class 1A	27	2	11	**14**

[1)] Shown are the chemicals in the ICCVAM validation dataset with positive human LOEL values and tested in KeratinoSens® (n=71)

[2)] Predicted by equation 7

[3)] Predicted by equation S5 in [4]

Table VII.5. Prediction of the dataset of 244 molecules for LLNA GHS classes based on the Equation 6 (global model) or by the combined analysis.

		DIP prediction (global model)			**DIP prediction (combined global and domain models)**		
LLNA prediction	n	**NS**	**GHS 1B**	**GHS 1A**	**NS**	**GHS 1B**	**GHS 1A**
Non-sensitiser	111	**84**	26	1	**84**	26	1
LLNA GHS 1B	79	13	**53**	13	13	**61**	5
LLNA GHS 1A	54	2	15	**37**	2	14	**38**
	244		Three-way Accuracy = 71%			Three-way Accuracy = 75%	

11. Consideration of uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

1. The DIP's structure

- Key event 3 is not included. However an analysis on the effect of adding Key event 3 was performed, indicating that in vitro data from Key event 2 and 3 give mostly redundant information (Natsch et al., 2015) and are not both needed.

- Key event 4 is not included due to lack of available tests. Assumption is that a given amount of modified protein and a given amount of induction of early cellular events will lead to a predictable lymphocyte proliferation, i.e. that lymphocyte proliferation depends on the ability of the chemical to trigger these earlier events. This assumption is not sufficiently tested currently.

- The confidence is high for chemicals reacting by addition – elimination, epoxides and Michael acceptors, where analysis indicates that coverage is clearly sufficient. Confidence is lower for aldehydes, where coverage is less good due to inadequate coverage of MIE by current reactivity assays (Natsch et al., 2012) (i.e. Quantitative reactivity of amine reactive chemicals not fully represented, no kinetic amine reactivity test) and limited confidence for quinone methide precursors due to limited metabolic capacity of test systems.

2. The information sources used within the defined approach

Reproducibility of peptide reactivity, cytotoxicity and Nrf2-induction measurements are very high. Not only within given test and within and between different laboratories, but recent analysis also indicates high correlation between different tests with the same readout. Reproducibility of the information source thus has minor impact as compared to incomplete coverage and incomplete in vitro-in vivo extrapolation.

3. Benchmark data used

- The variability of the *in vivo* data may limit accuracy of prediction, it is originating from two sources: test variability and variability between laboratories (the reference databases come from multiple centres and were built up over two decades for the LLNA and 3-4 decades for the human data). Repeated LLNA values in the ICCVAM database vary 1.7 – 2-fold on either side of the mean (ICCVAM, 2008) and therefore a 2-fold misprediction of an EC3 may in many cases still fall within the variability of the true *in vivo* outcome. The uncertainty is even higher for the human data obtained with different test protocols. Correlation of the LLNA with the human LOEL is far from perfect (Natsch et al., 2015), which can partly be attributed to these limitations in the human dataset.

11.2 Impact of uncertainty on the DIP's prediction

Variability of input data:

a) Due to very high reproducibility of input data from in chemico and in vitro assays, the test data uncertainty will have limited effect on final uncertainty.

b) LLNA data do have significant variation, however, due to the fact that the model is trained on a large dataset, this variability should not significantly affect the overall regression parameters obtained. (However, when assessing DIP accuracy of single chemical vs. known single LLNA value, somewhat limited predictivity has to be expected taking the 1.7 – 2 fold LLNA variability into account.)

Following from the two above points, the overall uncertainty when predicting a new chemical is mainly dominated by:

a) **limited AOP coverage / DIP – structure** (Whereby lack of addressing Key event 3 has relatively little impact, but analysis of a set including additionally h-CLAT data indicates that even when addressing KE 1,2 and 3 – lack of AOP coverage still limits predictiviy.)

b) **limited training data in domains with few LLNA and human data**.

To estimate uncertainty, for each new prediction, the domain attribution should be carefully checked. For chemicals predicted with the global model, the prediction accuracy of related chemicals with known in vitro and in vivo data should be checked. Confidence is high if similar chemicals are included in training set and predicted with less than 3 – 5 fold misprediction. Effect of limited AOP coverage is not known for chemicals completely outside of training set domains.

12. References

Ball N, et al. (2011). Evaluating the sensitization potential of surfactants: Integrating data from the local lymph node assay, guinea pig maximization test, and in vitro methods in a weight-of-evidence approach. Regul. Toxicol. Pharmacol, 60: p. 389-400.

Delaine T, et al. (2011). Structure-Activity Relationship between the in Vivo Skin Sensitizing Potency of Analogues of Phenyl Glycidyl Ether and the Induction of Nrf2-Dependent Luciferase Activity in the KeratinoSens in Vitro Assay. Chem. Res. Toxicol, 24(8): p. 1312-8.

Dennehy M.K., et al. (2006). Cytosolic and nuclear protein targets of thiol-reactive electrophiles. Chem Res Toxicol, 19(1): p. 20-9.

Dinkova-Kostova A.T, Holtzclaw W.D and Kensler T.W, (2005). The role of Keap1 in cellular protective responses. Chemical Research in Toxicology, 18(12): p. 1779-1791.

Dos Santos G.G et al. (2011). A potential in vitro epidermal equivalent assay to determine sensitizer potency. Toxicol In Vitro, 25(1): p. 347-57.

ECVAM, (2014). EURL ECVAM Recommendation on the KeratinoSens® assay for skin sensitisation testing: Attachments. http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations/file-kerati/Attachments%20merged.pdf

El Ali Z. et al. (2013). Allergic skin inflammation induced by chemical sensitizers is controlled by the transcription factor nrf2. Toxicol Sci, 134(1): p. 39-48.

Emter R, Ellis G and Natsch A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. Toxicol. Appl. Pharmacol, 245: p. 281-290.

Gerberick G.F. et al. (2005). Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. Dermatitis, 16(4): p. 157-202.

Haneke K.E. et al. (2001). ICCVAM evaluation of the murine local lymph node assay. Data analyses completed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. Regul Toxicol Pharmacol, 34(3): p. 274-86.

Hartung T. et al. (2012). Food for thought ... Systems toxicology. ALTEX, 29(2): p. 119-28.

ICCVAM, (2008). ICCVAM database, downloaded in December 2008 from http://iccvam.niehs.nih.gov/methods/immunotox/LLNA-pot/appx/LLNApotencyAppB18Jan08FD.xls

ICCVAM, (2011). ICCVAM Test Method Evaluation Report: Usefulness and Limitations of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans. NIH Publication Number 11-7709. Downloaded 23.1.2014 from http://ntp.niehs.nih.gov/iccvam/docs/immunotox_docs/LLNA-pot/TMER.pdf#search=Test%20Method%20Evaluation%20Report:%20Usefulness%20and%20Limitations%20of%20the%20Murine%20Local%20Lymph%20Node%20Assay

Jaworska J. et al. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. Journal of Applied Toxicology, 33(11): p. 1353-1364.

Kern P.S. et al. (2010). Local lymph node data for the evaluation of skin sensitization alternatives: A second compilation. Dermatitis, 21(1): p. 8-32.

McNamee P.M. et al. (2008). A review of critical factors in the conduct and interpretation of the human repeat insult patch test. Regul Toxicol Pharmacol, 52(1): p. 24-34.

Natsch A. and Haupt T. (2013). Utility of Rat Liver S9 Fractions to Study Skin-Sensitizing Prohaptens in a Modified KeratinoSens Assay. Toxicol Sci, 135(2): p. 356-68.

Natsch, A. and H. Gfeller, (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(2): p. 464-478.

Natsch A., Haupt T, and Laue H. (2011). Relating skin sensitizing potency to chemical reactivity: reactive Michael acceptors inhibit NF-kappaB signaling and are less sensitizing than S(N)Ar- and S(N)2-reactive chemicals. Chem Res Toxicol, 24(11): p. 2018-27.

Natsch A. et al. (2012). Chemical reactivity and skin sensitization potential for benzaldehydes: can schiff base formation explain everything? Chem Res Toxicol, 25(10): p. 2203-15.

Natsch A.E, Gfeller R, Haupt H, Ellis T. G. (2015). Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: Global vs. domain-based assessment. Toxicol Sci, 143(2): p. 319-32.

Neves B.M et al. (2011). Signal transduction profile of chemical sensitisers in dendritic cells: an endpoint to be included in a cell-based in vitro alternative approach to hazard identification? Toxicol. Appl. Pharmacol, 250(2): p. 87-95.

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

Roberts D.W and Natsch A. (2009). High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of michael acceptor electrophiles. Chem Res Toxicol, 22(3): p. 592-603.

Van der Veen J.W. et al. (2013). Induction of skin sensitization is augmented in Nrf2-deficient mice. Arch Toxicol, 87(4): p. 763-6.

13. Supporting information

All the details on the data and the training set as well as the calculation of the statistical models are given in the key publication (Natsch et al., 2015) which forms an integral part of this case study (http://toxsci.oxfordjournals.org/content/143/2/319/suppl/DC1).

14. Abbreviations and definitions

ARE antioxidant response element;

cLogP logarithmic calculated Octanol-water partition coefficient

DPRA Direct Peptide reactivity Assay

DIP Data interpretation procedure

DSA_{05} Induction dose per skin area, in µg/cm2, in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population;

EC3 Estimated concentration in % of a substance expected to produce a stimulation index of 3, the threshold value for a substance to be considered a sensitiser in the LLNA

$EC1.5_{KS}/EC3_{KS}$ Dose in µM for induction of a 1.5/2/3-fold induction of luciferase activity in KeratinoSens®

HMT Human maximization test

HRIPT Human repeat-insult patch test

$IC50_{KS}$ Dose in µM for 50% reduction in cell viability in KeratinoSens®;

LC-MS	liquid chromatography coupled to mass spectrometry
LLNA	local lymph node assay
Nrf2	nuclear factor-erythroid 2-related factor 2
QRA	quantitative risk assessment
VP	vapour pressure.

CASE STUDY VIII

Sensitiser potency prediction based on Key event 1+2+3 : The artificial neural network model for predicting LLNA EC3

1. Summary

The defined approach presented in this document describes an integrated testing strategy (ITS) for prediction of the skin sensitization potential and potency of a substance. The line of evidence predicted LLNA (primary target). The combination of test methods used covers the first three key events (KEs) of the adverse outcome pathway (AOP) leading to skin sensitization as formally described by the OECD: KE 1: protein binding (e.g. via the SH test or direct peptide reactivity assay (DPRA); OECD TG 442 C); KE 2: keratinocyte activation (e.g. via the ARE assay or KeratinoSensTM ; OECD TG 442D); and KE3: dendritic cell activation (e.g. via the human cell line activation test (h-CLAT); OECD TG 442E). Log P of the target molecule is used as an indicator of skin penetration. This ITS described here is based on 62 chemicals classified a Sensitisers / Non-Sensitisers (S/NS) in the LLNA. Artificial neural network (ANN) analysis, a nonlinear statistical data-modelling tool, was adopted in this model. The artificial neural network model can be used not only for hazard identification but also for potency estimations. Regarding EC3 predictive capacity, the R value (coefficient of correlation) was 0.838, and RMS error was 0.628. Predicted EC3 value which is calculated by this model can be used as a point of departure in quantitative risk assessment (QRA), applying adequate safety factors. For three classification (extreme and strong, moderate and weak, and non-sensitiser), accuracy is 79%. Therefore this approach may be used to subcategorise skin sensitisers in cat 1A and 1B for GHS/CLP.

2. General information

2.1 Identifier: Sensitiser potency prediction

2.2 Date:

13 April 2016

2.3 Author(s) and contact details:

Takao Ashikaga, Ph.D.

2-12-1 Hayabuchi, Tsuzuki-ku, Yokohama-shi 224-8558 Japan

takao.ashikaga@to.shiseido.co.jp

2.4 TA SRF update(s): Version 1

2.5 Reference to main scientific papers:

Tsujita-inoue K et al. (2014). Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays**,** Toxicology in Vitro, 28: p. 626-639.

Hirota M et al. (2015). Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization, J. Applied Toxicology, Mar 30. doi: 10.1002/jat.3105.

2.6 Proprietary aspects:

h-CLAT is covered by intellectual property right in Japan. However, Shiseido offered free of charge its proprietary core technology associated with h-CLAT to all outside institutions for use as a skin sensitisation test.

3. Endpoint addressed

3.1 Endpoint:

The ANN model assesses the skin sensitiser potency of chemicals in the LLNA (EC3 value).

3.2 Species:

Mouse (primary target), Human (secondary target). Hazard characterization (potency estimation).

3.3. Additional information about the endpoint:

Skin sensitisation as an allergic reaction is the outcome of a number of complex interactions at molecular, cellular and tissue levels. There are the key biological mechanisms of the induction phase of the skin sensitisation, including skin penetration, covalent binding of the chemical to the skin protein (KE 1), activation of epidermal keratinocytes (KE 2), activation of Dendritic cells (KE 3), and proliferation of antigen specific T cells. The model use data from in vitro assay, in vitro assay and physicochemical property addressing skin penetration, KE 1, KE 2 and KE 3.

4. Definition of the purpose and regulatory relevance

The artificial neural network model can be used not only for hazard identification but also for potency predictions. Predicted EC3 value can be used as a point of departure in quantitative risk assessment (QRA), applying adequate safety factors. In other words, predicted EC3 value can be applied for prediction of a safe level of human exposure using a Quantitative Risk Assessment (QRA) approach (Api et al., 2008). This approach can be used to subcategorise skin sensitisers in cat 1A and 1B for GHS/CLP.

5. Rationale underlying the construction of the defined approach

Skin sensitisation is the result of a complex multifactorial sequence of events and has long been the focus of research. The molecular initiating event is defined as the covalent binding of the hapten to skin proteins. This step is evaluated using the SH test or Direct Peptide Reactivity Assay (DPRA; OECD TG 442C). Inflammatory and protective responses by the first cells coming into contact with the substance, the keratinocytes, are essential for downstream events to take place. Keratinocyte activation is evaluated via the Nrf2-ARE-based ARE assay or KeratinoSens^{TM} (OECD TG 442D). Dendritic cells (DCs) transport the hapten to the regional lymph nodes, present the hapten on the cell surface and, when activated (mature DCs), are able to present the antigen in the proper context (upregulated cell surface markers, e.g. CD86 and CD54) to activate naïve T-cells thereby triggering their proliferation. The potential of a substance to cause DC activation is assessed using the h-CLAT (OECD TG 442E). Log P is closely related skin penetration. Each in vitro test is corresponding protein binding, Keratinocyte activation and Dendritic cell activation. These indicator and tests are covering the AOP key events. Our ANN model can contribute to building a new QRA evaluation system by predicting EC3 without animal testing. Due to the diversity of architectures and adaptation algorithms, the ANN approach is used in the broad spectrum including toxicological applications [e.g., mutagenicity (Valkova et al., 2004)]. Because the mechanisms of skin sensitisation are too complex, based on immune system, it is widely recognized that a single in vitro test is insufficient to replace and that integration of results from various in vitro tests, as well as in silico methods, is needed for prediction of skin sensitisation potency. Therefore, the ANN approach would be to play important role in this field, where the commonly used approaches hardly work. It is widely accepted that ANN approach is effective for estimation of complex reaction consisting of multi steps such a toxicological process (Valkova et al., 2004).

6. Description of the individual information sources used within the approach

A) Log P: distribution coefficient (in silico)

Log P of the target molecule is addressing a skin penetration. Log P is the important parameter of skin penetration, and one of the factors which might influence the potency of skin sensitisation seriously (Guy et al., 1993). Another risk assessment model incorporating Log P as a measure of water solubility and skin sensitisation has already been reported (Jaworska et al., 2013). Therefore, we use Log P as one of parameters because it can affect skin penetration and solubility to culture medium. Log P is calculated using a non-validated but widely in use software “KOWWINN” which was downloaded from US EPA web site. Log P is calculated as a continuum number. Log P of high molecule and metal is not able to be calculated.

B) SH test: cell-surface thiol test (in vitro)

SH test detects changes of cell-surface thiols on hapten-treated THP-1 cells. This test is addressing a protein binding as well as DPRA, and the outcome of SH test is closely correlated with that of DPRA (Suzuki et al., 2009). Binding of chemicals to the skin protein is an essential step for sensitiser to obtain allergenicity (Suzuki et al., 2009). For SH test, the amount of cell surface free thiol is measured. Maximum amount of change (MAC) of cell surface thiol is used as an indicator of protein binding. SH test a non-validated test method implemented by a relatively small number of institutes. Insoluble chemicals are out of these applicable domains.

C) ARE assay: Antioxidant Response Element assay (in vitro)

ARE assay is addressing a Keratinocyte activation. ARE assay was developed to assess activation of the Keap/Nrf2/ARE regulatory pathway as well as KeratinoSens (Natsch and Emter, 2008). Keratinocyte activation is one of the key events and it includes inflammatory responses such as cytokines release in the skin (Natsch and Emter, 2008). AREc32 cells are used for the ARE assay. ARE c32 is made from MCF7 cell line, which is a breast cancer cell line, not keratinocyte. However, many studies proved that it is an useful test to screen for the skin sensitisation potential of novel chemicals as well as Keratinosens. For ARE assay, luciferase activity is measured. The average maximal induction of luciferase activity (Imax) is used as an indicator of Keratinocyte activation. ARE assay is a non-validated test method but widely in use. Insoluble chemicals are out of these applicable domains.

D) h-CLAT: human Cell Line Activation Test (in vitro)

h-CLAT is addressing a Dendritic cell activation. Dendritic cell has a major role, which is antigen presenting, in the skin sensitisation process. Without the activation of dendritic cells, sensitisation to allergen is not formed. Obtaining DCs from peripheral blood not only takes several days, but also requires complicated procedures. Furthermore, the phenotype is unstable and the effects of chemicals on the surface phenotype of DCs were dependent on the source of peripheral blood, i.e., they varied from donor to donor. Therefore, we use a human leukemia cell line called THP-1 as a replacement of dendritic cell. The cell line is well characterized and is known to as surrogate for dermal dendritic cell. For h-CLAT, the amount of CD86 and CD54 protein on the cell surface is measured (Ashikaga et al., 2010). Minimum induction threshold of cell surface markers (CD86 and CD54) and cytotoxicity [cytotoxicity could be a relatively crude measure of danger signal formation, and cytotoxic concentration are related to the concentration at which DCs are activated (Emter et al., 2010)] is used as an indicator of Dendritic cell activation (Nukada et al., 2012). h-CLAT is a validated but non-regulatory accepted test yet. Insoluble chemicals are out of these applicable domains.

E) LLNA (in vivo)

EC3 values are used as in vivo indicators. In case of non-sensitiser, maximum applied doses were used as well.

In general, descriptors that are correlated with each other are not used in ANN analysis, because combinations of independent descriptors are expected to improve the predictive performance of ANN models, as compared with combinations of correlated descriptors (e.g., cytotoxicity of each in vitro test).

The author would like to point out that information source can be exchangeable (e.g., DPRA and KeratinoSens can be used as replacement of SH test and ARE assay, respectively).

7. Data interpretation procedure (DIP) applied

These individual tests give us several parameters. Then these parameters are integrated by using artificial neural network. Therefore, it is possible to say that our model is a kind of integrated approach in order to derive an interim conclusion. Please note that we cannot see how much each information source contributed to the decision.

The artificial neural network consists of input layer, hidden layer and output layer. Log P and parameters resulting from the three in vitro tests are input. All calculations were performed using QwikNet Ver. 2. 23. In this mode, there are two hidden layers. Published LLNA thresholds are used as the output layer. Our DIP model is quantitative.

The conceptual diagramis presented in Figure VIII.1.

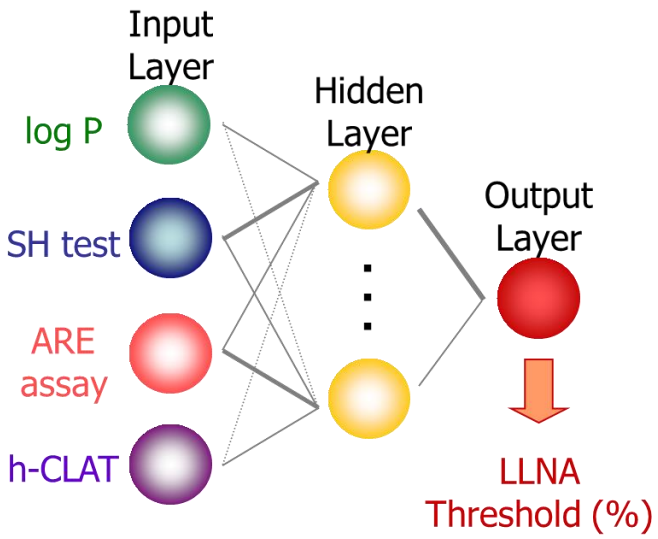


Figure VIII.1: Conceptual diagram of the artificial neural network consisting of input layer, hidden layer and output layer. The LLNA Threshold is EC3 value (%).

8. Chemicals used to develop and test approach

8.1 Availability of training and test set

A dataset of 62 chemicals was used as the defined approach training set. You can see the dataset in Kyoko Tsujita-inoue et al. 2014. Please see appendix 1 in Kyoko Tsujita-inoue et al. 2014, indicating endpoint values, chemical names and CAS numbers.

8.2 Selection of the training set and test set used to assess the approach

Regarding the data selection process, we carefully chose a variety of chemicals with various physicochemical properties and potency of sensitisation. All chemicals were selected on the premise that in vivo data for skin sensitisation was published in high quality journal. Instead of using a test set, a 10-fold cross-validation procedure was conducted. This method systematically removed data successively from the training set. A network model was then constructed on the basis of this reduced data set and was subsequently used to predict the removed data. This procedure was repeated for all data so that a complete set of predicted values was obtained. The goodness-of-fit was evaluated in terms of root-mean-square (RMS) error.

9. Limitations in the application of the defined approach

Chemicals that fall outside the applicability domains of each in vitro test adopted in this model cannot be applicable.

Technical limitations:

- Technical limitations exist, e.g. if substances or precipitates that are formed interfere with the detection system (bubbles formed by surfactants can interfere with flow cytometric detection in some cytometers, depletion of peptides not due to adduct formation, pigments could interfere with viability readouts)

- Physical state may preclude testing e.g. gases, highly lipophilic substances (cell culture). Substances with a high logP (e.g. exceeding 3.5 in the h-CLAT and 5.0 in the KeratinoSens assay) may pose problems due to the aqueous nature of the cell culture medium and solubility issues.

- Substances must be stable under test conditions e.g. the DPRA uses high alkaline conditions for lysine reactivity

- Complex mixtures, e.g. plant extracts or formulations are difficult to evaluate as molecular weights or molar equivalents are used in some tests

- Peptide depletion due to adduct formation cannot be differentiated from peptide depletion due to dimerization or oxidation of the peptide

Substance related limitations:

- Pre- and pro-haptens might not be reliably predicted due to lack of metabolic capacities in both the DPRA and h-CLAT.

- Substances that only react with lysine and not with cysteine can lead to false negative predictions as both the DPRA and KeratinoSens^{TM} use cysteine reactivity as a read-out.

In silico limitations also exist in this model.

In silico limitations:

- It is impossible to see how much each information source contributed to the decision.

- The number of the dataset is limited (62 chemicals). Increasing the size of the dataset might be effective to improve the predictive capacity.

10. Predictive capacity of the defined approach

Regarding the predictive capacity of our model, we can show not only the accuracy but also the linear correlation analysis.

The accuracy for the three classifications (extreme and strong, moderate and weak, and non-sensitiser) was 79%. In addition, over predictive performance was 9.7% and under predictive performance was 11.3%, respectively. This analysis is shown in Table VIII.1.

Table VIII.1. Prediction of the dataset of 62 chemicals for LLNA three categories

		LLNA classification		
		Extreme or strong (18)	Moderate or weak (30)	Non sensitiser (14)
DIP prediction	Extreme or strong	12	1	0
	Moderate or weak	6	28	5
	Negative	0	1	9

In this model, under-prediction of 8 chemicals was observed. In particular, Phthalic anhydride was under-predicted by 2 ranks (this strong sensitiser was estimated as weak). Phthalic anhydride has poor water solubility. This result suggested that water-insoluble chemical might cause under-prediction. Furthermore, phthalic anhydride is easily hydrolysed to phthalic acid. Therefore, the instability of the chemical might affect the estimation. Lastly, Phthalic anhydride exclusively reacts with lysine residues (Natsch, 2010) and is therefore, outside the applicability domain of Keratinosens. In this model, 3 of 4 chemicals were classified as a weak sensitiser, which were judged as positive in only one of the in vitro tests. Therefore, weak response in the in vitro assays might affect the predictive capacity of this model. On the other hand, over-prediction of 18 chemicals was observed. 15 chemicals were judged either a weak sensitiser, or a non-sensitiser. This result might reflect a sensitivity of these in vitro test adopted in our model. Among 10 weak sensitisers that were predicted to be moderate sensitisers, 8 chemicals were judged as positive in all of the

3 in vitro tests, h-CLAT, SH test and ARE assay. Furthermore, 5 non-sensitisers that were predicted to be either weak or moderate sensitisers were judged as positive in at least one of these in vitro tests.

The R value (coefficient of correlation) between actual and predicted EC3 was 0.838, and RMS error was 0.628.

Please note that the predictive capacity depends on the dataset (62 chemicals were used in this model).

11. Consideration of uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

1. The DIP's structure

- Key event 4 is not included due to lack of available tests.
- The confidence is lower for chemicals which are out of the technical limitations (e.g., log Kow > 3.5).
- The confidence is lower for prehaptens and prohaptens due to limited metabolic capacities of test methods.
- The predictive capacity can be varied depending on dataset we adopt.
- In this approach, there is a hidden layer between input layer and output layer. Therefore, it is impossible to see from outside how the different data were weighted and combined.
- There is not an agreed approach to the validation of in silico methods such as neural network analysis.

2. The information sources used within the defined approach

- The DPRA and h-CLAT has been validated under the ECVAM. Reproducibility of peptide reactivity and CD86/CD54 measurements are very high. However, reproducibility of SH test and ARE assay has not been formally validated.

3. Benchmark data used

- The Integrated Testing Strategy (ITS) for hazard identification is based on the data from LLNA. The variability of EC3 values of LLNA has been reported depending on vehicle used and laboratories. Therefore, the uncertainty in misprediction of EC3 values is taken into account.

11.2 Impact of uncertainty on the DIP's prediction

The sources of uncertainty could cause to under- or over-estimate skin sensitisation potential. However, when the limitations are taken into account and similar chemicals are included in training set, confidence in the predictions is high.

12. References

Api A. M. et al. (2008). Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients, Regulatory Toxicology and Pharmacology, 52: p.3–23.

Ashikaga T. et al. (2010). A comparable evaluation of in vitro skin sensitization tests: the human cell line activation test (h-CLAT) versus the local lymph node assay (LLNA), ATLA, 38: p. 1–10.

Emter, R. et al. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin, 2 Toxicology and Applied Pharmacology, 45: p281–290.

Guy R.H., Potts R.O. (1993). Penetration of industrial chemicals across the skin: a predictive model, Am. J. Ind. Med, 23: p711-719.

Hirota M. et al. (2015). Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization, J. Applied Toxicology, Mar 30. doi: 10.1002/jat.3105.

Jaworska J. et al. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. J. Appl. Toxicol, 33 (11): p.1353-1364.

Natsch A. and Emter R. (2008). Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitization potential of chemicals. Toxicol. Sci, 102: p. 110–119.

Natsch A. (2010). The Nrf2-Keap1-ARE Toxicity Pathway as a Cellular Sensor for Skin Sensitizers—Functional Relevance and a Hypothesis on Innate Reactions to Skin Sensitizers. Toxicol. Sci, 113(2) :p. 284–292.

Nukada Y. et al. (2012). Data integration of non-animal tests for development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicology in Vitro, 26: p. 1150–1160.

Suzuki et al. (2009). Evaluation of changes of cell-surface thiols as a new biomarker for in vitro sensitization test. Toxicology in Vitro, 23: p.687–696.

Tsujita-inoue K et al. (2014). Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays. Toxicology in Vitro, 28: p. 626-639.

Valkova I. et al. (2004). Modeling of structure-mutagenicity relationships: counter propagation neural network approach using calculated structural descriptors. Anal. Chem. Acta, 509: p. 179-186.

13. Supporting information

Detailed information are given in the key publication Kyoto Tsujita-inoue et al., Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays, Toxicology in Vitro, 2014. 28: p. 626-639.

14. Abbreviations and definitions

ANN: Artificial Neural Network

ARE: Antioxidant Response Element

DPRA: Direct Peptide Reactivity Assay

EC3: Estimated concentration to produce a stimulation index of 3, the threshold value for a substance to be considered a sensitiser in the LLNA

EC150/200: Effective Concentrations (EC)150/200: concentrations at which the test chemical induces a Relative Fluorecence Intensity of 150 or 200.

h-CLAT: human Cell Line Activation Test

Imax: the highest average value of hold induction of luciferase activity in the ARE assay

LLNA: Local Lymph Node Assay

Log P: Octanol-water partition coefficient

MAC: Maximum Amount of Change The maximum change in the three dose settings was taken in the SH-test

QRA: Quantitive Risk Assessment

RMS error: Root Mean Square error

SH test: cell-Surface Thiol test

CASE STUDY IX

Sensitiser potency prediction based on Key event 1+2+3: Bayesian Network ITS/DS for hazard and potency identification of skin sensitisers

1. Summary

This DIP (BN-ITS-3) is based on Bayesian Network methodology. The Skin Sensitisation AOP structure (i.e. sequence of events, MIEs) as well as data related to KEs 1 (DPRA), 2 (Keratinosens), 3 (h-CLAT) are encoded in the BN ITS-3. This allows result interpretation in the biological context and is chemical specific. Cysteine and Lysine reactivity are treated as two separate, independent molecular initiating events (MIEs). BN ITS-3 uses metabolic transformation and auto-oxidation from TIMES-SS in the prediction process. Bioavailability considerations are applied to both in vivo and in vitro assays to represent an estimate of the potential to penetrate the stratum corneum and the free concentration respectively. Since the BN ITS-3 can reason based on partial information, only relevant data are used for predictions. This allows explicit consideration of the applicability domains of individual assays related to 1. Bioavailability domains based on phys-chem properties: water solubility and fraction ionised, and 2. biological domains: pre and prohaptens, as well as cytotoxicity. Data outside of domains are not included in the integrated prediction or treated with caution according to the prediction process. The prediction is given as potency probability distribution, the pEC3, in 4 potency classes: nonsensitisers (NS), weak (W), moderate (M), strong and extreme (S). Expressing prediction as a probability distribution naturally quantifies prediction uncertainty. In turn, it allows conversion of the prediction into a decision based on the strength of the evidence which is done using Bayes factors. Since the process of adding in vitro assay data to the BN ITS-3 can be cumulative it can also be used to guide and optimize testing strategy before testing is commenced.

2. General information

2.1 Identifier:

Bayesian network decision support system for a) sensitisation potency prediction based on bioavailability and Key event 1 + 2 +3 and b) for testing strategy.

2.2 Date:

10 April2016

2.3 Author(s) and contact details:

Joanna Jaworska, Procter & Gamble, Brussels Innovation Center, Temselaan 100, 1853 Strombeek – Bever, 1853 Belgium**;** Jaworska.j@pg.com**,** tel +32 2 456 2076

2.4 Template update(s): Version 1.

2.5 Reference to main scientific papers:

Jaworska, J., Gabbert, S., Aldenberg, T. (2010). Towards optimization of chemical testing under REACH: Bayesian network approach to Integrated Testing Strategies. Reg. Tox. Pharm 57, 157-167.

Jaworska. J, Dancik Y. Kern P. Gerberick F., Natsch A. (2013). A Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. J. App. Toxicology 33 (11): 1353–1364**.**

Pirone JR, Smith M, Kleinstreuer NC, Burns TA, Strickland J, Dancik Y, Morris R, Rinckel LA, Casey W, Jaworska JA. (2014). Open source software implementation of an integrated testing strategy for skin sensitization potency based on a Bayesian network. ALTEX 31:336-340

Jaworska JS, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M. (2015). . Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. Archives of Toxicology 89:2355-2383.

2.6 Proprietary aspects:

ACDlabs is a commercial software from ACDlabs. Inc. TIMES is a commercial software from LMC Bourgas. The recombinant cell-line used in the KeratinoSens™ assay requires a material transfer agreement and a license fee if the assay will be offered for commercial testing. The DPRA, h-CLAT and KerationoSens™ assays are offered by CRO's. Bayesialab is a commercial software. A previous version of BN ITS-3, namely BN ITS-2, is available open source in R via https://ntp.niehs.nih.gov/pubhealth/evalatm/integrated-testing-strategies/index.html and documented in Pirone et al 2014. Development of BN ITS-3 in R is under way.

3. Endpoint addressed

3.1 Endpoint:

The BN DIP estimates skin sensitisation potency in the LLNA, TG 429, expressed as probability distribution of LLNA pEC3, among 4 potency classes: nonsensitisers (NS), weak (W), moderate (M), and combined strong and extreme (S) sensitisers. In addition, it guides testing strategy by value of information.

3.2 Species:

Mouse

3.3 Additional information about the endpoint:

Skin sensitisation potency classes are often set based on EC3% values (NS, W EC3% 100-10, M EC3%10-1, S EC3% <1 %) (Kimber, 2003). However, from a potency prediction perspective, the use of molar concentrations is more relevant because effects resulting from exposure to chemicals are a function of number of molecules present at target sites, not the mass of these molecules. To this end we converted LLNA EC3 value usually expressed as % (w/v) to $\mathrm{pEC3} = \mathrm{Log}(\frac{\mathrm{MW}}{\mathrm{EC3\%}})$. The pEC3 cut-offs for classes representing C1, C2, C3, C4 classes are -1.9, -1.1, -0.35. These cutoffs were chosen to follow as closely as possible the classification based on EC3% (NS, W, M and S) in the training set.

Some molecules with very low/high MW are assigned to a higher/lower class respectively. The training set (n = 147) includes 36 NS, 28 W, 35 M, and 25 S sensitisers. The test set (n = 60) contains 12 NS, 21 W, 13 M, and 14 S chemicals. Distribution in the pEC3 space, i.e., C1–C4 classes, is 39, 39, 40, 29 for the training set and 14, 19, 12, 15 for the test set.

To facilitate the interpretation of results, equivalence of C1 and NS, C2 and W, C3 and M, and C4 and S is used below as it does not compromise the interpretation. NS, W, M and S for the 4 potency categories will be used throughout the DIP. For more details, see Jaworska 2015. Further, all the input data are mol/L concentrations.

4. Definition of the purpose and regulatory relevance

The BN DIP purpose is to provide skin sensitisation hazard and potency information for chemicals considering their mode of action. The information can be used:

a) for screening of chemicals for priority setting

b) for classification and labeling of chemicals under the GHS scheme (as grouping into 1A/1B can be made)

c) in quantitative risk assessment (as 4 potency categories can be provided)

d) for the development of an efficient testing strategy, when data are initially limited. There is no one best, predefined, testing strategy for all chemicals, as the optimal sequence of tests is chemical specific and depends on the information at hand. The testing strategy is driven by Value of Information (VoI) and progress is measured by uncertainty reduction.

5. Rationale underlying the construction of the defined approach

Inputs

The BN ITS-3 DIP uses the following data streams as input data:

1. Bioavailability in vivo and in vitro (physicochemical properties: $logD_{pH=7}$, water solubility $Ws_{pH=7}$, fraction ionised, serum albumin binding PB – ACDlabs)

2. In silico metabolism, potential for oxidation, potency prediction (TIMES)

3. KE 1: Peptide reactivity [OECD 442 C: Direct peptide reactivity test (DPRA)]

4. KE 2: Keratinocyte activation [OECD 442 D: ARE-Nrf2 luciferase test method (KeratinoSens^{TM})]

5. KE 3: Dendritic cell activation [human cell-line activation test (h-CLAT).

DIP structure

This DIP (BN ITS-3) is based on Bayesian Network methodology. A Bayesian network is a probabilistic graphical model (a type of statistical model) that graphically represents a set of variables and their conditional dependencies. These graphical structures are used to represent uncertain knowledge about a domain. In particular, each node in the graph represents a variable relevant to a domain/ process, while the arcs between the nodes represent probabilistic dependencies among the

corresponding variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics. More information on the computational details of BNs and why they are well suited as ITS framework can be found in Jaworska et al. 2010.

A Bayesian network approach allows converting integrated prediction to transparent, consistent decisions. It has the following features:

a) adaptive, as it can run with only partial evidence; it allows to add more evidence if needed to make a decision;

b) quantifies uncertainty for individual prediction based on the evidence entered. As such quantifies confidence in decisions and allows them to be fit for purpose;

c) assesses consistency in evidence and identifies conflict between input data.;

d) guides potential additional testing by quantifying the additional test information value before testing is commenced.

The structure of the BN ITS-3 model was developed manually from mechanistic knowledge of the endpoint following the approach outlined in Lucas et al. (2004). The AOP structure (i.e. sequence of events, MIEs) as well as data related to KEs 1, 2, and 3 are encoded in the BN ITS-3 which allows chemical specific result interpretation in the biological context. The hypothesis generated by the BN ITS-3 model can be explained based on known mechanisms. Mechanistic models are more robust and extrapolate better beyond data used to develop the model.

Both the construction method and the resulting structure of BN ITS-3 are similar to a previous version BN ITS-2 (Jaworska et al., 2013) but with several important refinements (for details refer to publications on BN ITS-2 and ITS-3) and its structure can be seen in Figure IX.1. Here for BN ITS-3 the mechanistic scheme of the skin sensitisation induction process (Basketter and Kimber, 2009) with the KEs of stratum corneum penetration, protein binding, keratinocyte activation and DC activation (Basketter and Kimber, 2010) were translated into a Naïve Bayes network structure. Naïve Bayes structure assumes that these events are independent. In the network the Bioavailability latent node relates to stratum corneum penetration potential as well as free concentration in vitro. The Cys latent node and Lys nodes relate to KE 1, peptide binding, and 2, keratinocyte activation (for Cys only). The h-CLAT latent node relates to KE3, DC activation, and combines information from all h-CLAT readouts. Second, the tests used to observe the above process were mapped onto the initial network as manifest variables. There are tests that clearly measure different key events and there are also tests that measure the same KE or part of the process but in different ways. Capturing this information is critical to the proper mapping of tests onto the initial network structure and is described below.

There are two possible MIEs: reaction with cysteine (Cys) and reaction with lysine (Lys), which are represented by two independent nodes. This allows identification of chemicals that act via both MIEs as well as only through one MIE. The Cys latent variable represents the event of cysteine haptenation that can be observed via the DPRA-Cys measurement and/or the KeratinoSens™ assay [a bias toward cysteine-reactive chemicals in Nrf2-dependent assays has been discussed previously (Natsch, 2010)]. Reactivity towards cysteine is also measured indirectly in TIMES as electrophilicity molecular descriptors. Further, it has been postulated that the molecular basis of DC stimulation by electrophilic chemicals is a reflection of their ability to bind to sensor proteins (such as Keap1 or others). Therefore, it was even argued that DC-based assays might be a complicated measure of cysteine reactivity (Kimber et al., 2011). To reflect this, arcs connecting Cys latent with h-CLAT, as well as Cys latent and TIMES were introduced.

BN ITS-3 also relates to bioavailability and cytotoxicity. Despite the obvious fact that a chemical must pass through the skin's stratum corneum barrier, most authors did not find bioavailability, usually

expressed as log Kow, to be a significant contributor to explain skin sensitisation hazard (Alves et al., 2015) or even potency (Roberts and Aptula, 2008). Our own efforts to express bioavailability in BN ITS-2 using absorbed dose, as well as maximum epidermal concentration, from the skin penetration simulation of the LLNA dosing scenario using the model developed by Kasting and coworkers (Dancik et al., 2013) showed a somewhat stronger relationship, especially for maximum epidermal concentration, but the effect was still small except for weak sensitisers (Jaworska et al., 2013).

While the role of skin penetration kinetics in in vivo skin sensitisation potency remains to be further elucidated, another kinetic component, kinetics in vitro, should also be considered in the ITS framework. Kinetics in vitro aims to assess the free concentration of a tested chemical in an in vitro test. The need for consideration of in vitro kinetics and the importance of using free instead of nominal concentration in the interpretation of the in vitro result has been demonstrated (Groothuis et al., 2015; Kramer et al., 2012) but remains to be routinely used. To this end, we decided to generalise the Bioavailability latent variable to consider both skin penetration in vivo and kinetics in vitro in the BN ITS-3 framework structure. The Bioavailability latent variable is constructed from the following physico-chemical properties: water solubility at pH=7, distribution coefficient, log D at pH=7, fraction ionised at pH=7, and % Plasma Protein Binding (PB). These variables are relevant determinants of skin penetration, cell membrane penetration, and free concentration. The Bioavailability latent variable is connected by arcs to LLNA pEC3, Cys, Lys, and h-CLAT nodes. The pEC3-Bioavailability arc represents bioavailability in vivo while the arcs with Cys, Lys and h-CLAT represent the respective bioavailabilities in vitro and in chemico.

In order to trigger the sensitisation response in vivo there is, after hapten formation, the need for a danger signal in the form of local trauma triggering the emigration of DC. This danger signal appears to involve the formation of extracellular ATP and breakdown products of hyaluronic acid generated by sensitisers (Esser et al., 2012; Weber et al., 2010). The release of ATP from cells is, at least under certain circumstances, triggered by cytotoxicity. For example, cytotoxic surfactants have the ability to provide this local trauma. In the LLNA, which we model in our analysis, no such adjuvant is given. Thus, in the LLNA, a chemical must provide both the hapten and the danger signal in order to trigger the response. Therefore, the LLNA measures both the haptenic potential and the danger signal provided by the chemical, and a chemical with stronger danger signal potential in principle will generate a stronger LLNA response. To account for the presence of the danger signal in the network, we connect the Cytotoxicity and pEC3 nodes. The Cytotoxicity latent variable is constructed from cytotoxicity measured in the h-CLAT (CV75) and the KeratinoSens™ assay (IC50). The arcs connecting IC50 with the KeratinoSens™ data inputs KEC1.5, KEC3, as well as CV75 with h-CLAT data inputs EC150 and EC200, inform about cell viability in relation to the sensitisation-specific response. Cytotoxicity in cell-based assays to a certain extent may mimic the ‘danger’ signals elicited by skin sensitisers in vivo, which might explain why cytotoxicity appears to be related to LLNA potency for some chemicals. However, it is important to keep in mind that this reasoning specifically applies to the experimental situation of the LLNA which is modeled in this work.

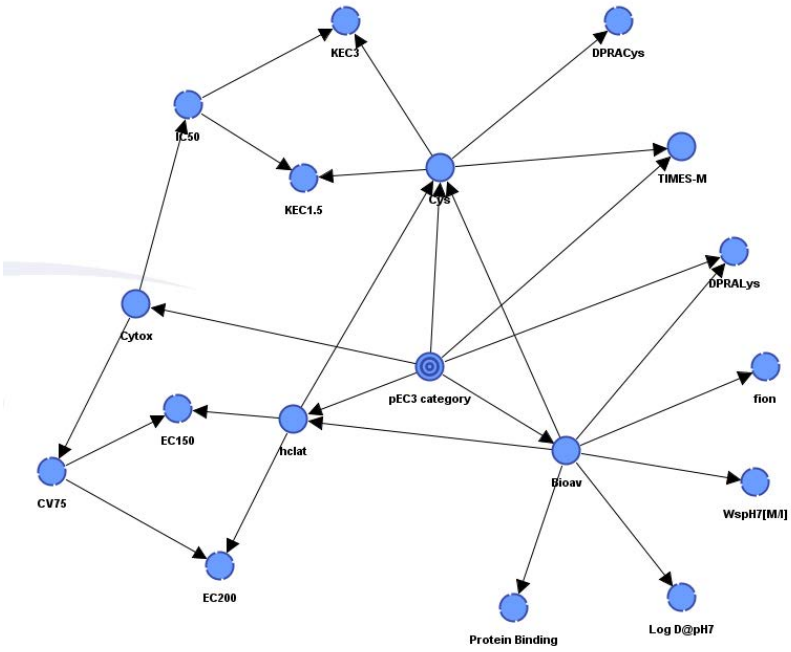


Figure IX.1: Structure of the BN ITS-3. The structure of the BN ITS model represents abstracted AOP and is developed based on mechanistic knowledge with the aim to follow sequence of the mechanistic events in the existing AOP. Two possible MIEs: reactions with Cysteine and/or DPRA are represented by two independent nodes. Bioav, Cys, hclat o and Cytox are latent variables that correct for dependency among their child nodes (arrow pointing to the node) in predicting pEC3 class

6. Description of the individual information sources used (see Annex II)

Table IX.1. Individual data inputs

Input type	**Endpoint**	**unit**
Bioavailability	1.Ws-Water solubility at pH=7 2. Log D- Distribution coefficient at pH=7 3. Plasma protein binding fraction 4. Fraction ionised	M/l [-] [-] [-]
KE 1:TIMES-SS	1. Mechanistic alert for direct reactivity (including direct Michael Acceptor) 2. Theoretical metabolites and their mechanistic alerts for direct reactivity 3. Prediction of 3 classes (Non-sensitiser (1), weak (2) or moderate/strong (3)) based on the most potent among parent and metabolites.	Classes (NS, W, S)
KE 1: DPRACys, DPRALys	% of the cysteine- (Cys), and lysine- (Lys) peptide remaining in the DPRA-assay	% remaining peptide
KE 2: KEC1.5, KEC3, IC50	1.5-fold (KEC1.5) ; 3-fold (KEC3) Induction of Nrf2-dependent luciferase activity in the KeratinoSens™ assay; 50% reduction in cell viability in the KeratinoSens™ assay	µM/l
KE 3: EC150, EC200, CV75	150% induction of the cell surface activation marker CD86 in the h-CLAT; 200% induction of the cell surface activation marker CD54 in the h-CLAT; 25% reduction in cell viability in the h-CLAT.	µM/l

Source & relevance

1. **ACDlabs v. 12.0** is used to calculate phys-chem properties related to Bioavailability. Bioavailability in vivo/in vitro converts external/nominal dose to the internal/free concentration for in vitro and potential to penetrate skin stratum corneum in vivo.

2. **TIMES SS Software** TIMES-SS software (V.2.27.13) (Dimitrov et al., 2005; Patlewicz et al., 2014) is an *in silico* hybrid expert system that 1) generates reactivity alert, b) assesses potential autooxidation, c) assesses metabolic transformation potential d) semi-quantitatively classifies chemicals into non-, weak sensitisers and strong potency categories in vivo. Intrinsic or metabolically triggered reactivity has a key biological relevance.

3. **DPRA** (OECD TG442c) addresses KE 1: the peptide binding. Binding of chemicals to the skin protein is an essential step for sensitiser to obtain allergenicity (OECD, 2012). Relevance: Reactivity is key for the MIE, intrinsic or metabolically triggered reactivity has a key biological relevance. MIEs assessed: Cys and Lys.Peptide reactivity is reported as percent of free peptide remaining in the sample, which is opposite to the original method reporting percent depletion. The data were generated at Procter & Gamble laboratories (Cincinatti, OH, USA).

4. **KeratinoSens™ ARE-Nrf2 Luciferase Test Method (OECD TG 442d)** addresses KE 2. Based on recent data, the Nrf2 pathway is a key pathway of defence triggered by sensitisers in vivo. According to current knowledge it is the key common molecular pathway which triggers gene expression in response to electrophilic chemicals at sub-toxic concentrations (Ade et al., 2009, Kim et al., 2008). Sensitisers with an exclusive reactivity towards lysine might be negative in the KeratinoSens™ assay.

Data were generated using the HaCaT keratinocyte-based cell line KeratinoSens™. The average concentrations (in µM) inducing a 1.5-fold or a 3-fold enhanced luciferase activity (KEC1.5 or KEC3.0, respectively) and the concentration leading to 50% cytotoxicity after 24 h (IC50, in µM) are determined. KEC3 data are used in addition to KEC1.5 data, because KEC1.5 may be too low a threshold for some reactive chemicals, (Emter et al., 2010). KeratinoSens™ data were obtained from Emter et al. (Emter et al., 2010) or generated at Givaudan laboratories.

5. **h-CLAT** addresses KE 3: DC activation. When a hapten is applied to the skin, surface molecules (i.e. CD54, CD86) on skin DCs are up-regulated through the maturation process. Since CD54 is involved in DC migration to draining lymph nodes and CD86 stimulates T cell activation during antigen-presentation by DC, both molecules are essential in the induction of skin sensitisation.

Data were generated with the Human Cell Line Activation test (h-CLAT) which uses the THP-1 cell line. The average concentrations (in µM) inducing a 150-fold increase in the CD86 cell surface marker or a 200-fold increase in cell surface CD54 expression (EC150 or EC200, respectively) and the concentration leading to 25% cytotoxicity after 24 h (CV75, in µM) were determined. Data were generated in Kao and Shiseido laboratories.

7. Data interpretation procedure applied

The process of deriving a hazard or potency prediction for a new chemical consists of two steps: gathering evidence and developing a quantitative hypothesis. This process was applied to all chemicals in the test set and in the case studies and it is summarized below:

1) Gathering evidence

   a) Calculation of physico-chemical properties of chemicals

   b) Prediction of sensitisation potency category using TIMES:

      i) Potency is based on the highest potency among parent molecule and predicted metabolites;

      ii) Assessment of potential of metabolic activation (pro-hapten) and auto-oxidation (pre-hapten) to facilitate interpretation of DPRA, KeratinoSens™ and h-CLAT assay results;

      iii) Determine whether a chemical is a direct MA based on reactivity alerts.

   c) Evaluation of the completeness of the evidence for MIEs: does the dataset have evidence for both cysteine and lysine reactivity?

   d) Assessment of applicability domains:

      i) If the chemical is deemed a potential pre- or pro-hapten via TIMES prediction, then DPRA, KeratinoSens™, and h-CLAT data are examined with caution, against potential conflict with other data. A hypothesis without these data is considered.

      ii) Solubility domain. Only data records not exceeding solubility cutoffs are considered in the analysis (Table IX.2). For chemicals with water solubility <2.5e-08 M, only TIMES and physico-chemical inputs characterizing bioavailability are used.

      iii) Ionization: chemicals that are completely ionised were not considered suitable for the in vitro assays.

Table IX.2. Water solubility at pH=7 cutoffs for DPRA, KeratinoSens™, hCLAT

Ws at pH=7 [M/l]	**DPRA**	**Keratinosens**	**hCLAT**
<2.5e-08	x	x	x
2.5e-08 - 1.7e-04	ok	x	x
1.7e-04 - 2.1e-04	ok	ok	x
> 2.1e-04	ok	ok	ok

2) Integration of all relevant in domain evidence via BN ITS-3 and prediction of the pEC3 probability distribution. (In case of missing evidence on MIE=Lys apply additional caution to the prediction (see reference for details).

   a) Post-processing correction of the probability distribution for MA, if applicable. See supplementary information.

   b) Analysis of the hypothesis based on cumulative evidence from combinations of relevant assays.

   c) Conversion of probability distribution to Bayes factors for final interpretation and acceptance of prediction.

Bayes Factors:

The use of Bayes factor removes biases in the predicted probability distribution introduced by distribution of a training set. Next, use of Bayes factors allows transparent expression of uncertainty in the prediction and eventually a consistent decision. The conversion is done using the following formula:

$$B = \frac{P(H = x|e)/P(H = not\_x|e)}{P(H = x/P(H = not\_x)} = \frac{posterior\ odds}{prior\ odds}$$

Where:

Prior distribution

P(H=x) - probability of a chemical to be in class x (x=NS, W, M, S) in the training set

P(H=not x) probability of a chemical to not to be in class x

Posterior distribution

P(H=x|e)- probability of a chemical to be in class x (x=NS, W, M, S) given the evidence provided to ITS-3

P(H=not x) probability of a chemical to not to be in class x given the evidence provided to BN ITS-3.

Table IX.3. Interpretation of Bayes factors in terms of strength of evidence (Jeffreys, 1961)

Bayes Factor	**Strength of evidence**
<1	Negative (supports alternative)
1-3	Barely worth mentioning (weak)
3-10	Substantial
>30	Strong

Additional comments to the process of collecting biological data

The BN ITS-3 uses quantitative weight of evidence based on Bayesian statistics to update the hypothesis about potency in LLNA after any new information is provided. The individual information sources are not used as stand-alone assays, but the outcomes are used to derive interim conclusions and to select, on the basis of value of information, which assays could be conducted next to increase evidence for the prediction.

The BN ITS-3 DIP generates a probabilistic hypothesis about skin sensitisation hazard and potency even with only 1 piece of information (i.e. data input). This hypothesis is updated upon the introduction of additional data inputs into the BN. The testing strategy is developed according to hypothesis driven Value of Information, which is expressed as mutual information (MI) values. In

addition, by considering dependencies between tests, the BN ITS also proposes alternative tests if the most optimal test is not available. The BN ITS-3 DIP also identifies which tests are not effective to be used simultaneously. This is a big asset in practical application, in which chemical safety data may be incomplete and the assessor must make decisions.

8. Chemicals used to develop and test the DIP

8.1 Availability of training and test sets:

The training and test sets, including SMILES experimental data and in silico predictions, are available in Jaworska et al. 2015 publication and can be accessed through this link http://link.springer.com/article/10.1007%2Fs00204-015-1634-2.

8.2 Selection of the training set and test set used to assess the DIP:

The training and test sets are composed 147 and 60 chemicals, respectively (total of 207) coming from: Jaworska et al. 2013, Urbisch et al. 2015, RIFM data generation project, and additional testing to fill in data gaps conducted by Kao, Shiseido, Givaudan, and P&G. The database was built by first selecting quality LLNA data and later filling in data gaps for the alternative assays (i.e. DPRA, h-CLAT and KeratinoSens™). Attention was paid to balance number of chemicals in each potency class as well as structural diversity that represent chemicals used in consumer products. Training and test set were determined by balancing structural diversity, and as even as possible representation of chemicals in all 4 chemical classes.

8.3 Supporting information on the training and test sets:

The whole database, including SMILES experimental data and in silico predictions, is available in the Jaworska et al. 2015 publication. The training set with predictions is also provided.

8.4 Other information on the training and test sets:

Not applicable

9. Limitations in the application of the defined approach

BN ITS-3 system requires biological (in vitro) data input of reliable consistent quality. The data need to come from within the applicability domains of the individual assays (DPRA, KeratinoSens™, h-CLAT):

a) In vitro assays are applicable to test chemicals soluble in either water or DMSO and test chemicals that form a stable dispersion;

b) Highly cytotoxic chemicals cannot be tested in the in vitro assays

c) Prohaptens: Experimental assays have varying metabolic capacity in the following order: DPRA (none) < h-CLAT < Keratinosens. These assays may yield underestimation of potency.

d) Prehaptens: experimental assay results may yield underestimation of potency.

10. Predictive capacity of the defined approach

The BN ITS-3 estimates skin sensitisation potency in the LLNA, TG 429, expressed as probability distribution of LLNA pEC3, among 4 potency classes: nonsensitisers (NS), weak (W), moderate (M), and combined strong and extreme (S) sensitisers.

Accuracy and precision with full and partial evidence

The strength of the BN ITS-3 is its ability to provide a prediction with either all or partial data inputs. Therefore, only results that are within the applicability domains of the individual assays are recommended to be used when predicting potency of new molecules. When all evidence is entered to the system the hazard prediction accuracy (Yes/ No) for the test set expressed as a balanced accuracy (bac) is 100% (Table IX.4). Bac accounts for uneven distribution of positive and negative chemicals in a dataset. For a binary classification the formula is:

$$bac = \frac{Se + Sp}{2}$$

Where Se = sensitivity and Sp = specificity. For 4 class prediction bac is calculated per class.

Extending bac to multiple classes, one gets bac for GHS C&L = 96%; bac for four-class potency = 89%. The four-class potency accuracy of 89% is in excellent agreement with accuracy for the training set (85%). It demonstrates that the BN ITS-3 model is not overfitted. The higher accuracy for the test set reflects the fact that we have a pre-processing step of selecting data only from their physico-chemical applicability domains and a post-processing step of MA correction. The BN ITS-3 achieved 100% accuracy for 14 NS chemicals. It also reliably predicted the weak (89%) and strong (87%) class chemicals. When two problematic acrylates are removed, it predicted W chemicals 100% correctly. However, the model has a drop in accuracy for M, correctly predicting 75% of the chemicals.

In Table IX.4 predictions with varying degree of uncertainty are accepted as long Bayes factor (B) >1. One can easily modify criteria for acceptance. For example a non-sensitiser prediction can be accepted only when B>3 (strong evidence), while B>1 can be deemed sufficient to accept chemical to be a sensitiser.

Table IX.4. Predictive capacity of the approach given as a contingency matrix based on the highest Bayes factor. Numbers in parentheses indicate number of chemicals.

GHS category	Observed									
	Training set (147)					Test set (60)				
	Class	NS(39)	W(39)	M(40)	S(29)	Class	NS(14)	W(19)	M(12)	S(15)
none	NS	**36**	2	1	0	NS	**14**	0	0	0
1B	W	2	**32**	3	3	W	0	**17**	3	0
	M[a]	0	3	**38**	5	M	0	2	**9**	2
1A[a]	S	1	2	8	**21**	S	0	0	0	**13**

[a]Since the GHS classification cutoff for 1A is ≤2% EC3 value in LLNA (not fully in line with our cutoff between the potency groups), the table provides a more conservative classification. Further differences are to be expected due to the conversion from weight to molar units.

When using all information, the following seven chemicals in the test set were mispredicted (Table IX.5). One should look at the mispredictions from the side of in vivo data and alternative data inputs. Among the seven chemicals mispredicted there are four cases where the in vivo data are not reliable: 2 acrylates, tocopherol and anhydride. Of the remaining three, two chemicals are out of the in vitro assay domains due to poor water solubility.

Table IX.5. List of chemicals in the test set mispredicted by BN ITS-3. Bolded numbers denote class based on experimental data.

Chemical	EC3%	B (NS)	B (W)	B (M)	B (S)	Explanation
Ethyl acrylate	28.0	0.2	**0.8**	2.2	1.5	High vapor pressure, in vivo results likely under-predicted due to evaporation
Methyl acrylate	20.0	0.0	**0.8**	2.6	1.6	High vapor pressure, in vivo results likely under-predicted due to evaporation
Dihydroeugenol(2-methoxy-4-propyl-phenol)	6.8	0.0	5.4	**0.7**	0.6	Pro-hapten, however removal of DPRA yields class S
Farnesol	4.1	0.3	2.0	**1.6**	0.4	Predicted by TIMES as pre-hapten, however removal of DPRA yields class S. KeratinoSens™ and h-CLAT out of solubility domain
Tocopherol	7.4	0.4	5.1	**0.4**	0.5	logP=10.6, result based on TIMES only, DPRA, KeratinoSens™ and h-CLAT out of solubility domain. Tocopherol / Vitamine E is not a human sensitiser and LLNA may be false-positive
1,2-cyclohexane dicarboxylic anhydride (hexahydrophthalic anhydride)	0.8	0.1	0.3	4.6	**1.3**	This chemical quickly hydrolyzes in water. However, in DPRA reactivity is so fast that it is even faster than hydrolysis (if peptide added first). KeratinoSens™ and h-CLAT out of solubility domain for the parent molecule however it is more likely that an acid is tested. Because the acid is very strong it will fall out from applicability domain based on f_ion.[a]

Squaric acid diethyl ester	0.9	0.4	1.1	3.9	**0.1**	This chemical quickly hydrolyzes in water, in vitro assays test not the parent chemical but acid and alcohol (Cohen and Cohen 1966)

[a]Similar chemical (phthalic anhydride [CAS# 85-44-9] , a known misclassified extreme sensitiser hydrolyzes in water at pH 6.8 - 7.24 with half-lives of 0.5 - 1 min at 25 °C, forming phthalic acid and is therefore not within the applicability domain of the in vitro assays (UNEP 2005). Phthalic acid [CAS# 88-99-3] is classified as a non-sensitiser by a modification of the Maguire method and the LLNA (ECHA database on registered substances, searched on 25.07.2014).

Predicting potency without one of the AOP KE assays

From a mechanistic point of view the three assays have fairly large information overlap. Many models were published, using information on subsets of KEs with good results. We (Jaworska et al., 2013) and others (Natsch et al., 2015) have shown previously that a correct prediction of potency does not always require entering information from all possible assays. Our results in Table IX.6 further confirm this observation.

Table IX.6. Accuracy of potency predictions for the test set in % for either full data input or with omission of one of the KE assays.

GHS C&L	Potency class	All data	w/o DPRA	w/o Ksens	w/o h-CLAT	w/o TIMES	n
	all	89	82	77	75	74	60
none	NS	100	100	93	93	87	14
1B	W	90	83	83	83	89	19
	M	75	58	50	45	58	12
1A	S	87	87	80	80	60	15

However, the loss of accuracy when information from one of the assays is not provided is more prominent for M and S classes than for NS and W. In other words our results indicate that the highest information overlap is in predicting NS. This may not be too surprising given that all three in vitro assays (and the LLNA) were initially developed to discriminate sensitisers from non-sensitisers – so with the same primary goal to provide this specific information. Omitting TIMES has the biggest effect overall driven by the loss of accuracy for NS, M, and S. This result is a combination of the high TIMES predictivity for NS but also the fact that six of 15 chemicals in the S class have very poor solubility that makes the in vitro data out of the applicability domain. Thus, these predictions w/o TIMES use only bioavailability. Omitting h-CLAT and KeratinoSens^{TM} has the same effect on accuracy for all the classes while DPRA has the smallest effect.

Accuracy presents only one aspect of the predictive value of the system. Accuracy refers to the agreement between measured and predicted value. The other aspect of the predictive value, independent from accuracy, is provided by precision. It tells us about uncertainty of the prediction. Bayes factors are expressions of precision. Only while analyzing accuracy and precision together one can make a choice about the evidence needed to make a decision. Details on considerations around precision can be found in the referenced publications.

From prediction to decision: Quantification of uncertainty for decision making- converting probability-based predictions to Bayes factors

BN ITS-3 probability distribution results are converted to and expressed as Bayes factors. The use of Bayes factors corrects for the distribution of the chemicals in the training set and therefore provides a more objective prediction than a posterior probability distribution. In the BN ITS-3 training set, S chemicals are slightly underrepresented (compare the probability of the prior distribution of pEC3 S with those of the other three classes in Table IX.7). This results in the deflated posterior probabilities for this class. In the example of benzo(a) pyrene this leads to the conclusion that this is a M chemical based on probabilities (Pr(C3)=0.37). However, when predictions are based on the highest Bayes factor (B), benzo(a) pyrene is predicted as S, which is concordant with experimental data. Since B(S) is only 1.75, we conclude that the evidence for this chemical to be S is weak. Similarly, for chemical classes that are over-represented, the prediction probabilities are inflated.

Table IX.7. Prior and posterior distribution probabilities and Bayes factors for benzo(a)pyrene, CAS# 50-32-8.

Prior distribution as in the training set				Posterior distribution predicted by BN ITS-3				Bayes factors			
pEC3 NS	pEC3 W	pEC3 M	pEC3 S	pEC3 NS	pEC3 W	pEC3 M	pEC3 S	B (NS)	B (W)	B (M)	B (S)
0.27	0.27	0.27	0.19	0.04	0.29	**0.37**	0.30	0.11	1.11	1.60	**1.75**

11. Consideration of uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

The DIP structure and information sources used within the defined approach.

Assumption	**Direction & Magnitude**
Bayesian network structure correctly represents the biological mechanism of the induction of skin sensitisation	The structure of the DIP model represents abstracted AOP and is developed purely based on mechanistic knowledge with the aim to follow sequence of the mechanistic events in the existing AOP.
The dataset robustly characterizes parameters of the network, the conditional probability tables.	The x-validation done using bootstrapping shows a very stable network.
The phys-chem properties sufficiently characterize bioavailability in vitro and in vivo.	These parameters are key inputs to skin penetration model as well key parameters for in vitro kinetics.

Metabolic activation and autooxidation are sufficiently characterized by TIMES	false positive will yield overestimation, false negative will yield underestimation.

Variability in vitro is expected to be smaller than variability in vivo. In vitro assay systems are more biologically simple and more standardized than the in vivo systems. The between-laboratory reproducibilities for nonsensitiser/sensitiser outcomes for DPRA, KeratinoSens™, and h-CLAT were: 75%, 86%, and 80% respectively (Joint Research Centre of the European Union, 2013; Joint Research Centre of the European Union, 2014; Joint Research Centre of the European Union, 2015).

Benchmark data used.

The variability of the *in vivo* data should be taken into account, originating from two sources: test variability and variability between laboratories (the reference databases come from multiple centres and were built up over two decades for the LLNA and 3-4 decades for the human data). Repeated LLNA values in the ICCVAM database vary 1.7 – 2-fold on either side of the mean (ICCVAM, 2008) and therefore a 2-fold misprediction of an EC3 may in many cases still falls within the variability of the true *in vivo* outcome. The uncertainty is even higher for the human data obtained with different test protocols. Correlation of the LLNA with the human LOEL is far from perfect (ICCVAM, 2011), which can partly be attributed to these limitations in the human dataset.

11.2 Impact of uncertainty on the DIP's prediction

Deterministic models have very limited scope for correctly handling intrinsic data uncertainty while probabilistic models have a naturally built in capability to handle it. The ITS-3 prediction for a new chemical, being probabilistic, inherently includes assessment of uncertainty associated with this prediction. Further, conversion to Bayes factors allows for a consistent acceptance of uncertainty in predictions based on fit for purpose criteria. This uncertainty reflects the combined uncertainty associated with ITS-3 structure and, in part, uncertainty due to the variability of input information sources as well as the target, i.e. LLNA pEC3.

12. References

Ade N, Leon F, Pallardy M, Peiffer J. L, Kerdine-Romer S, Tissier, M. H, Bonnet P.A, Fabre I and Ourlin J. C. (2009). HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: Role of the Keap1/Nrf2 pathway.Toxicol. Sci, 107, 451-460.

Alves VM, Muratov E, Fourches D, Strickland J, Kleinstreuer N, Andrade CH, Tropsha A. (2015). Predicting chemically-induced skin reactions. Part II: QSAR models of skin permeability and the relationships between skin permeability and skin sensitization. Toxicol Appl Pharmacol, 284(2):273–280.

Basketter DA, Kimber I. (2009). Updating the skin sensitization in vitro data assessment paradigm in 2009. J Appl Toxicol, 29(6):545-550.

Basketter D, Kimber I. (2010). Re: Updating the skin sensitization in vitro data assessment paradigm in 2009 - A chemistry and QSAR perspective. J Appl Toxicol, 30(3):289

Dancik Y, Miller MA, Jaworska J, Kasting GB. (2013). Design and performance of a spreadsheet-based model for estimating bioavailability of chemicals from dermal exposure. Adv Drug Deliv Rev, 65(2):221-36

Dimitrov SD, Low LK, Patlewicz GY, Kern PS, Dimitrova GD, Comber MH, Phillips RD, et al. (2005). Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. Int J Toxicol, 24(4):189-204

Emter R, Ellis G, and Natsch A. (2010). Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. Toxicol. Appl. Pharmacol, 245: p. 281-290.

Esser PR, Wolfle U, Durr C, von Loewenich FD, Schempp CM, Freudenberg MA, Jakob T, et al. (2012). Contact sensitizers induce skin inflammation via ROS production and hyaluronic acid degradation. PLoS ONE, 7(7):e41340

Groothuis FA, Heringa MB, Nicol B, Hermens JL, Blaauboer BJ, Kramer NI. (2015). Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. Toxicology, 332:30-40.

Haneke K.E. et al. (2001). ICCVAM evaluation of the murine local lymph node assay. Data analyses completed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. Regul Toxicol Pharmacol, 34(3): p. 274-86.

ICCVAM, (2008). ICCVAM Independent Scientific Peer Review Panel Report: Validation Status of New Versions and Applications of the Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals and Products. National Institute of Environmental Health Sciences, Research Triangle Park.

ICCVAM, (2011). ICCVAM Test Method Evaluation Report: Usefulness and Limitations of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans. National Institute of Environmental Health Sciences, Research Triangle Park.

Jaworska J, Dancik Y, Kern P, Gerberick F, Natsch A. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. J Appl Toxicol, 33(11):1353-64.

Jaworska JS, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M. (2015). Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. Archives of Toxicology, 89:2355-2383

Jeffreys H.(1961). The Theory of Probability (3 ed.), p. 432

Joint Research Centre of the European Union, (2013). EURL ECVAM Recommendation on the Direct Peptide Reactivity Assay (DPRA) for Skin Sensitisation Testing. Publications Office of the European Union.

Joint Research Centre of the European Union, (2014). EURL ECVAM Recommendation on the KeratinoSens™ assay for skin sensitisation testing. Publications Office of the European Union.

Joint Research Centre of the European Union, (2015). EURL ECVAM Recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitisation testing. Publications Office of the European Union.

Kim H. J, Barajas B, Wang, M.and Nel A. E. (2008). Nrf2 activation by sulforaphane restores the age-related decrease of T(H)1 immunity: role of dendritic cells. J Allergy Clin Immunol, 121, 1255-1261 e7.

Kimber I, Basketter DA, Butler M, Gamer A, Garrigue JL, Gerberick GF, Newsome C et al. (2003). Classification of contact allergens according to potency: proposals. Food Chem Toxicol, 41(12):1799–1809

Kimber I, Basketter DA, Gerberick GF, Ryan CA, Dearman RJ. (2011). Chemical allergy: Translating biology into hazard characterization. Toxicol Sci, 120(SUPPL.1):S238-S268.

Kramer NI, Krismartina M, Rico-Rico A, Blaauboer BJ, Hermens JL. (2012). Quantifying processes determining the free concentration of phenanthrene in Basal cytotoxicity assays. Chem Res Toxicol, 25(2):436-45

Lucas PJ, van der Gaag LC, Abu-Hanna A. (2004). Bayesian networks in biomedicine and health-care. Artif Intell Med, 30(3):201-14.

Natsch A. (2010). The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers--functional relevance and a hypothesis on innate reactions to skin sensitizers. Toxicol Sci, 113(2):284-92.

Natsch A., Haupt T, and Laue H. (2011). Relating skin sensitizing potency to chemical reactivity: reactive Michael acceptors inhibit NF-kappaB signaling and are less sensitizing than S(N)Ar- and S(N)2- reactive chemicals. Chem Res Toxicol, 24(11): p. 2018-27.

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(2):319-332.

OECD, (2010). Test No. 429. Skin Sensitisation: Local Lymph Node Assay OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. OECD Publishing. http://www.oecd.org/env/test-no-429-skin-sensitisation-9789264071100-en.htm

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

OECD, (2015b). Draft Proposal for a New Test Guideline. *In Vitro* Skin Sensitisation: human Cell Line Activation Test (h-CLAT). OECD Publishing. http://www.oecd.org/env/ehs/testing/Draft-Proposal-for-a-new-Test-Guideline-on-in-vitro-skin-sensitisation-h-CLAT.pdf

OECD, (2015c). Test No. 442C. *In Chemico* Skin Sensitization: Direct Peptide Reactivity Assay (DPRA) OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. OECD Publsihing. http://www.oecd.org/env/test-no-442c-in-chemico-skin-sensitisation-9789264229709-en.htm

OECD, (2015d). Test No. 442D. I*n Vitro* Skin Sensitisation: ARE-Nrf2 Luciferase Test Method OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. OECD Publishing. http://www.oecd.org/env/test-no-442d-in-vitro-skin-sensitisation-9789264229822-en.htm

Patlewicz G, Kuseva C, Mehmed A, Popova Y, Dimitrova G, Ellis G, Hunziker R, et al. (2014). TIMES-SS--recent refinements resulting from an industrial skin sensitisation consortium. SAR QSAR Environ Res, 25(5):367-91

Pirone JR, Smith M, Kleinstreuer NC, Burns TA, Strickland J, Dancik Y, Morris R et al. (2014). Open source software implementation of an integrated testing strategy for skin sensitization potency based on a Bayesian network. ALTEX, 31(3):336–340

Roberts DW, Aptula AO. (2008). Determinants of skin sensitisation potential. J Appl Toxicol, 28(3):377-387

Roberts, D.W. and Natsch A. (2009). High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of michael acceptor electrophiles. Chem Res Toxicol, 22(3): p. 592-603.

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(4):599-609.

UNEP, (2005). OECD SIDS. Phthalic anhydride. CAS No: 85-44-9. UNEP Publishing.

Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, et al. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul Toxicol Pharmacol, 71(2):337-51

Weber FC, Esser PR, Muller T, Ganesan J, Pellegatti P, Simon MM, Zeiser R, et al. (2010). Lack of the purinergic receptor P2X(7) results in resistance to contact hypersensitivity. J Exp Med, 207(12):2609-19.

13. Supporting information

Water Solubility cutoffs

A unifying limitation of the cell based assays and to a lesser degree in chemico assay is water solubility (Joint Research Centre of the European Union, 2013; Joint Research Centre of the European Union, 2014; Joint Research Centre of the European Union, 2015). Traditionally solubility cutoff of an assay has been expressed as a function of log Kow. The published cutoff for hCLAT were log Kow>3.5 (Takenouchi et al., 2013) and log Kow >5 for KeratinoSens™ (OECD, 2015). Till now DPRA assay was considered not to have a solubility cutoff because for poorly soluble chemicals there is always a possibility to add DMSO up to 10% of solution.

However, Log Kow is a good surrogate of water solubility only for neutral chemicals. Partially or fully ionised chemicals are much more soluble in water than their neutral counterparts. In general water solubility is pH dependent. In order to generalise the cutoff also to chemicals that are ionised at physiological pH we calculated water solubility at pH=7 and express the cutoff based on this variable. Water solubility at pH=7 was calculated using ACD labs software. It is worth noting that majority of softwares offers only intrinsic solubility, i.e. solubility of the neutral molecule, calculations without clearly explaining this fact. As expected chemicals with log Kow>3.5 for hCLAT and log Kow>5 for Keratinosens™ revealed a wide range of solubility values as these chemicals have a diverse degrees of ionization. The cutoffs were chosen as the largest solubility among the chemicals with log Kow>3.5 and log Kow>5 for hCLAT and KeratinoSens™ respectively. This resulted in a very close cutoff values for hCLAT and KeratinoSens™ (Table IX.8). Very similar solubility cutoffs makes sense as both assays require similar medium composition but with slightly different buffer capacity. Only data records not exceeding solubility cutoff were considered in the analysis of the test set. Solubility cutoffs were not considered in the training set as this would require either all the chemicals with solubility less than 2.1e-4 [M/l] that would result in a loss of valuable information. Instead, we chose to keep all the chemicals and retain records below solubility limits that may introduce noise to the data.

Table IX.8. Water solubility at pH=7 cutoffs for DPRA, KeratinoSens™, hCLAT

Ws at pH=7 [M/l]	**DPRA**	**Keratinosens**	**hCLAT**
<2.5e-08	x	x	x
2.5e-08 - 1.7e-04	ok	x	x
1.7e-04 - 2.1e-04	ok	ok	x
> 2.1e-04	ok	ok	ok

For chemicals with Ws<2.5e-08M/l only TIMES is used together with phys-chem inputs.

Fraction ionised

Chemicals that were 100% ionised at pH=7 were deemed not to be suitable for cell based assays due to lack of bioavailability, i.e. not crossing cell wall. For partially ionised chemicals we assumed that while bioavailability is impaired in terms of rate of crossing cell wall, testing period is sufficient for the chemical to cross cell wall and reach the target. Fraction ionised was calculated from the logKow and $logD_{pH=7.4}$

Direct Michael Acceptor (MA) correction

Natsch et al. 2011 noted that chemicals with this substructure are less sensitising in vivo than would be inferred from chemical reactivity data, due to the anti-inflammatory action of MA. Further, Natsch et al. 2011 showed that the anti-inflammatory activity increases with chemical reactivity for this class of molecules. Since this MA alert does not translate to potency a priori, it was not included it in the structure of the network. However, it was decided to use information about this alert in the predictions by manually modifying the hypothesis toward a weaker class in the following manner: if a chemical is predicted S (strong sensitiser), 70% probability mass was moved from S to M (moderate); if predicted M, 40% probability mass was moved from M to W (weak) class; if predicted W, 40% probability mass was moved from W to NS (non-sensitiser); if predicted NS, there was no change in the probability distribution. This modification corresponds to the direct MA effect presented in Natsch et al. 2011 graphic for Table of Contents and was applied only to the test set in Table 3.

14. Abbreviations and definitions

Not applicable

CASE STUDY X

Sequential testing strategy (STS) for sensitising potency classification based on *in chemico* and *in vitro* data

1. Summary

The defined approach presented in this document is constructed as a sequential testing strategy (STS) for prediction of the skin sensitisation potential and potency of a substance. The test methods used address two key events (KEs) 1 and 3 as defined in OECD Adverse Outcome Pathway (AOP) of skin sensitisation: KE 1 of protein binding is evaluated using the Direct Peptide Reactivity Assay (DPRA; OECD TG 442c); KE 3 of dendritic cell activation is evaluated using the human cell line activation test (h-CLAT). This STS described here is based on 139 chemicals classified a Sensitisers / Non-Sensitisers (S/NS) in the LLNA. The underlying rationale of this STS is that any substance which is judged as positive by either DPRA or by h-CLAT is a sensitiser, i.e. one positive result in either DPRA or h-CLAT indicated a sensitiser, and two negative results indicated a non-sensitiser. The quantitative dose-response outcomes in the h-CLAT do correlate to sensitising potency based on the EC3 values in the LLNA. The strong class in the h-CLAT is available to predict EC3<1% in LLNA (Strong). Either the weak class in the h-CLAT or the positive result in the DPRA is available to predict EC3≥1% (Weak). The potency prediction is given as three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS), but is not used for GHS sub-classification into sub-category 1A (EC3<2% in LLNA) and 1B (EC3≥2%).

2. General information

2.1 Identifier:

Sensitising potency classification based on AOP KEs 1 and 3

2.2 Date: 3 May2016

2.3 Author(s) and contact details:

Masaaki Miyazawa, Ph.D., R&D – Core Technology – Safety Science Research, Kao Corporation, 2606 Akabane, Ichikai-machi, Haga-gun, Tochigi 321-3497, Japan miyazawa.masaaki@kao.co.jp; Tel +81-285-68-7342

2.4 Template update(s): Version 1.

2.5 Reference to main scientific papers:

Key paper:

Takenouchi et al. (2015). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. J Appl Toxicol 35, 1318-1332.

Nukada et al. (2013). Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicol in Vitro, 27, 609-618.

Nukada et al. (2012). Prediction of skin sensitization potency of chemicals by human Cell Line Activation Test (h-CLAT) and an attempt at classifying skin sensitization potency. Toxicol in Vitro, 26:1150-60.

2.6 Proprietary aspects:

None

3. Endpoint addressed

3.1 Endpoint:

This STS predicts skin sensitisation potential (hazard identification: S/NS) and potency (three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS)), based on readouts from assays addressing AOP KEs 1 and 3. Thus, it is addressing the endpoint of existing test guidelines e.g. Local Lymph Node Assay (LLNA) (OECD TG 429), Guinea Pig Maximisation Test (GPMT) and Buehler Test (OECD TG 406).

3.2 Species:

Mouse (primary target), Human (secondary target).

3.3 Additional information about the endpoint:

Skin sensitisation is the result of a complex multifactorial sequence of events. There are four KEs in the chemical and biological pathways (protein binding, keratinocyte activation, dendritic cell activation, and proliferation of antigen-specific T-cells) as the AOP for skin sensitisation defined by the OECD (2012). This STS uses quantitative data from in chemico assay addressing the intrinsic chemical reactivity (KE 1) and in vitro assay addressing capacity to induce dendritic cell activation (KE 3).

4. Definition of the purpose and regulatory relevance

This STS was developed to predict the skin sensitisation potential (Sensitisers or Non-Sensitisers) of a substance for the purpose of classification and labelling under the GHS scheme without animal testing. The potency prediction of this STS will be useful to obtain three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS), but is not used for GHS subclassification into subcategory 1A (EC3<2% in LLNA) and 1B (EC3≥2%).

5. Rationale underlying the construction of the defined approach

This STS is developed based on the molecular initiating event (KE 1) and the cellular response of dendritic cells (KE 3) in the skin sensitisation AOP to predict the sensitising potential and potency. The molecular initiating event (KE 1) leading to skin sensitisation is postulated to be covalent binding of electrophilic chemical species to selected nucleophilic molecular sites of action in skin proteins.

The covalent binding to skin proteins is evaluated using the Direct Peptide Reactivity Assay (DPRA), which is now the OECD TG 442C. The activation of dendritic cells (DC) is typically assessed by expression of specific cell surface markers, chemokines and cytokines. The h-CLAT is proposed to address the KE 3 (dendritic cell activation) of the skin sensitisation AOP and is now the OECD TG 442E.

In this STS, the assay related to KE 2 is not included, but DPRA cysteine depletion (KE 1) and KeratinoSens™ covering KE 2 are mechanistically relevant (Joanna et al., 2013). The key molecular pathway (Nrf2-ARE pathway) induced in KeratinoSens™ corresponds to cysteine reactivity with the Keap1 sensor protein. In addition, the Nrf2 activation is induced by sensitisers and not by non-sensitisers in THP-1 cells, and could function as one of the danger signals to lead to the phenotypic alterations on THP-1 cells (Migdal et al., 2013; Ade et al., 2009). Thus, there is a mechanistic rationale that DPRA and h-CLAT could be linked to KeratinoSens™ (KE 2).

The underlying rationale of this STS is that any substance which is judged as positive by either DPRA OR by h-CLAT is a sensitiser, i.e. one positive result in either DPRA or h-CLAT indicated a sensitiser, and two negative results indicated a non-sensitiser.

6. Description of the individual information sources used

1. DPRA (OECD TG 442C) is addressing the peptide binding. Haptens applied to the skin are covalently binding to nucleophilic residues (i.e. cysteine, lysine) in dermal proteins. Binding of chemicals to the skin protein is an essential step for sensitiser to obtain allergenicity (OECD, 2012).

   Substances that induced mean peptide depletion of cysteine- and lysine-containing peptide above 6.38% are considered to have peptide reactivity of sensitiser. Substances predicted as sensitiser are defined to have potential to be strong or weak sensitiser based on 22.62% of mean peptide depletion.

2. h-CLAT is addressing DC activation. When a hapten is applied to the skin, surface molecules (i.e. CD54, CD86) on skin DCs were up-regulated through the maturation process. Since CD54 is involved in DC migration to draining lymph nodes and CD86 stimulates T cell activation during antigen-presentation by DC, both molecules are essential in the induction of skin sensitisation.

   Substances inducing a fold induction greater than 2-fold for CD54 and/or 1.5-fold increase for CD86 at cell viabilities above 50% are predicted to have a DC activating potential of sensitiser (Ashikaga et al., 2010). From the dose-dependency curves of experiments, the median concentration(s) inducing 1.5- and/or 2-fold induction of CD86 and/or CD54 are calculated and the resulting lower value is defined as minimal induction threshold, MIT. Substances predicted as positive are defined to have potential to be strong or weak sensitiser based on 10 µg/mL of MIT (Nukada et al., 2013).

7. Data interpretation procedure (DIP) applied

In this STS, in order to predict skin sensitising potential, the conservative decision is conducted by weighing one positive result in the individual assay. In fact, the conservative decision approach using two assays (DPRA and h-CLAT) vs three assays (DPRA, KeratinoSens™, and h-CLAT) could be compared to predict skin sensitising potential based on the 101 chemical dataset (Urbisch et al., 2015), but it was found that it only slightly improves sensitivity and markedly decrease specificity. Moreover, both DPRA and h-CLAT, but not KeratinoSens™, can provide the information related to sensitising

potency (three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS)) in the individual assay.

Regarding the sequential order of testing, h-CLAT, which has higher sensitivity and accuracy than DPRA in the training data set, is designated as a first step for the STS, taking into account the time and cost effectiveness by reducing the number of false negatives; positive results in h-CLAT are classified as strong or weak class based on an cut-off of 10 μg/ml (Figure X.1). When the MIT is above 10 and below 5000 μg/ml, the substance is classified as a weak sensitiser. In the case of negative results in h-CLAT, DPRA is performed as a second test. If DPRA yields positive results, the chemical is classified as a weak sensitiser. Test chemicals that were negative in both tests were considered as not-classified. This yields a qualitative result (positive/negative and three rank classes).

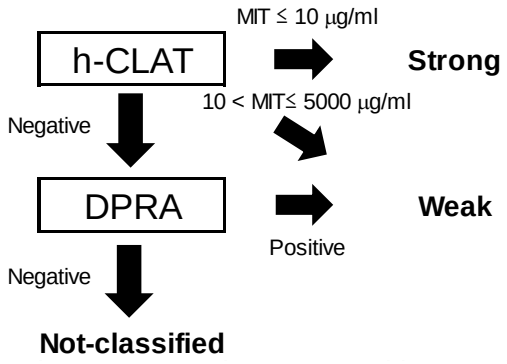


Figure X.1. Sequential Testing Strategy using h-CLAT and DPRA for potency prediction.

8. Chemicals used to develop and test the approach

8.1 Availability of training and test sets:

The training set of 139 chemicals is published in J Appl Toxicol and can be accessed through this link http://onlinelibrary.wiley.com/doi/10.1002/jat.3127/full.

8.2 Selection of the training set and test set used to assess the approach:

The training set were selected based on the availability of good quality animal data, coverage of a range of sensitising potency observed in LLNA, a variety ofcontains the physic-chemical and chemical structural properties, and commercial availability of test substances.

8.3 Supporting information on the training and test sets:

None

8.4 Other information on the training and test sets:

None

9. Limitations in the application of the approach

The strengths and limitations on individual test methods are described in the individual data sources (see Annex II). Chemicals that fall outside the applicability domains of the DPRA and h-CLAT cannot be applicable to the STS.

Technical limitations:

- Low water soluble chemicals with log Kow > 3.5. For the DPRA, test chemicals should be soluble in an appropriate solvent such as acetonitrile or water. For the h-CLAT, test chemicals should be soluble or form a stable dispersion in DMSO or saline.

Substance related limitations:

- Pre- and pro-haptens might not be reliably predicted due to lack of metabolic capacities in both the DPRA and h-CLAT.

When information from the different individual data sources is integrated in the STS, the individual limitation can be minimized and the STS can lead to correct classification of pre- /pro-haptens and low water soluble chemicals.

10. Predictive capacity of the defined approach

For prediction of sensitising potential based on the STS, the sensitivity, specificity, and accuracy for hazard identification are 90%, 54%, and 81%, respectively, for the collection of 139 chemicals (Table X.2a).

The STS categorised 19 of 29 extreme and strong sensitisers (EC3<1%) in the strong class and 57 of 73 moderate and weak sensitisers (EC3≥1%) in the weak class (Table X.2b). As for the three-rank classification by the STS, the over-prediction rate, under-prediction rate, and overall accuracy were 17%, 14%, and 69%, respectively.

When the analysis is repeated while excluding the negative results for 10 chemicals (8 sensitisers and 2 non-sensitisers) with log Kow > 3.5, the sensitivity, specificity, and accuracy of the STS were 98% (92 of 94 sensitisers), 51% (18 of 35 non-sensitisers), and 85% (111 of 129 chemicals), respectively. With these exclusions, the three-rank classification achieved overall accuracies (compared to LLNA) of 73% (94 of 129 chemicals) for the STS.

Table X.2. Predictive performance of Sequential Testing Strategy (STS) in determining sensitising (a) potential and (b) potency.

(a) Potential

Hazard identification		STS	
		Positive	Negative
LLNA	102 sensitizers	92	10
	37 non-sensitizer	17	20
	Sensitivity (%)	90 (92/102)	
	Specificity (%)	54 (20/37)	
	Accuracy (%)	81 (112/139)	

(b) Potency

Potency identification		STS		
		Strong	Weak	Not classified
LLNA	Extreme/Strong	19	10	0
	Moderate/Weak	6	57	10
	Non-sensitizer	0	17	20
	Over prediction rate (%)	17 (23/139)		
	Under prediction rate (%)	14 (20/139)		
	Accuracy (%)	69 (96/139)		

11. Consideration of uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

1. **The DIP's structure,**

- KE 4 is not included due to lack of available tests.
- The Sequential Testing Strategy (STS) covers KE 1 and 3 of AOP and is based on a dataset of 139 chemicals. The confidence in the prediction for hazard identification is high, when similar chemicals are available in this data set and the limitations are taken into account.
- The confidence is lower for chemicals with log Kow > 3.5.
- The confidence is lower for prehaptens and prohaptens due to limited metabolic capacities of test methods.

2. **The information sources used within the defined approach**

- DPRA and h-CLAT has been validated under the ECVAM. Reproducibility of peptide reactivity and CD86/CD54 measurements are very high.

3. Benchmark data used

- The Sequential Testing Strategy (STS) for hazard identification is based on the data from LLNA. The variability of EC3 values of LLNA has been reported depending on vehicle used and laboratories. Therefore, the uncertainty in misprediction of EC3 values is taken into account.

11.2 Impact of uncertainty on the DIP's prediction

Some uncertainty might cause to under- or over-estimation of hazard identification and potency classification for the Sequential Testing Strategy (STS). For a new test chemical, the similar chemicals with *in vitro* or *in vivo* data should be checked. Confidence in the predictions is high if the limitations are taken into account and similar chemicals are included in training set.

12. References

Ade N, Leon F, Pallardy M, Peiffer JL, Kerdine-Romer S, Tissier MH, Bonnet PA, Fabre I, Ourlin JC. (2009). HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: role of the Keap1/Nrf2 pathway. Toxicol Sci, 107:451-460.

Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y, Miyazawa M, Ito Y, Nishiyama N, Itagaki H. (2010). A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). Altern Lab Anim, 38:275-84.

Langton K, Patlewicz GY, Long A, Marchant CA, Basketter DA. (2006). Structure-activity relationships for skin sensitization: recent improvements to Derek for Windows. Contact Dermatitis, 55:342-7.

Migdal C, Botton J, El Ali Z, Azoury ME, Guldemann J, Giménez-Arnau E, Lepoittevin JP, Kerdine-Römer S, Pallardy M. (2013). Reactivity of chemical sensitizers toward amino acids in cellulo plays a role in the activation of the Nrf2-ARE pathway in human monocyte dendritic cells and the THP-1 cell line. Toxicol Sci, 133:259-274.

Nukada Y, Ashikaga T, Miyazawa M, Hirota M, Sakaguchi H, Sasa H, Nishiyama N. (2012). Prediction of skin sensitization potency of chemicals by human Cell Line Activation Test (h-CLAT) and an attempt at classifying skin sensitization potency. Toxicol In Vitro, 26:1150-60.

Nukada Y, Miyazawa M, Kazutoshi S, Sakaguchi H, Nishiyama N. (2013). Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicol In Vitro, 27:609-18.

Takenouchi O, Fukui S, Okamoto K, Kurotani S, Imai N, Fujishiro M, Kyotani D, Kato Y, Kasahara T, Fujita M, Toyoda A, Sekiya D, Watanabe S, Seto H, Hirota M, Ashikaga T, Miyazawa M. (2015). Test battery with h-CLAT, DPRA, and DEREK based on a 139 chemical dataset for predicting skin sensitizing potential and potency of chemicals. J Appl Toxicol, 35:1318-1332.

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, 71:337-351.

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

OECD, (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins Part 2: Use of the AOP to Develop Chemical and Integrated Assessment and Testing Approaches. Series on Testing and Assessment No. 168. ENV/JM/MONO(2012)10/PART2 http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part2&doclanguage=en

13. Supporting information

None

14. Abbreviations and definitions

AOP: Adverse outcome pathway

ARE: antioxidant response element

DC: Dendritic cell

DIP: Data interpretation procedure

DPRA: Direct peptide reactivity assay

EC3: estimated concentration of a test substance needed to produce a stimulation index of three in LLNA, the threshold value for the substance to be considered a sensitiser

GHS: Globally harmonised System for Classification and Labelling

h-CLAT: human Cell Line Activation Test

MIT: Minimal induction threshold

Nrf2: nuclear factor-erythroid 2-related factor 2

STS: Sequential testing strategy

LLNA: Local lymph node assay

CASE STUDY XI

Integrated testing strategy (ITS) for sensitising potency classification based on *in silico*, *in chemico*, and *in vitro* data

1. Summary

The defined approach presented in this document is constructed as an integrated testing strategy (ITS) for prediction of the skin sensitisation potential and potency of a substance. The test methods used include in silico prediction (DEREK Nexus) and address two key events (KEs) 1 and 3 as defined in OECD Adverse Outcome Pathway (AOP) of skin sensitisation: KE 1 of protein binding is evaluated using the Direct Peptide Reactivity Assay (DPRA; OECD TG 442c); KE 3 of dendritic cell activation is evaluated using the human cell line activation test (h-CLAT). This ITS described here is based on 139 chemicals classified a Sensitisers / Non-Sensitisers (S/NS) in the LLNA. DEREK Nexus predicts the probability that a substance will be a Sensitisers / Non-Sensitisers (S/NS) by an alert. The peptide depletion in the DPRA and the quantitative dose-response outcome in the h-CLAT do correlate to sensitising potency based on the EC3 values in the LLNA. The alert in DEREK Nexus and the quantitative outcomes in the DPRA and h-CLAT are converted to a score of 0 to 3. The summed score of three test methods can be used to predict the skin sensitising potential (hazard identification; S/NS) and potency of a substance. The potency prediction is given as three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS), but is not used for GHS sub-classification into sub-category 1A (EC3<2% in LLNA) and 1B (EC3≥2%).

2. General information

2.1 Identifier:

Sensitising potency classification based on AOP KEs 1 and 3

2.2 Date: 3 May2016

2.3 Author(s) and contact details:

Masaaki Miyazawa, Ph.D., R&D – Core Technology – Safety Science Research, Kao Corporation, 2606 Akabane, Ichikai-machi, Haga-gun, Tochigi 321-3497, Japan miyazawa.masaaki@kao.co.jp; Tel +81-285-68-7342

2.4 Template update(s): *Version 1*

2.5 Reference to main scientific papers:

Key paper;

Takenouchi et al., (2015) Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals, J Appl Toxicol 35, 1318-1332.

Nukada et al., (2013) Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicol in Vitro 27, 609-618.

2.6 Proprietary aspects:

A license agreement is needed for Derek Nexus, which is commercially available software from Lhasa Limited.

3. Endpoint addressed

3.1 Endpoint:

This ITS predicts skin sensitisation potential (hazard identification: S/NS) and potency (three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS)), based on readouts from assays addressing AOP KEs 1 and 3 and in silico prediction. Thus, it is addressing the endpoint of existing test guidelines e.g. Local Lymph Node Assay (LLNA) (OECD TG 429), Guinea Pig Maximisation Test (GPMT) and Buehler Test (OECD TG 406).

3.2 Species:

Mouse (primary target), Human (secondary target).

3.3 Additional information about the endpoint:

Skin sensitisation is the result of a complex multifactorial sequence of events. There are four KEs in the chemical and biological pathways (protein binding, keratinocyte activation, dendritic cell activation, and proliferation of antigen-specific T-cells) as the AOP for skin sensitisation defined by the OECD (2012). This ITS uses data from in silico, in chemico, and, in vitro assays addressing the intrinsic chemical reactivity (KE 1) and capacity to induce dendritic cell activation (KE 3).

4. Definition of the purpose and regulatory relevance

This ITS was developed to predict the skin sensitisation potential (Sensitisers or Non-Sensitisers) of a substance for the purpose of classification and labelling under the GHS scheme without animal testing. The potency prediction of this ITS will be useful to obtain three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS), but is not used for GHS subclassification into subcategory 1A (EC3<2% in LLNA) and 1B (EC3≥2%).

5. Rationale underlying the construction of the defined approach

Based on the adverse outcome pathway of skin sensitisation defined by OECD, the molecular initiating event (KE 1) and the cellular response of dendritic cells (KE 3) are taken into account in this ITS.

The molecular initiating event (KE 1) leading to skin sensitisation is postulated to be covalent binding of electrophilic chemical species to selected nucleophilic molecular sites of action in skin proteins. The covalent binding to skin proteins is evaluated using the DEREK Nexus and the Direct Peptide Reactivity Assay (DPRA), which is now the OECD TG 442C.

The activation of dendritic cells (DC) is typically assessed by expression of specific cell surface markers, chemokines and cytokines. The h-CLAT is proposed to address the KE 3 (dendritic cell activation) of the skin sensitisation AOP and isOECD TG 442E.

In this ITS, the assay related to KE 2 is not included, but DPRA cysteine depletion (KE 1) and KeratinoSens™ covering KE 2 are mechanistically relevant (Joanna et al., 2013). The key molecular pathway (Nrf2-ARE pathway) induced in KeratinoSens™ corresponds to cysteine reactivity with the Keap1 sensor protein. In addition, the Nrf2 activation is induced by sensitisers and not by non-sensitisers in THP-1 cells, and could function as one of the danger signals to lead to the phenotypic alterations on THP-1 cells (Migdal et al., 2013; Ade et al., 2009). Thus, there is a mechanistic rationale that DPRA and h-CLAT could be linked to KeratinoSens™ (KE 2).

In this ITS, the outcomes or quantitative parameters in each of the individual test methods are assigned to scores, by modifying the weight of evidence approach proposed by Jowsey et al. (2006) and Natsch et al. (2009) in order to define a sensitising potential (hazard identification; sensitisers vs non-sensitisers) and potency (three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS)) of a substance.

The underlying rationale of this ITS is that either a medium score (2) in the individual test (i.e., DPRA or h-CLAT) or a low score (1) in two test methods out of three is considered enough evidence for judging a substance as a sensitiser.

6. Description of the individual information sources used

1. DEREK Nexus: in silico knowledge-based toxicity alerting software comprising alerts on skin sensitisation (version 2.0 from Lhasa Limited).

DEREK Nexus is mainly addressing structural features and whether a hapten has a potential for electrophilic binding to skin proteins either directly or following metabolism (Langton et al., 2006). To each alert. a certainty level is associated. Substances with causative structural alert(s) (i.e., certain, probable, plausible, equivocal, and doubted) are conservatively considered to be a potential sensitiser.

2. DPRA (OECD TG442c) is addressing the peptide binding. Haptens applied to the skin are covalently binding to nucleophilic residues (i.e. cysteine, lysine) in dermal proteins. Binding of chemicals to the skin protein is an essential step for sensitiser to obtain allergenicity (OECD, 2012).

Substances that induced mean peptide depletion of cysteine- and lysine-containing peptide above 6.38% are considered to have peptide reactivity of sensitiser. Substances predicted as sensitiser are defined to have potential to be strong or weak sensitiser based on threshold of 22.62% mean peptide depletion.

3. h-CLAT is addressing DC activation. When a hapten is applied to the skin, surface molecules (i.e. CD54, CD86) on skin DCs were up-regulated through the maturation process. Since CD54 is involved in DC migration to draining lymph nodes and CD86 stimulates T cell activation during antigen-presentation by DC, both molecules are essential in the induction of skin sensitisation.

Substances inducing a fold induction greater than 2-fold for CD54 and/or 1.5-fold increase for CD86 at cell viabilities above 50% are predicted to have a DC activating potential of sensitiser (Ashikaga et al., 2010). From the dose-dependency curves of experiments, the median concentration(s) inducing 1.5- and/or 2-fold induction of CD86 and/or CD54 are calculated and the resulting lower value is defined as minimal induction threshold, MIT. Substances predicted as positive are defined to have potential to be strong or weak sensitiser based on a threshold of 10 µg/mL of MIT (Nukada et al., 2013).

7. Data interpretation procedure (DIP) applied

The quantitative parameters or outcomes of the individual test methods are assigned to scores, by modifying the weight of evidence approach proposed by Jowsey et al. (2006) and Natsch et al. (2009) in order to define a sensitising potential and potency of a substance. The quantitative parameters of h-CLAT and DPRA are converted into a score from 0 to 3 as shown Table XI.1. The thresholds for the scores from 0 to 3 were set in order to span the whole dynamic range on the individual assays and were also derived from the values needed for significant results. For h-CLAT, the minimum induction thresholds (MITs) are converted to a score from 0 to 3 based on the cutoffs of 10 and 150 µg/ml. For DPRA, the mean percent depletion for the cysteine and lysine peptides is converted to a score from 0 to 3, based on OECD TG 442C. In cases where co-elution occurs only with the lysine peptide, the depletion for only cysteine peptides is converted to a score from 0 to 3. For DEREK Nexus, an alert is assigned a score of 1; absence of an alert was assigned a score of 0. Having only an alert outcome is regarded as not sufficient evidence to predict a test substance as a sensitiser. When the sum of these scores have been assessed, a total battery score from 0 to 7, calculated by summing the individual scores, is used to predict the sensitising potential (hazard identification; sensitisers vs non-sensitisers) and potency (three rank classes: EC3<1% in LLNA (Strong), EC3≥1% (Weak), Non-Sensitisers (NS)). The positive criteria are set as a total battery score of 2 or greater. Furthermore, a total battery score is classified into three ranks: score of 7 is defined as a strong sensitiser; score of 6, 5, 4, 3, or 2, weak sensitiser; score of 1 or 0, not-classified. The summed score yields a qualitative result (positive/negative and three rank classes).

Table XI.1. Conversion of the outcome in h-CLAT, DPRA, and DEREK for Integrated Testing Strategy

Score	h-CLAT MIT (µg/ml)	DPRA Depletion (%) (Cysteine-only)	DEREK Alert
0	Not calculated (Negative)	<6.376 (<13.89)	No alert
1	>150, ≤5000	≥6.376, <22.62 (≥13.89, <23.09)	Alert
2	>10, ≤150	≥22.62, <42.47 (≥23.09, <98.24)	-
3	≤10	≥42.47 (≥98.24)	-

8. Chemicals used to develop and test the approach

8.1 Availability of training and test sets:

The training set of 139 chemicals ispublished in J Appl Toxicol and can be accessed through this link http://onlinelibrary.wiley.com/doi/10.1002/jat.3127/full.

8.2 Selection of the training set and test set used to assess the approach:

The training set were selected based on the availability of good quality animal data, coverage of a range of sensitising potency observed in LLNA, a variety of the physic-chemical and chemical structural properties, and commercial availability of test substances.

8.3 Supporting information on the training and test sets:

None

8.4 Other information on the training and test sets:

None

9. Limitations in the application of the approach

The strengths and limitations on individual test methods are described in the individual data sources (see Annex II). Chemicals that fall outside the applicability domains of the DPRA and h-CLAT cannot be applicable to the ITS.

Technical limitations:

- Low water soluble chemicals with log Kow > 3.5. For the DPRA, test chemicals should be soluble in an appropriate solvent such as acetonitrile or water. For the h-CLAT, test chemicals should be soluble or form a stable dispersion in DMSO or saline.

Substance related limitations:

- Pre- and pro-haptens might not be reliably predicted due to lack of metabolic capacities in both the DPRA and h-CLAT.

When information from the different individual data sources is integrated in the ITS, the individual limitation can be minimized and the ITS can lead to correct classification of pre- /pro-haptens and low water soluble chemicals.

10. Predictive capacity of the defined approach

For prediction of sensitising potential based on the ITS, the sensitivity, specificity, and accuracy were 89%, 70%, and 84%, respectively, for the collection of 139 chemicals (Table XI.2a).

The ITS categorised 15 of 29 extreme and strong sensitisers (EC3<1%) in the strong class and 57 of 73 moderate and weak sensitisers (EC3≥1%) in the weak class (Table XI.2b). As for the three-rank classification by the ITS, the over-prediction rate, under-prediction rate, and overall accuracy were 12%, 18%, and 71%, respectively.

When the analysis was repeated while excluding the negative results for 11 chemicals (8 sensitisers and 3 non-sensitisers) with log Kow > 3.5, the sensitivity, specificity, and accuracy of the ITS were 97% (91 of 94 sensitisers), 68% (23 of 34 non-sensitisers), and 89% (114 of 128 chemicals), respectively. With these exclusions, the three-rank classification achieved overall accuracies (compared to LLNA) of 74% (95 of 128 chemicals) for the ITS.

Table XI.2. Predictive performance of Integrated Testing Strategy (ITS) in determining sensitising (a) potential and (b) potency.

(a) Potential

Hazard identification		ITS	
		Positive	Negative
LLNA	102 sensitizers	91	11
	37 non-sensitizer	11	26
	Sensitivity (%)	89 (91/102)	
	Specificity (%)	70 (26/37)	
	Accuracy (%)	84 (117/139)	

(b) Potency

Potency identification		ITS		
		Strong	Weak	Not classified
LLNA	Extreme/Strong	15	14	0
	Moderate/Weak	5	57	11
	Non-sensitizer	0	11	26
	Over prediction rate (%)	12 (16/139)		
	Under prediction rate (%)	18 (25/139)		
	Accuracy (%)	71 (98/139)		

11. Consideration of uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

1. **The DIP's structure**

- KE 4 is not included due to lack of available tests.
- The Integrated Testing Strategy (ITS) covers KE 1 and 3 of AOP and is based on a dataset of 139 chemicals. The confidence in the prediction for hazard identification is high, when similar chemicals are available in this data set and the limitations are taken into account.
- The confidence is lower for chemicals with log Kow > 3.5.
- The confidence is lower for prehaptens and prohaptens due to limited metabolic capacities of test methods.

2. **The information sources used within the defined approach**

- The DPRA and h-CLAT has been validated under the ECVAM. Reproducibility of peptide reactivity and CD86/CD54 measurements are very high.

3. Benchmark data used

- The Integrated Testing Strategy (ITS) for hazard identification is based on the data from LLNA. The variability of EC3 values of LLNA has been reported depending on vehicle used and laboratories. Therefore, the uncertainty in misprediction of EC3 values is taken into account.

11.2 Impact of uncertainty on the DIP's prediction

Some uncertainty might cause to under- or over-estimation of hazard identification and potency classification for the Integrated Testing Strategy (ITS). For a new test chemical, the similar chemicals with in vitro or in vivo data should be checked. Confidence in the predictions is high if the limitations are taken into account and similar chemicals are included in training set.

12. References

Ade N, Leon F, Pallardy M, Peiffer JL, Kerdine-Romer S, Tissier MH, Bonnet PA, Fabre I, Ourlin JC. HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: role of the Keap1/Nrf2 pathway. Toxicol Sci. 2009. 107:451-460.

Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y, Miyazawa M, Ito Y, Nishiyama N, Itagaki H. A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). Altern Lab Anim. 2010. 38:275-84.

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

Jowsey IR, Basketter DA, Westmoreland C, Kimber I. A future approach to measuring relative skin sensitising potency: a proposal. J Appl Toxicol. 2006. 26:341-50.

Langton K, Patlewicz GY, Long A, Marchant CA, Basketter DA. Structure-activity relationships for skin sensitization: recent improvements to Derek for Windows. Contact Dermatitis. 2006. 55:342-7.

Migdal C, Botton J, El Ali Z, Azoury ME, Guldemann J, Giménez-Arnau E, Lepoittevin JP, Kerdine-Römer S, Pallardy M. Reactivity of chemical sensitizers toward amino acids in cellulo plays a role in the activation of the Nrf2-ARE pathway in human monocyte dendritic cells and the THP-1 cell line. Toxicol Sci. 2013. 133:259-274.

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

Nukada Y, Miyazawa M, Kazutoshi S, Sakaguchi H, Nishiyama N. Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicol In Vitro. 2013. 27:609-18.

Takenouchi O, Fukui S, Okamoto K, Kurotani S, Imai N, Fujishiro M, Kyotani D, Kato Y, Kasahara T, Fujita M, Toyoda A, Sekiya D, Watanabe S, Seto H, Hirota M, Ashikaga T, Miyazawa M. Test battery with h-CLAT, DPRA, and DEREK based on a 139 chemical dataset for predicting skin sensitizing potential and potency of chemicals. J Appl Toxicol. 2015. 35:1318-1332.

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 ENV/JM/MONO(2012)10/PART1. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en

OECD, (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins Part 2: Use of the AOP to Develop Chemical and Integrated Assessment and Testing Approaches. Series on Testing and Assessment No. 168. ENV/JM/MONO(2012)10/PART2 http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part2&doclanguage=en

13. Supporting information

None

14. Abbreviations and definitions

AOP: Adverse outcome pathway

ARE: antioxidant response element

DC: Dendritic cell

DIP: Data interpretation procedure

DPRA: Direct peptide reactivity assay

EC3: estimated concentration of a test substance needed to produce a stimulation index of three in LLNA, the threshold value for the substance to be considered a sensitiser

GHS: Globally harmonised System for Classification and Labelling

h-CLAT: human Cell Line Activation Test

ITS: lntegrated testing strategy

MIT: Minimal induction threshold

Nrf2: nuclear factor-erythroid 2-related factor 2

LLNA: Local lymph node assay

CASE STUDY XII

DIP for Skin Allergy Risk Assessment (SARA)

1. Summary

Our objective was to build a chemically and biologically plausible model of skin sensitisation induction in humans, using 2,4-Dinitrochlorobenzene (DNCB) as our initial case study. An ordinary differential equation (ODE) approach was employed to model the underlying chemical and biological processes of mass transport, reaction kinetics, Dendritic cell (DC) Class I MHC antigen presentation pathways and T cell receptor (TcR) binding events. The SARA model predicts the rate of naïve $CD8^+$ T cell receptor triggering (component of AOP step 7, KE 4) following a given skin exposure to a chemical, which we hypothesise can be thresholded to inform a prediction of the likelihood of skin sensitisation induction/allergic contact dermatitis in humans (AOP step 11, adverse outcome).

2. General information

2.1 Identifier:

DIP for Skin Allergy Risk Assessment (SARA)

2.2 Date:

22 April 2016

2.3 Author(s) and contact details:

Gavin Maxwell, Nikki Gellatly, Richard Cubberley, Ruth Pendlington, Juliette Pickles, Joe Reynolds and Cameron MacKay; Unilever - Safety Environmental Assurance Centre (SEAC), Colworth Science Park, Sharnbrook, Bedford, UK, MK44 1LQ

Corresponding author: Gavin Maxwell (gavin.maxwell@unilever.com: +44-(0)-1234-264-888)

2.4 Template update(s): Version 1

2.5 Reference to main scientific papers:

Davies M, Pendlington RU, Page L, et al. (2011). Determining epidermal disposition kinetics for use in an integrated nonanimal approach to skin sensitization risk assessment. Toxicological Sciences, 119: p.308–18.

MacKay C, Davies M, Summerfield V and Maxwell G. (2013). From Pathways to People: Applying the Adverse Outcome Pathway (AOP) for Skin Sensitization to Risk Assessment. ALTEX. 30, p.473-486.

Reynolds J, Pickles J, Cubberley R. et al. (2016). Toxicokinetic/Toxicodynamic (TK/TD) modelling of skin sensitisation. Part I: Toxicokinetics. Tox. Sci, Manuscript in Preparation.

MacKay C, Reynolds J, Dhadra S. et al. (2016). Toxicokinetic/Toxicodynamic (TK/TD) modelling of skin sensitisation. Part II: Toxicodynamics and hazard characterisation. Tox. Sci, Manuscript in Preparation

2.6 Proprietary aspects:

The SARA model is proprietary at present, however full disclosure manuscripts are in the process of being prepared for publication (i.e. MacKay et al. 2016; Reynolds et al. 2016).

3. Endpoint addressed

3.1 Endpoint:

Probability of human skin sensitisation to a given exposure of chemical (e.g. outcome of Human Repeat Insult Patch Test (HRIPT)).

3.2 Species: Human

3.3 Additional information about the endpoint:

The chemical and biological mechanisms represented within the SARA model have been qualitatively described within the recent OECD report 'The Adverse Outcome Pathway for Skin Sensitisation initiated by Covalent Binding to Proteins. Part 1. Scientific Evidence' (2012). Quantitative data to enable mathematical modelling of these mechanisms have been obtained from the published literature; in many cases assumptions have been made that the biological mechanisms underpinning human sensitiser-induced immune responses are analogous to those characterised for pathogen-driven immune responses.

4. Definition of the purpose and regulatory relevance

The purpose of the SARA model is to inform human health risk assessment decisions through direct prediction of the probability of an individual becoming allergic following a given skin exposure to a sensitising chemical to underpin the marketing of a new consumer product. Uncertainty in the SARA model prediction is explicit and therefore can be used to highlight the need for additional information generation (i.e. to refine or benchmark the model prediction) and/or risk management measures (i.e. where information generation is infeasible or impractical).

5. Rationale underlying the construction of the defined approach

The proposed model is described according to three major events leading to sensitisation induction in an individual following DNCB exposure: distribution in skin and formation of hapten-protein complexes; antigen processing and presentation by skin-derived dendritic cells; and activation of naïve

hapten-specific $CD8^+$ T cells. Our model predicts the rate of naïve $CD8^+$ T cell receptor triggering (component of AOP step 7, KE 4 in OECD-defined AOP), which

The SARA model is a qualitative (mechanistic) and quantitative explanation of the induction of naïve $CD8^+$ T cell response in humans, which encompasses what we regard as the most relevant events of the skin sensitisation AOP for making a quantitative prediction. The model has been constructed to predict the following events over time:

- diffusion/partitioning of sensitising chemicals within human skin (AOP step 1)
- haptenation of protein within viable layers of skin (AOP steps 3 and 4, KE 1)
- Class I MHC processing & presentation of haptenated skin protein by Dendritic cells (DC) (component of AOP step 6, KE 3)
- extent of human naïve $CD8^+$ T cell activation following interaction with dendritic cells presenting haptenated peptides in the context of Class I MHC (component of AOP step 7, KE 4)

We hypothesise that a metric derived from modelling an aspect of KE 4 (rate of naïve $CD8^+$ TcR triggering) can be thresholded to inform a prediction of the likelihood of skin sensitisation induction/allergic contact dermatitis elicitation in humans (AOP step 11 and defined Adverse Outcome).

These chemical and biological pathways represent a subset of the KEs described within the Skin Sensitisation AOP (see Figure XII.1). They were judged appropriate for representing the simplest route to predicting the probability that a specific CD8+ T cell response will be induced following a given skin exposure to a direct-acting sensitising chemical (such as DNCB).

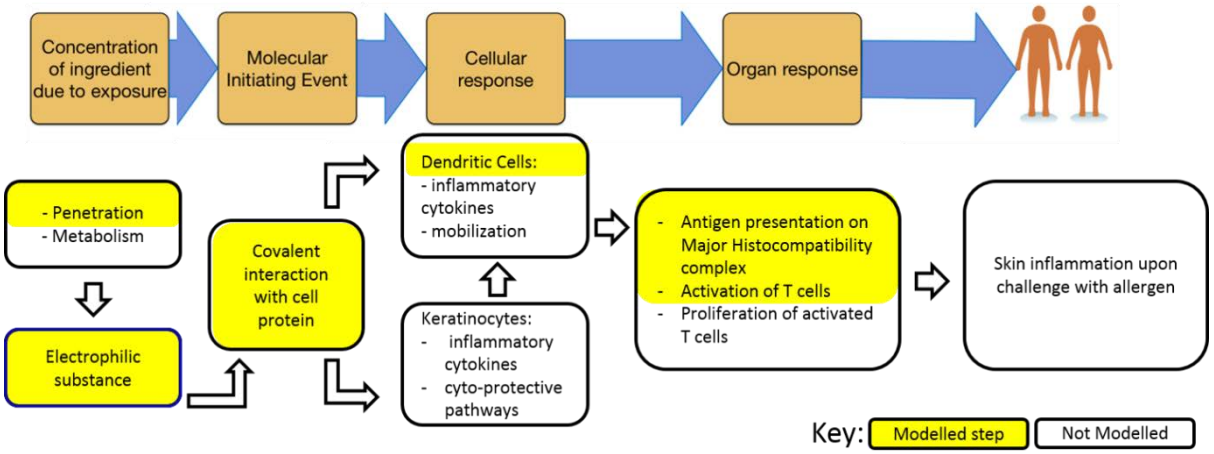


Figure XII.1: SARA model scope relative to Skin Sensitisation AOP

Under these assumptions, the following AOP steps/KEs are excluded:

- Keratinocyte and DC activation (components of steps 5 and 6, KE 3) – assumed to be sufficient to enable optimal co-stimulation of naïve $CD8^+$ T cell, on the basis that DC will undergo activation/maturation upon direct exposure to DNCB

- DC migration (component of step 7, KE 4) – assumed to be sufficient to ensure an antigen presentation event can take place in the draining lymph node on the basis that DC migration will follow direct exposure to DNCB

These assumptions may or may not be judged as reasonable, however they are designed to be conservative as all steps not explicitly modelled are assumed to present and sufficient. Furthermore we are aware that multiple other pathways (e.g. skin metabolism, other innate immune processes and other T cell subtypes) not modelled here also play a role in the induction of skin sensitisation and that there is an intrinsic level of variability in the overall system. Consequently, we have performed uncertainty analysis on the model to determine to what extent we believe that we have accurately captured the key chemical and biological events and the underlying variability in the system with our current model structure and parameter set. This uncertainty analysis enables the model output to be visualised as a prediction with defined credible intervals to ensure that model-informed risk assessment decisions are made in the context of this uncertainty. In addition, case-by-case re-assessment of the model scope and parameterisation is necessary to ensure the SARA model is fit for purpose for a given application.

6. Description of the individual information sources used (see Annex II)

SARA model parameters essentially fall into three categories: exposure, chemical-specific and biological. The chemical-specific parameters are related to diffusion, partitioning and reactivity and are determined by the information sources and data interpretation procedures specified below, with remaining parameters taken from sensitiser-specific data where available. The exposure parameters (dose applied and area exposed) for our initial DNCB case study are taken from literature on the clinical study under investigation (Friedmann et al., 1983). However, in general these would be defined by the exposure scenario of the risk assessment under investigation. Biological parameters, e.g. 'rate of protein turnover' and 'number of MHCI presentable peptides per protein' are taken either directly from published literature or derived from bio-informatic analysis of publically available datasets. Human-relevant data are used wherever possible. Where human data are not available, data obtained using other species are considered. A global sensitivity analysis has been performed for all model parameters to determine how sensitive the uncertainty in model output is to their contribution.

Chemical-specific parameters are obtained from experimental data on the kinetics of skin bioavailability and protein haptenation are obtained via a modification to OECD Skin Penetration test guideline 428 to include additional termination time points and measurement of free and bound test item 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.

Parameters are inferred from the 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, Richardson and Spiegelhalter, 1996) with 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.

The SARA model prediction can be informed by other types of skin bioavailability and protein reactivity input data (e.g. in silico predictions, in vitro or in vivo data) however this will have consequences for model prediction uncertainty. At present the impact of different types of input data on the overall model prediction has not yet been fully assessed, however this represents an important area for future research.

7. Data interpretation procedure applied

The SARA model structure can be summarised using the following schematic (see Figure XII.2) or as follows: a compartmental approach is applied for modelling mass transfer of chemical by Fickian diffusion from an applied vehicle through skin (previously described in Davies et al., 2011). Vehicle, stratum corneum and viable skin (epidermis and dermis) compartments are all explicitly represented with viable skin assumed to be the target site for formation of hapten-protein complexes. Chemical loss can occur through evaporation from vehicle, loss from skin to blood or loss from skin through haptenation of proteins. We hypothesise that skin sensitisation is predominantly driven by $CD8^+$ T cells and therefore consider only antigen processing and presentation via Class I MHC pathways within DC and assume that the resulting hapten-peptide complexes presented on Class I MHC are equally immunogenic. The formation of the hapten-protein complex can occur at nucleophilic amino acid residues and the N-terminus of the protein and we assume only one modification per binding site is possible. Cytosolic protein degradation by the proteasome is considered to be our primary pathway for generation of peptide fragments for presentation on Class I MHC. This results in a pool of peptides (both haptenated and non-haptenated) that can bind to Class I MHC and be presented on the DC surface. To calculate the TcR triggering rate we adopt the confinement time hypothesis (reviewed in Zarnitsyna and Zhu, 2012) and consider the signals exchanged during a DC: T cell kinapse interaction to simulate the impact of peptide-MHC complex densities on extent of naïve $CD8^+$ TcR triggering.

Our SARA model predicts human, naïve $CD8^+$ T cell activation as a surrogate metric for sensitisation induction in humans. A full dose-response (chemical dose vs. rate of human, naïve $CD8^+$ T cell receptor triggering) is predicted from which a point of departure is derived for use in risk assessment decision making.

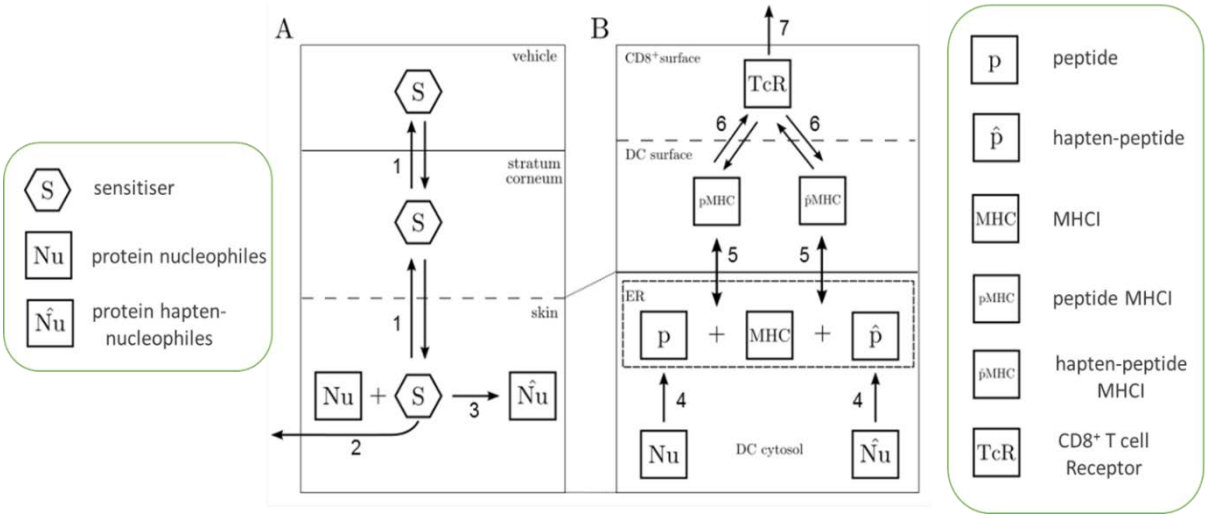


Figure XII.2: SARA model schematic: **Panel A**; Toxicokinetic model (described in Reynolds, 2016): (1) diffusion and partitioning into the stratum corneum and skin; (2) sensitiser clearance by dermal capillaries; (3) covalent modification of protein nucleophiles by hapten; **Panel B**; Toxicodynamic model (described in MacKay, 2016): **(4)** proteasome processing of protein nucleophiles to form small peptides and transport to the endoplasmic reticulum (ER); **(5)** binding of peptides and hapten-peptide complexes to Class I MHC and transport to plasma membrane; **(6)** binding of pMHC and hapten-pMHC to $CD8^+$ T cell receptors and **(7)** activation of naïve specific $CD8^+$ T cells.

The chemical and biological mechanisms that are considered relevant are encoded in a mathematical model that consists of both ODEs and model parameters. The structure of the ODEs prescribe, qualitatively, the underlying kinetic and dynamic behaviour of the KEs and their causal relations. Specification of the parameter values and solution of the ODE system yields a quantitative description of the toxicokinetic and toxicodynamic behaviour of the system. The parameterisation of the model reflects the specifics of the exposure scenario, the subsequent interaction of the chemical with the biological system (e.g. reaction with proteins, diffusion, partitioning) and the subsequent biological response to the chemical (e.g. processing and presentation of haptenated proteins and naïve CD8+ T cell activation). Significant literature review was performed for each of these events which extended beyond the sensitiser-specific literature. Although not formal or prescriptive, this should be interpreted as a weight-of-evidence approach to determining model structure and parameterisation.

Following model construction and model parameterisation, the ODEs are solved in order to provide a quantitative description of the kinetics and dynamics of the chemical and biological pathways of the system. The ODEs are solved with 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).

8. Chemicals used to develop and test the DIP

8.1 Availability of training and test sets:

For our initial SARA model case study we have simulated the probability of skin sensitisation/allergic contact dermatitis in humans following a single $7.1cm^2$ volar forearm exposure for 48hrs under

occlusion to the experimental sensitiser DNCB. To do so, we have benchmarked the model output against a previously published study by Friedmann *et al.* (1983) in which 165 healthy volunteers were exposed to one of four doses of DNCB and their allergic status checked by diagnostic patch test applied 4 weeks later. Further published datasets describing the induction and incidence of allergic contact dermatitis following exposure to different experimental and consumer-relevant chemical sensitisers (e.g. HRIPT data) will be used to continue the evaluation the SARA model.

8.2 Selection of the training set and test set used to assess the DIP:

Literature data were evaluated to identify human clinical benchmark data for chemical sensitisers where the exposure scenario had been clearly defined and sensitisation threshold had been measured (e.g. HRIPT or diagnostic patch test data). DNCB was selected as our initial case study chemical due to the existence of a large clinical dataset meeting the aforementioned criteria. A further benefit of DNCB is the large amount of historical research performed to delineate its mechanism of action as a skin sensitiser.

8.3 Supporting information on the training and test sets:

The SARA model is proprietary at present, however full disclosure manuscripts are in the process of being prepared for publication (i.e. MacKay et al. 2016; Reynolds et al. 2016) that fully describe the model (i.e. ODEs) and how it was parameterised for our initial case study (i.e. DNCB-specific model input data and biological parameter ranges).

8.4 Other information on the training and test sets:

Additional case studies are underway to build confidence in model predictions through benchmarking to relevant clinical datasets (e.g. published human repeat insult patch test (HRIPT) data).

9. Limitations in the application of the defined approach

The biological scope of the SARA model has been set to address chemical sensitisers that induce skin sensitisation/allergic contact dermatitis through the direct haptenation of proteins (i.e. without the need for prior oxidation or metabolism).

At present the chemical sensitiser of interest needs to be radiolabelled to enable model input data (i.e. skin penetration and protein binding kinetic information) to be generated. To address this limitation we are currently exploring whether other forms of skin bioavailability and protein binding information can be used to parameterise the model to address this limitation.

10. Predictive capacity of the defined approach

To date the SARA model has been evaluated against a previously published study by Friedmann et al. 1983 in which 165 healthy volunteers were exposed to one of four doses of the experimental sensitiser DNCB. Uncertainty analysis of the SARA model (in collaboration with John-Paul Gosling, Univ. Leeds – work funded by UK NC3Rs) has enabled us to evaluate and visualise the impact of parameter uncertainty on our model prediction.

Further sensitiser and exposure case studies are required to evaluate the predictive capacity of such an approach, therefore additional case studies, model development and fundamental research are all underway to explore model assumptions, reduce uncertainty in model predictions and build confidence in model predictions through benchmarking to relevant clinical datasets (e.g. published HRIPT data).

11. Consideration of uncertainties associated with the application of the defined approach

11. 1 Sources of uncertainty

The uncertainties relating to structure and weighting arise from the validity of the assumptions that have been made during construction of the SARA model.

Assumptions judged to be major or founding were as follows:

- Extent of naïve CD8+ T-cell receptor (TcR) triggering is the key determinant of human allergic status
- Existence of at least one T-cell specific to the 'antigen' (i.e. Class I MHC peptide derived from haptenated skin protein)
- Required T-cell co-stimulatory signals are sufficient
- Accompanying CD4+ T-cell response is optimal
- DC migration from exposure site is sufficient

Assumptions judged to be minor were as follows:

- That stratum corneum, viable epidermis and dermis are well represented by the ODE approach.
- That transport occurs by passive diffusion driven by a concentration gradient with partitioning modelled by applying Fick's first law.
- That the transport and partitioning are dose independent and thus the model has a linear dependency when extrapolating across the doses under consideration.
- That the only transport processes acting on the ingredient-bound nucleophiles in intracellular space is loss by proteasome processing.
- That the rate of loss of ingredient-bound nucleophiles to proteasome processing is proportional to the concentration of ingredient-bound nucleophiles.
- That, once an ingredient-bound nucleophile is processed by the proteasome, it is either presented on Class I MHC or lost from the system by some other process e.g. excretion.
- That the nucleophiles available on proteins can be treated as well mixed throughout viable skin. That, for the purposes of modelling the reaction kinetics, the available nucleophiles considered are Cysteine, Tyrosine, Lysine, Histidine, Arginine, Methionine and the protein N-terminus.
- The reaction stoichiometry is one haptenation event per nucleophilic binding site.
- The fraction of nucleophilic biding sites are unavailable for binding remains constant over time.

- That the reaction kinetics can be considered concentration dependent and that the reaction rate constant (but not the reaction rate) is concentration independent and time independent in the timescale under consideration.

- That free and ingredient-bound nucleophiles associated with proteasome generated peptides can be treated as a well-mixed homogeneous population for the purposes of describing their dynamics.

- That the pMHC expressed on the surface of exposure-site derived DC is representative of the state peptide pool at the point of leaving the skin.

- That the epitopes of the response are haptenated-nucleophiles within peptides presented on MHCI.

- That the affinity of the interaction between the epitopes of the response and their cognate TCRs can be well represented by an average affinity for TcR. The average affinity (kon/koff) can be conceptualised in a thought experiment where all the distinct epitopes in a population are taken (variation over allele, peptide sequence, number of haptenated nucleophiles and specificity of haptenated nucleophiles) together with the set of TCRs which would respond to these epitopes. In such an experiment, the average affinity for TcR would be obtained by averaging the kon/koff values over all of these interactions.

- That TcR triggering rate is well described by the 'confinement time' hypothesis of Dushek *et al*, 2009.

- That cross-presentation (presentation of LC/DC endocytosed antigen via Class I MHC) does not significantly contribute to the pool of Class I MHC: peptide complexes presented on skin-derived DC.

- That the volume of internal skin exposure to the chemical is given by the surface area exposed multiplied by the depth of the skin (i.e. the chemical does not move laterally).

- That the depth at which diffusion from viable skin to blood occurs is the same as the total skin depth.

- That chemical sensitiser only enters the draining lymph node when bound to protein within a DC.

To date we have evaluated the SARA model using one case study (i.e. simulation of Friedmann *et al*. 1983 exposure to DNCB). Confidence in SARA model results for other chemicals and exposure scenarios will vary depending on:

- how well the chemical mechanism of action has been elucidated

- how closely the hypothesised mechanism of action aligns to our SARA model scope (i.e. relevance of the modelled pathways for that chemical/exposure scenario)

- quality and relevance of the information used to parameterise the model

Model parameters are expressed as distributions in order to reflect accuracy and precision in underlying data sources. Given the parameter distributions, model output is also calculated as a distribution and thus the confidence in the model is quantified.

The target data, human dose response information, is not used to inform the SARA model but instead used to evaluate or benchmark the model prediction. As such, variability in the target data does not affect prediction confidence beyond taking it into account when making the comparison with model simulations.

11.2 Impact of uncertainty on the DIP's prediction

Uncertainty in SARA model predictions for new chemicals will be characterised in the same manner as has been illustrated for the DNCB case study. Model parameters would be expressed as distributions in order to reflect accuracy and precision in underlying data sources specific to the new chemical. Structural, coverage and/or weighting uncertainties would be interrogated by assessing the appropriateness of the structural assumptions for the new chemical under consideration. Model predictions of human, naïve $CD8^+$ T cell activation, would be provided as a full dose-response with credible intervals indicating the uncertainty in model prediction.

As an example of this analysis, the SARA model structure assumptions (documented above) were used to identify potential sources of uncertainty in the prediction of naïve $CD8^+$ TcR triggering rate for the DNCB case study. All potential sources of uncertainty were consolidated and evidence for and against each source of uncertainty/assumption was captured by all individuals who felt competent to score it, with the direction and magnitude of the potential model over (+/++/+++) or underprediction (-/--/---) judged based upon the stated evidence. All evidence was then reviewed as a team and a consensus score agreed, which is shown below.

Source of Uncertainty	**Direction & Magnitude**
1. That the vehicle volume and area is given by the volume and area of filter paper to which the dose is applied in formulation.	-/+
2. That the ingredient remains stable in the viable skin. No transformation apart from the molecular initiating event itself.	-/++
3. That stratum corneum, viable epidermis and dermis are well represented by the ODE approach.	-/+
4. That transport occurs by passive diffusion driven by a concentration gradient with partitioning modelled by applying Fick's first law.	-/+
5. That the transport and partitioning are dose independent and thus the model has a linear dependency when extrapolating across the dose under consideration.	---/+++
6. That the only transport processes acting on the ingredient-bound nucleophiles in intracellular space is loss by proteasome processing	-/+
7. That the rate of loss of ingredient-bound nucleophiles to proteasome processing is proportional to the concentration of ingredient-bound nucleophiles.	---/+++
8. That, once an ingredient-bound nucleophile is processed by the proteasome, it is either presented on Class I MHC or lost from the	--

system by some other process e.g. excretion.	
9. That the nucleophiles available on proteins can be treated as well mixed throughout viable skin.	--/++
10. That, for the purposes of modelling the reaction kinetics, the distinct nucleophiles considered are Cysteine, Tyrosine, Lysine, Histidine, Arginine, Methionine and the protein N-terminus.	-/+
11. The reaction stoichiometry is one haptenation event per nucleophilic binding site.	-/+
12. The fraction of nucleophilic biding sites are unavailable for binding remains constant over time	-/+
13. The reaction kinetics can be considered concentration dependent.	---/+++
14. That the reaction rate constant (but not the reaction rate) is concentration independent.	---/+++
15. That the reaction rate constant is time independent in the timescale under consideration.	---/+++
16. That free and ingredient-bound nucleophiles associated with proteasome generated peptides can be treated as a well-mixed homogeneous population for the purposes of describing their dynamics.	-/+
17. That the pMHC expressed on the surface of exposure-site derived DC is representative of the state peptide pool at the point of leaving the skin.	--/++
18. That the epitopes of the response are haptenated-nucleophiles within peptides presented on MHCI.	--/++
19. That the affinity of the interaction between the epitopes of the response and their cognate TCRs can be well represented by an average affinity for TcR. The average affinity (kon/koff) can be conceptualised in a thought experiment where all the distinct epitopes in a population are taken (variation over allele, peptide sequence, number of haptenated nucleophiles and specificity of haptenated nucleophiles) together with the set of TCRs which would respond to these epitopes. In such an experiment, the average affinity for TcR would be obtained by averaging the kon/koff values over all of these interactions.	--/++
20. That TcR triggering rate is well described by the 'confinement time' hypothesis.	---/+++
21. That cross-presentation (presentation of LC/DC endocytosed antigen via Class I MHC) does not significantly contribute to the pool of Class I MHC: peptide complexes presented on skin-derived DC	-/+
22. That the volume of internal skin exposure to the chemical is given by the surface area of the filter paper multiplied by the depth of the	++

skin (i.e. the chemical does not move laterally)	
23. The depth at which diffusion from viable skin to blood occurs is the same as the total skin depth	++
24. DNCB only enters the LN when it is bound within a DC	---

12. References

Davies M, Pendlington RU, Page L, et al. (2011). Determining epidermal disposition kinetics for use in an integrated nonanimal approach to skin sensitization risk assessment. Toxicological Sciences, 119: p.308–18.

Friedmann P, Moss C, Shuster S, Simpson J. M. (1983). Quantitative relationships between sensitizing dose of DNCB and reactivity in normal subjects. Clin. Exp. Immunol, 53: 709-7015

Gelman A, Carlin J. B, Stern H. S, Dunson D. B, Vehtari A and Rubin D. B. (2013). Bayesian Data Analysis, Third Edition. Taylor & Francis.

MacKay C, Davies M, Summerfield M and Maxwell G. (2013). From Pathways to People: Applying the Adverse Outcome Pathway (AOP) for Skin Sensitization to Risk Assessment. *ALTEX*. **30**. p.473-486.

MacKay C, Reynolds J, Dhadra S, Gellatly N, Stark R, Tang D and Maxwell G. (2016). Toxicokinetic/Toxicodynamic (TK/TD) modelling of skin sensitisation. Part II: Toxicodynamics and hazard characterisation. Tox. Sci, Submitted

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

Maxwell G, MacKay C, Cubberley R et al. (2014). Applying the skin sensitisation adverse outcome pathway (AOP) to quantitative risk assessment. Toxicol. In Vitro. 28: p.8-12.

Pendlington RU, Minter HJ, Stupart L, MacKay C, Roper CS, Sanders DJ, Pease CK. (2008). Development of a modified in vitro skin absorption method to study the epidermal/dermal disposition of a contact allergen in human skin. Cutan. Ocul. Toxicol. 27, 283–94.

Reynolds J, Pickles J, Cubberley R, Dhadra S, Gellatly N, Maxwell G, Pendlington RU, Saib O, and MacKay C. (2016). Toxicokinetic/Toxicodynamic (TK/TD) modelling of skin sensitisation. Part I: Toxicokinetics. Tox. Sci, Submitted.

Toni T, Welch D, Strelkowa N, Ipsen A and Stumpf M P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface 6, 187–202.

Zarnitsyna V and Zhu C. (2012). T cell triggering: insights from 2D kinetics analysis of molecular interactions. Phys. Biol. 9. 045005.

13. Supporting information

N/A

14. Abbreviations and definitions

AOP	Adverse outcome pathway
DC	Dendritic cell
dLN	Draining lymph node
DNCB	2,4-Dinitrochlorobenzene
HRIPT	Human repeat-insult patch test
LC	Langerhans' cell
LN	Lymph node
MHC	Major histocompatibility complex
pMHC	Peptide – Major histocompatibility complex
QRA	Quantitative risk assessment
TcR	T cell receptor

Appendix II.1: Details on the chemicals included in the training set

		Classification		LLNA
Chemical Name	**CAS**	**Human**	**LLNA**	**EC3 value**
5-Chloro-2-methyl-4-isothiazolin-3-one	26172-55-4	S	S	0,009
Diphenylcyclopropenone	886-38-4	S	S	0,003
Oxazolone	15646-46-5	S	S	0,003
p-Benzoquinone	106-51-4	S	S	0,0099
2,4-Dinitrochlorobenzene	97-00-7	S	S	0,05
nitrobenzyl bromide	100-11-8	S	S	0,05
1,4-Phenylenediamine	106-50-3	S	S	0,11
Cobalt(II) chloride	7646-79-9	S	S	0,6
Isoeugenol	97-54-1	S	S	1,5
2-Mercaptobenzothiazole	149-30-4	S	S	1,7
Cinnamic aldehyde	104-55-2	S	S	3
Tetramethylthiuram disulphide	137-26-8	S	S	5,2
Resorcinol	108-46-3	S	S	5,5
Diethyl maleate	141-05-9	S	S	5,8
Citral	5392-40-5	S	S	9,2
α-Hexylcinnamaldehyde	101-86-0	S	S	9,7
Eugenol	97-53-0	S	S	10,1
Cinnamyl alcohol	104-54-1	S	S	21
Benzocaine	94-09-7	S	S	22
Hydroxycitronellal	107-75-5	S	S	23
Imidazolidinyl urea	39236-46-9	S	S	24
Ethylene glycol dimethacrylate	97-90-5	S	S	28
Ethyl acrylate	140-88-5	S	S	28
Methyl methacrylate	80-62-6	S	S	90
tert-Butylhydroquinone	1948-33-0	S	S	NA
Nickel(II)chloride	7718-54-9	S	NS	FN
Triisobutylphosphate	126-71-6	S	NS	FN
Chlorobenzene	108-90-7	NS	NS	
2-Propanol	67-63-0	NS	NS	
Lactic acid	50-21-5	NS	NS	
Methyl salicylate	119-36-8	NS	NS	
Salicyclic acid	69-72-7	NS	NS	
Hexane	110-54-3	NS	NS	
Dextran	9004-54-0	NS	NS	
Propyleneglycol	57-55-6	NS	NS	
Tween80	9005-65-6	NS	NS	
Xylene	1330-20-7	NS	S	95,8 (FP)
Benzalkonium chloride	63449-41-2	NS	S	0,07 (FP)
Sodium dodecyl sulfate	151-21-3	NS	S	4 (FP)
Maleic acid	110-16-7	NS	S	7 (FP)
hexaethylene glycol monodecyl ether	5168-89-8	NS	S	FP

Abbreviations: S: skin sensitiser; NS: non-sensitiser; NA: no data available; FP: false-positive; FN: false-negative compared to human data

Appendix II.2: Predictions for the training set in each individual information source

Substance	Prediction*						
	Human	LLNA	Bayesian QSARS	DPRA	HaCaT gene signature	Keratinosens	hCLAT
5-Chloro-2-methyl-4-isothiazolin-3-one	S	S	S	S	S	S	S
Diphenylcyclopropenone	S	S	O	S	S	S	S
Oxazolone	S	S	O	S	S	S	S
p-Benzoquinone	S	S	O	S	S	S	S
2,4-Dinitrochlorobenzene	S	S	S	S	S	S	S
nitrobenzyl bromide	S	S	S	S	S	S	S
1,4-Phenylenediamine	S	S	O	S	S	S	S
Cobalt(II) chloride	S	S	O	S	S	S	S
Isoeugenol	S	S	S	S	S	S	O
2-Mercaptobenzothiazole	S	S	S	S	S	S	O
Cinnamic aldehyde	S	S	S	S	S	S	S
Tetramethylthiuram disulphide	S	S	O	S	S	S	S
Resorcinol	S	S	O	NS	S	NS	S
Diethyl maleate	S	S	S	S	S	S	S
Citral	S	S	S	S	S	S	S
α-Hexylcinnamaldehyde	S	S	S	NS	S	S	O
Eugenol	S	S	S	S	S	NS	S
Cinnamyl alcohol	S	S	O	S	S	NS	S
Benzocaine	S	S	O	NS	S	S	S
Hydroxycitronellal	S	S	S	S	S	S	S
Imidazolidinyl urea	S	S	O	S	S	S	O
Ethylene glycol dimethacrylate	S	S	S	S	S	S	S
Ethyl acrylate	S	S	S	S	S	S	X
Methyl methacrylate	S	S	S	S	S	NS	S
tert-Butylhydroquinone	S	S	O	S	S	S	X
Nickel(II)chloride	S	NS	O	S	S	S	O
Triisobutylphosphate	S	NS	O	S	S	NS	X
Chlorobenzene	NS	NS	O	NS	NS	NS	S
2-Propanol	NS	NS	NS	NS	NS	NS	NS
Lactic acid	NS	NS	NS	NS	NS	NS	NS
Methyl salicylate	NS	NS	NS	NS	NS	NS	NS
Salicyclic acid	NS	NS	NS	NS	NS	NS	O
Hexane	NS	NS	NS	NS	NS	NS	NS
Dextran	NS	NS	O	NS	NS	NS	NS
Propyleneglycol	NS	NS	NS	NS	NS	NS	NS
Tween80	NS	NS	O	NS	S	S	NS
Xylene	NS	S	NS	NS	NS	NS	NS
Benzalkonium chloride	NS	S	O	NS	S	S	NS
Sodium dodecyl sulfate	NS	S	NS	S	S	NS	NS
Maleic acid	NS	S	O	NS	NS	S	X
hexaethylene glycol monodecyl ether	NS	S	O	S	S	NS	X

Substances marked "S" are rated as positive in the individual assays;

Substances marked "NS" are rated as negative in the individual assays;

Substances marked "X" represent missing data in the h-CLAT

Substances marked "O" are equivocal calls in the assay.

Predictions are based on comparison with human evidence. Misclassifications are depicted in red.

Appendix II.3: Cooper statistics of the individual data sources

	Bayesian QSARs	DPRA	HaCaT gene signature	Keratinosens	hCLAT	LLNA[1]
Balanced accuracy	100.0%	88.3%	85.7%	80.0%	95.5%	78.4%
Sensitivity	100.0%	88.9%	100.0%	81.5%	100.0%	92.6%
Specificity	100.0%	85.7%	71.4%	78.6%	90.9%	64.3%
Balanced NPV	100.0%	88.5%	100.0%	80.1%	100.0%	89.7%
Balanced PPV	100.0%	86.2%	77.8%	79.2%	91.7%	72.2%
No prediction[2]	19	0	0	0	11	0

[1]Predictivity of the LLNA is shown only for the 41 substances in the training set; [2] Number of substances for which no prediction could be derived in the individual information source. These substances were excluded from the calculation of the predictivities; Balanced accuracy is calculated using the formula: =(specificity+sensitivity)/2; NPV: Balanced Negative Predictive Value (number of correctly predicted non-sensitisers as a fraction of all negative predictions; formula used: = specificity/(specificity + (1-sensitivity)); PPV: Positive Predictive Value (number of correctly predicted skin sensitisers as a fraction of all positive predictions; formula used: = sensitivity/(sensitivity + (1-specificity)).

Appendix II.4: Details on the predictions for the training set of the RIVM STS

			TIER 1			TIER 2				TIER 3	
Substance	Human	LLNA	Bayesian QSARS	DPRA	Interim Prediction	Follow-up assay	HaCaT gene signature	Keratinosens	Final or Interim decision	hCLAT	Final decision
5-Chloro-2-methyl-4-isothiazolin-3-one	S	S	S		S	Keratinosens		S	S		S
Diphenylcyclopropenone	S	S	O	S	S	Keratinosens		S	S		S
Oxazolone	S	S	O	S	S	Keratinosens		S	S		S
p-Benzoquinone	S	S	O	S	S	Keratinosens		S	S		S
2,4-Dinitrochlorobenzene	S	S	S		S	Keratinosens		S	S		S
nitrobenzyl bromide	S	S	S		S	Keratinosens		S	S		S
1,4-Phenylenediamine	S	S	O	S	S	Keratinosens		S	S		S
Cobalt(II) chloride	S	S	O	S	S	Keratinosens		S	S		S
Isoeugenol	S	S	S		S	Keratinosens		S	S		S
2-Mercaptobenzothiazole	S	S	S		S	Keratinosens		S	S		S
Cinnamic aldehyde	S	S	S		S	Keratinosens		S	S		S
Tetramethylthiuram disulphide	S	S	O	S	S	Keratinosens		S	S		S
Resorcinol	S	S	O	NS	NS	Gene signature	S		O	S	S
Diethyl maleate	S	S	S		S	Keratinosens		S	S		S
Citral	S	S	S		S	Keratinosens		S	S		S
α-Hexylcinnamaldehyde	S	S	S		S	Keratinosens		S	S		S
Eugenol	S	S	S		S	Keratinosens		NS	O	S	S
Cinnamyl alcohol	S	S	O	S	S	Keratinosens		NS	S		S
Benzocaine	S	S	O	NS	NS	Gene signature	S		O	S	S
Hydroxycitronellal	S	S	S		S	Keratinosens		S	S		S
Imidazolidinyl urea	S	S	O	S	S	Keratinosens		S	S		S
Ethylene glycol dimethacrylate	S	S	S		S	Keratinosens		S	S		S
Ethyl acrylate	S	S	S		S	Keratinosens		S	S		S
Methyl methacrylate	S	S	S		S	Keratinosens		NS	O		S
tert-Butylhydroquinone	S	S	O	S	S	Keratinosens		S	S		S
Nickel(II)chloride	S	NS	O	S	S	Keratinosens		S	S		S
Triisobutylphosphate	S	NS	O	S	S	Keratinosens		NS	O	X	O
Chlorobenzene	NS	NS	O	NS	NS	Gene signature	NS		NS		NS
2-Propanol	NS	NS	NS		NS	Gene signature	NS		NS		NS
Lactic acid	NS	NS	NS		NS	Gene signature	NS		NS		NS
Methyl salicylate	NS	NS	NS		NS	Gene signature	NS		NS		NS
Salicyclic acid	NS	NS	NS		NS	Gene signature	NS		NS		NS
Hexane	NS	NS	NS		NS	Gene signature	NS		NS		NS
Dextran	NS	NS	O		NS	Gene signature	NS		NS		NS
Propyleneglycol	NS	NS	NS		NS	Gene signature	NS		NS		NS
Tween80	NS	NS	O	NS	NS	Gene signature	S		O	NS	NS
Xylene	NS	S	NS		NS	Gene signature	NS		NS		NS
Benzalkonium chloride	NS	S	O	NS	NS	Gene signature	S		O	NS	NS
Sodium dodecyl sulfate	NS	S	NS		NS	Gene signature	S		O	NS	NS
Maleic acid	NS	S	O	NS	NS	Gene signature	NS		NS		NS
hexaethylene glycol monodecyl ether	NS	S	O	S	S	Keratinosens		NS	O	X	O

Substances marked "S" are rated as positive in the individual assays;

Substances marked "NS" are rated as negative in the individual assays;

Substances marked "X" represent missing data in the h-CLAT

Substances marked "O" are equivocal calls in the assay.

Predictions are based on comparison with human evidence. Misclassifications are depicted in red.

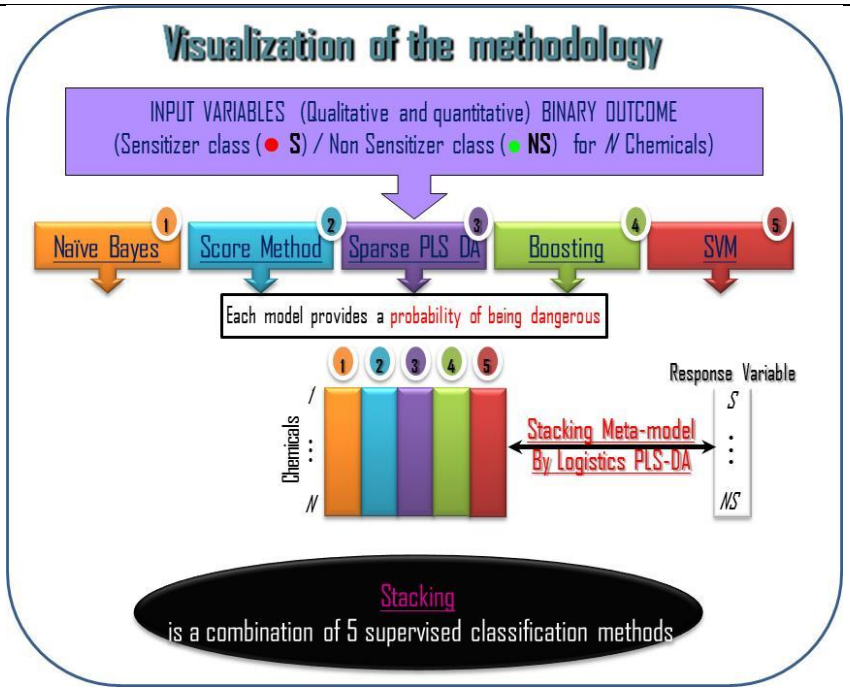
Appendix IV.1: Description of the statistical tools and the stacking meta-model

Although the template was not designed for this specific purpose, the case study submitter decided to use the *template for reporting individual information sources* to document in more detail and in a consistent way the different methodologies used to process the data

Stacking Methodology for combining Classifiers

Name of the information source	Stacking for a binary outcome
Mechanistic basis including AOP coverage	Not applicable
Description	Stacking (Wolpert (1992), Breiman (1996)) is known to be a successful way of linearly combining several models. We modify the usual stacking methodology when the response is binary and predictions highly correlated, by combining predictions with Logistic PLS-Discriminant Analysis. Our Stacking methodology combines linearly 5 models. Weights are obtained by PLS Logistic Discriminant analysis (Gomes et al., 2012) : Five prediction families have been addressed, covering a wide range of methods: sparse PLS discriminant analysis among linear models, tree boosting among decision trees, an internally developed expert scoring, SVM among machines learning and naïve Bayes classifier among probabilistic models.
Response(s)	Prediction of 2 groups (Non-sensitizer/Sensitizer) with a probability.
Prediction model	Stacking involves training a learning algorithm to combine the probability of predictions of five models (Sparse PLS-DA, Boosting, Scoring, Naïve Bayes, SVM and Boosting). First, all of the other algorithms are trained using the available data, then a combiner algorithm is trained to make a final prediction using all the probability of predictions of the other algorithms as additional inputs:

Stacking Methodology for combining Classifiers

	
Metabolic competence (if applicable)	Not applicable
Status of information source development, standardisation, validation	Developed in house – (Gomes *et al*., 2012)
Technical limitations and limitations with regard to applicability	Not applicable
Weaknesses and Strengths	Strengths: Instead of choosing one particular technique, a meta model combining several of them (efficient and complementary in terms of performance) will lead to an improved robust decision rules. Limitations: a major downside is that we lose the simple interpretability of rules.

Stacking Methodology for combining Classifiers

Reliability	The method has been tested on a validation sample and the results were similar to those obtained with the training sample.
Predictive capacity (if applicable)	Stacking leads to the conclusion over a greater number of chemicals, as defined below: • Chemicals with probability to be sensitizer ≥ 70% are predicted "Sensitizer"; • Chemicals with probability to be sensitizer ≤ 30% are predicted "Non Sensitizer"; • Chemicals with probability between those two thresholds are predicted "Equivocal". Training set 113 Chemicals (66 S / 47 NS) : • Sensitivity = 95.16% | Specificity = 90% | Kappa = 0.85 • Stacking prediction is conclusive on 90.27% (102/113) External validation Set 52 Chemicals (31 S/ 21 NS) : • Sensitivity = 92.86% | Specificity = 90% | Kappa = 0.83 • Stacking prediction is conclusive on 92.31% (48/52)
Proprietary aspects	R environment (open source software with package *PlsRglm (*Bertrand et al., 2014)
Proposed regulatory use	By combining five known classifiers of very different kinds, we obtained a prediction model with better performances than each of the five initial models taken separately. This result is important for the development of alternative approaches in safety evaluation of chemicals in cosmetic industry.
Potential role within the IATA	Not applicable

References

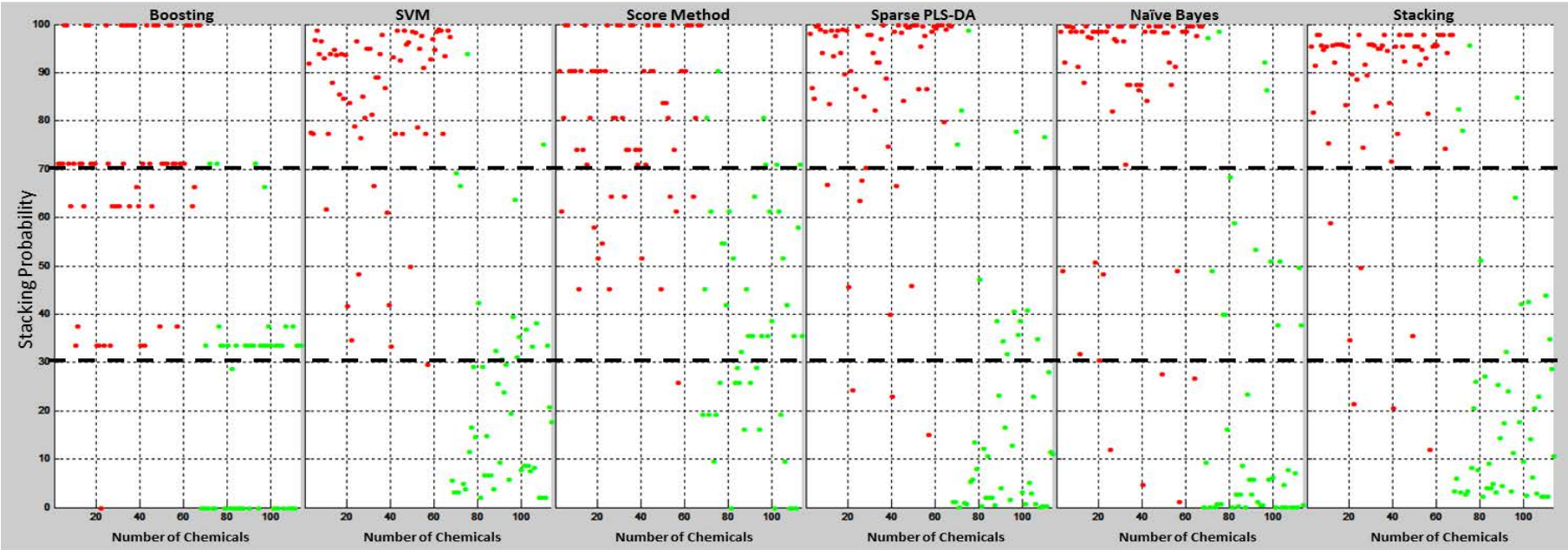
Bertrand F, Magnanensi J, Meyer N and Maumy-Bertrand M. (2014). plsRglm: Algorithmic insights and applications. Edited: June 2014; Compiled: July 17, 2014.

Breiman L. (1996). Stacked regressions. Machine Learning, 24 : 49-64.

Gomes C, Noçairi H, Thomas M, Ibanez F, Collin J and Saporta G. (2012). Stacking prediction for a binary outcome. Compstat 2012, August 2012, pp.271-282.

Wolpert D. (1992). Stacked Generalization. Neural Networks, 5, 41-259.

Training Set: Performances of all models



Boosting

	LLNA-based S / NS		
	S	NS	Total
S	43	3	46
NS	1	20	21
Total	44	23	67

Performances en %	
Concordance	94.03%
SENSITIVITY	97.73%
SPECIFICITY	86.96%
Kappa	0.86
% conclut	59.29%

SVM

	LLNA-based S / NS		
	S	NS	Total
S	56	2	58
NS	1	32	33
Total	57	34	91

Performances en %	
Concordance	96.70%
SENSITIVITY	98.25%
SPECIFICITY	94.12%
Kappa	0.93
% conclut	80.53%

Score Method

	LLNA-based S / NS		
	S	NS	Total
S	52	6	58
NS	1	19	20
Total	53	25	78

Performances en %	
Concordance	91.03%
SENSITIVITY	98.11%
SPECIFICITY	76.00%
Kappa	0.78
% conclut	69.03%

Sparse PLSDA

	LLNA-based S / NS		
	S	NS	Total
S	56	5	61
NS	3	33	36
Total	59	38	97

Performances en %	
Concordance	91.75%
SENSITIVITY	94.92%
SPECIFICITY	86.84%
Kappa	0.83
% conclut	85.84%

Naïve Bayes

	LLNA-based S / NS		
	S	NS	Total
S	55	4	59
NS	5	32	37
Total	60	36	96

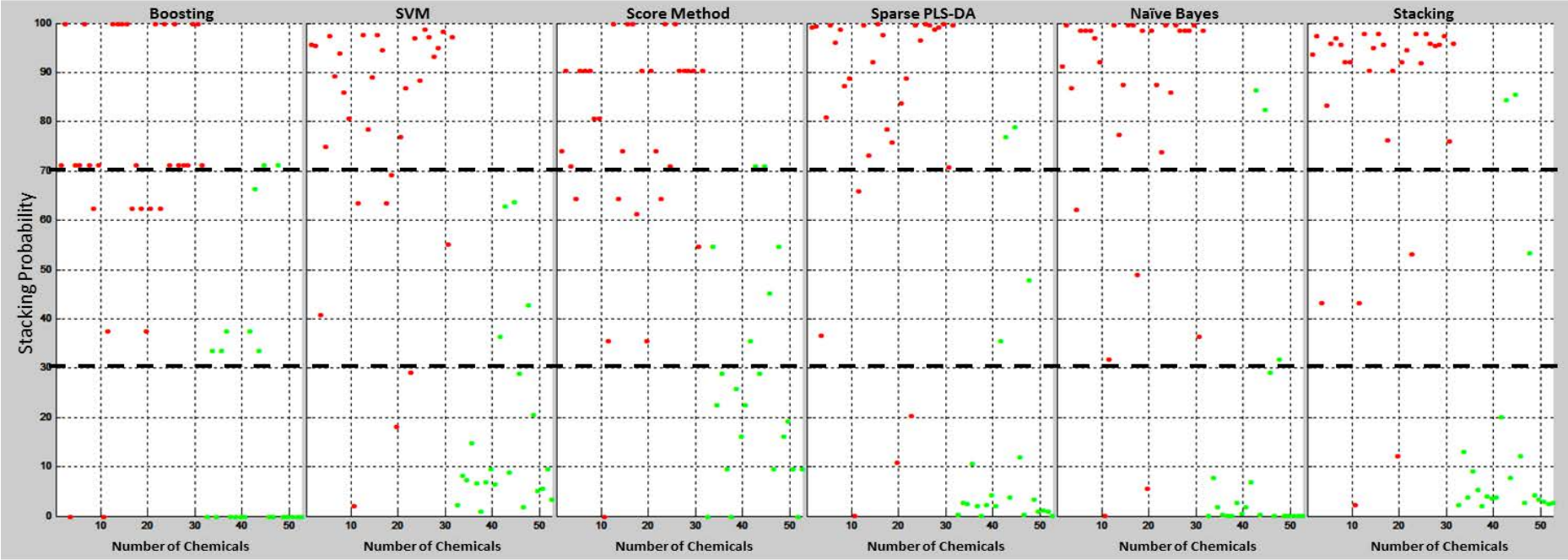
Performances en %	
Concordance	90.63%
SENSITIVITY	91.67%
SPECIFICITY	88.89%
Kappa	0.80
% conclut	84.96%

Stacking

	LLNA-based S / NS		
	S	NS	Total
S	59	4	63
NS	3	36	39
Total	62	40	102

Performances en %	
Concordance	93.14%
SENSITIVITY	95.16%
SPECIFICITY	90.00%
Kappa	0.86
% conclut	90.27%

Test Set: Performances of all models



Boosting

LLNA-based S / NS	S	NS	Total
S	22	2	24
NS	2	13	15
Total	24	15	39

Performances en %	
Concordance	89.74%
SENSITIVITY	91.67%
SPECIFICITY	86.67%
Kappa	0.78
% conclut	75.00%

SVM

LLNA-based S / NS	S	NS	Total
S	23	0	23
NS	3	17	20
Total	26	17	43

Performances en %	
Concordance	93.02%
SENSITIVITY	88.46%
SPECIFICITY	100.00%
Kappa	0.86
% conclut	82.69%

Score Method

LLNA-based S / NS	S	NS	Total
S	23	2	25
NS	1	15	16
Total	24	17	41

Performances en %	
Concordance	92.68%
SENSITIVITY	95.83%
SPECIFICITY	88.24%
Kappa	0.85
% conclut	78.85%

Sparse PLSDA

LLNA-based S / NS	S	NS	Total
S	26	2	28
NS	3	17	20
Total	29	19	48

Performances en %	
Concordance	89.58%
SENSITIVITY	89.66%
SPECIFICITY	89.47%
Kappa	0.78
% conclut	92.31%

Naïve Bayes

LLNA-based S / NS	S	NS	Total
S	25	2	27
NS	2	18	20
Total	27	20	47

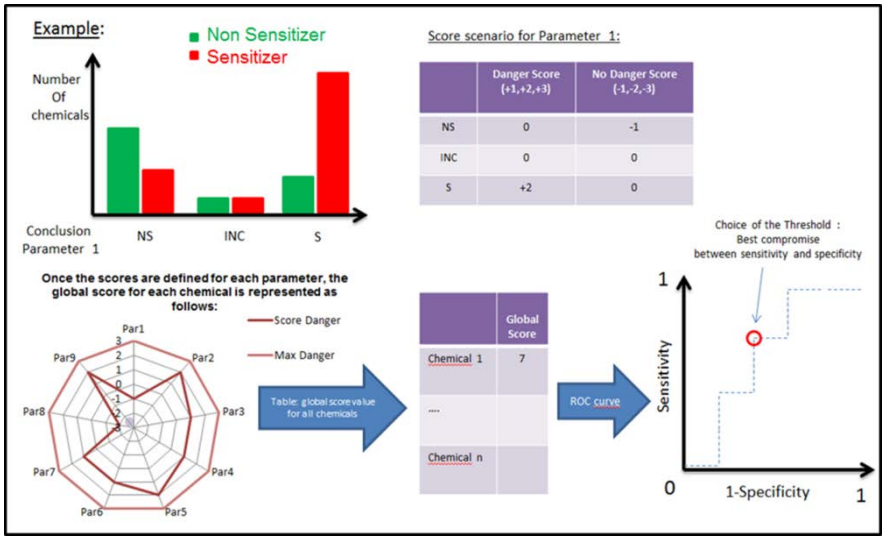
Performances en %	
Concordance	91.49%
SENSITIVITY	92.59%
SPECIFICITY	90.00%
Kappa	0.83
% conclut	90.38%

Stacking

LLNA-based S / NS	S	NS	Total
S	26	2	28
NS	2	18	20
Total	28	20	48

Performances en %	
Concordance	91.67%
SENSITIVITY	92.86%
SPECIFICITY	90.00%
Kappa	0.83
% conclut	92.31%

Expert ScoringTechnique

Name of the information source	Scoring method
Mechanistic basis including AOP coverage	Not applicable
Description	A new simple scoring technique used for in a binary supervised classification context when only a few observations are available. This method needs two steps: in the first one partial scores are obtained, one for each predictor, either categorical or continuous. Each partial score is a discrete variable with 7 values ranging from -3 to 3, based upon an empirical comparison of the distributions for each class. In a second step the partial scores are added and standardised into a global score, which allows a decision rule. 
Response(s)	Prediction of 2 groups (Non-sensitizer/Sensitizer) with a probability.
Prediction model	Like Boosting where a set of weak learners may produce a single stronger learner, our method obtains a strong prediction rule based upon a linear combination of weak rules, each rule being associated to a single predictor by a simple scoring technique. The expert Score provided a probability of being dangerous. This model was calibrated on the training set (113 chemicals) and validated on the test set (52 Chemicals).

Expert ScoringTechnique

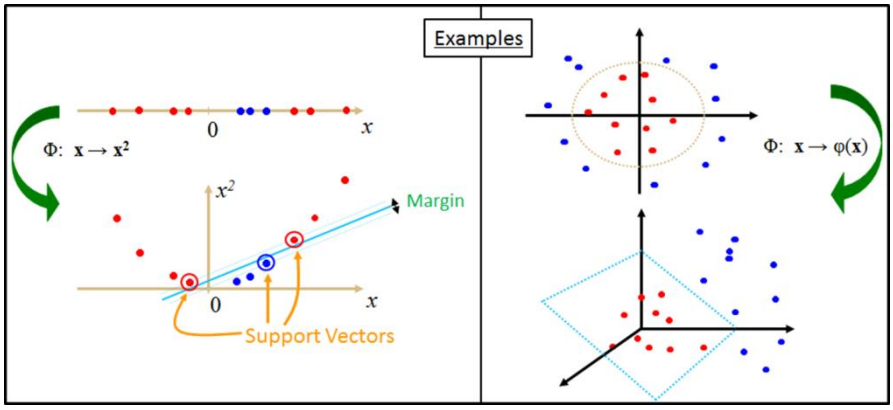
Metabolic competence (if applicable)	Not applicable
Status of information source development, standardisation, validation	Developed in house – (Gomes et al., 2014)
Technical limitations and limitations with regard to applicability	Not applicable
Weaknesses and Strengths	Strengths: Most of statistical model need parameter estimation and a large enough number of observations, this method combines some simple rules associated to different scores Limitations: is not applicable for a multiple class response.
Reliability	The method has been tested on a validation sample and the results were similar to those obtained with the training sample.
Predictive capacity (if applicable)	Training set 113 Chemicals (66 S / 47 NS) : • Sensitivity = 98.11% | Specificity = 76% | Kappa = 0.78 • Expert Score prediction is conclusive on 69.03% (78/113) External validation Set 52 Chemicals (31 S/ 21 NS) : • Sensitivity = 95.83% | Specificity = 88.24% | Kappa = 0.85 • Expert Score is conclusive on 78.85% (41/52)
Proprietary aspects	In house
Proposed regulatory use	The Score method may be combined with other models (Gomes et al., 2012).
Potential role within the IATA	Not applicable

References

Gomes C, Noçairi H, Thomas M, Ibanez F, Collin J and Saporta G. (2012). Stacking prediction for a binary outcome. Compstat 2012, August 2012, pp.271-282.

C., Noçairi, H., Thomas, M., Collin, J., and Saporta, G. (2014). A simple and robust scoring technique for binary classification. Artificial Intelligence Research, vol. 3(1), 52-58.

Support Vector Machines (SVM)

Name of the information source	Support Vector Machines (SVM)
Mechanistic basis including AOP coverage	Not applicable
Description	Support Vector Machines (Cortes and Vapnik, 1995) is widely used in machine learning for binary decision. This approach takes into account the fact that the predictors are potentially non-linearly related with the response variable. When data are linearly separable, the primary idea consists in finding the "thick" hyperplane which separates the data perfectly with a maximal margin. When data are not separable by a hyperplane, they may be linearly separated after a transformation, which maps the data into an extended "feature space". 
Response(s)	Prediction of 2 groups (Non-sensitizer/Sensitizer) with a probability.
Prediction model	The SVM provided a probability of being dangerous. This model was calibrated on the training set (113 chemicals) and validated on the test set (52 Chemicals).
Metabolic competence (if applicable)	Not applicable
Status of information source development, standardisation, validation	To achieve this methodology, we have chosen the R environment (open source software) and we used the package e1071 for SVM (Dimitriadou et al., 2011).

Support Vector Machines (SVM)

Technical limitations and limitations with regard to applicability	Not applicable
Weaknesses and Strengths	Strengths: Robustness towards small number of data points. Limitations: not directly applicable to predict more than two classes.
Reliability	The method has been tested on a validation sample and the results were similar to those obtained with the training sample.
Predictive capacity (if applicable)	Training set 113 Chemicals (66 S / 47 NS) : • Sensitivity = 98.25% | Specificity = 94.12% | Kappa = 0.93 • SVM prediction is conclusive on 80.53% (91/113) External validation Set 52 Chemicals (31 S/ 21 NS) : • Sensitivity = 88.46% | Specificity = 100% | Kappa = 0.86 • SVM prediction is conclusive on 82.69% (43/52)
Proprietary aspects	R environment (open source software with Package e1071)
Proposed regulatory use	The SVM classifier model may be combined with other models (Gomes et al., 2012).
Potential role within the IATA	Not applicable

References

Cortes C and Vapnik V. (1995). Support-vector network. Machine Learning, 20, 1-25.

Dimitriadou E, Hornik K., Leisch F, Meyer D and Weingessel A. (2011). e1071: Misc Functions of the Department of Statistics (e1071), TU Wien, R package version 1.5-26. http://CRAN.R-project.org/package=e1071.

Gomes C, Noçairi H, Thomas M, Ibanez F, Collin J and Saporta G. (2012). Stacking prediction for a binary outcome. Compstat 2012, August 2012, pp.271-282.

Naïve Bayes Classifier

Name of the information source	Naïve Bayes Model for binary classification
Mechanistic basis including AOP coverage	Not applicable
Description	Naïve Bayes method is a probabilistic method based on Bayes Theorem (Bayes, 1763): this approach calculates a probability that a result is true given a specific test outcome and then adding the probabilities together into a combined or posterior probability. In order to refine the "a priori" probability on each tests, a quality criterion (Quality Factor) is used, based on Klimisch-like (Buist et al., 2013 ; Kilimisch et al., 1997) code 1, 2 and 3 (noted *QF*) : • *Klimisch-like* "code 1" : reliable results → *QF* = 1 • *Klimisch-like* "code 2" : Doubtful results → *QF* = 0.8 • *Klimisch-like* "code 3" : Not reliable results → *QF* = 0.2 The aim of this criterion is to correct the observed "raw" prediction by taking into account the reliability on the test in the following way: • *Corrected Sensitivity* = 0.5+*QF*(*Sensitivity*-0.5) • *Corrected Specificity* = 0.5+*QF*(*Specificity*-0.5)
Response(s)	Prediction of 2 groups (Non-sensitizer/Sensitizer) with a probability.
Prediction model	The modified Naïve Bayes provided a probability of being dangerous. This model was calibrated on the training set (113 chemicals) and validated on the test set (52 Chemicals).
Metabolic competence (if applicable)	*Not applicable*
Status of information source development, standardisation, validation	To achieve this methodology, we have chosen the R environment (open source software) and we used the package e1071 for Naïve Bayes (Dimitriadou et al., 2011)

Naïve Bayes Classifier

Technical limitations and limitations with regard to applicability	Not applicable
Weaknesses and Strengths	Strengths: a prior information on the reliability of each test known as the klimisch score. Limitations: input must be qualitative
Reliability	The method has been tested on a validation sample and the results were similar to those obtained with the training sample.
Predictive capacity (if applicable)	Training set 113 Chemicals (66 S / 47 NS) : • Sensitivity = 91.67% | Specificity = 88.89% | Kappa = 0.80 • Bayesian prediction is conclusive on 84.86% (96/113) External validation Set 52 Chemicals (31 S/ 21 NS) : • Sensitivity = 92.59% | Specificity = 90% | Kappa = 0.90 • Bayesian prediction is conclusive on 90.38% (47/52)
Proprietary aspects	R environment (open source software)
Proposed regulatory use	The Naïve Bayes model may be combined with other models (Gomes et al., 2012).
Potential role within the IATA	Not applicable

References

Bayes T, (1763). An Essay towards solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London, 53:370-418.

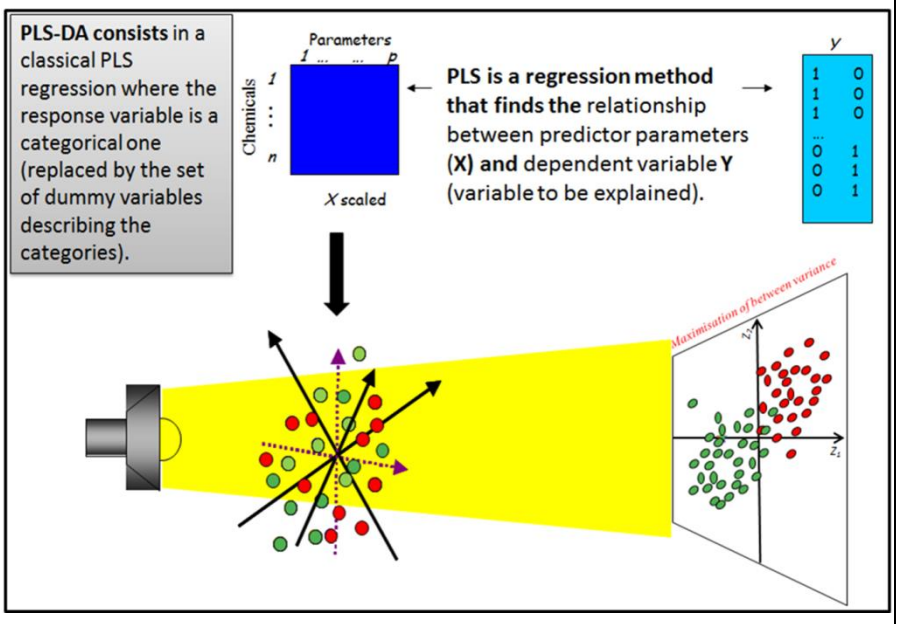
Buist H, Albenberg T, Baltke M, Escher S, Entink R.K, Künee R, Marquat H, Paune E, Rorije E, Schüürmann G and Kroese D. (2013). The OSIRIS Weight of Evidence approach. ITS mutagenicity and ITS Regulatory toxicology and pharmacology. Vol. 67, Issue 2, p 170-181.

Dimitriadou E, Hornik K., Leisch F, Meyer D and Weingessel A. (2011). e1071: Misc Functions of the Department of Statistics (e1071). TU Wien, R package version 1.5-26. http://CRAN.R-project.org/package=e1071.

Gomes C, Noçairi H, Thomas M, Ibanez F, Collin J and Saporta G. (2012). Stacking prediction for a binary outcome. Compstat 2012, August 2012, pp.271-282.

Klimisch H.J, Andreae M and Tillmann U. (1997). A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regulatory Toxicology and Pharmacology, 25, 1-5.

Sparse PLS-DA

Name of the information source	Sparse PLS Discriminant Analysis
Mechanistic basis including AOP coverage	Not applicable
Description	PLS D-A (Noçairi et al., 2005) is a regression method that finds the linear relationship between predictor parameters (inputs) and dependent variable (variable to be explained). The Sparse PLS-DA is based on the same concept as sparse PLS (Chun and Keles, 2010) to allow variable selection, except that for this method, the variables are only selected in the original predictors and in a supervised framework. The X-variables are selected with respect to different categories of the samples. 
Response(s)	Prediction of 2 groups (Non-sensitizer/Sensitizer) with a probability.
Prediction model	Sparse PLS-DA provided a probability of being dangerous. This model was calibrated on the training set (113 chemicals) and validated on the test set (52 Chemicals).
Metabolic competence (if	Not applicable

Sparse PLS-DA

applicable)	
Status of information source development, standardisation, validation	To achieve this methodology, we have chosen the R environment (open source software) and we used the package *spls* (Chun et al., 2010).
Technical limitations and limitations with regard to applicability	Not applicable
Weaknesses and Strengths	Strengths: Data visualization is one of the clear advantages of projection-based methods, the original Sparse PLS-DA, compared to the other approaches. For example, the latent variables can be used to represent the similarities and dissimilarities between the samples. Limitations: This method is optimal only in case of linear relationships
Reliability	The method has been tested on a validation sample and the results were similar to those obtained with the training sample.
Predictive capacity (if applicable)	Training set 113 Chemicals (66 S / 47 NS) : • Sensitivity = 94.92% | Specificity = 86.84% | Kappa = 0.82 • Sparse PSL-DA prediction is conclusive on 85.84% (97/113) External validation Set 52 Chemicals (31 S/ 21 NS) : • Sensitivity = 89.66% | Specificity = 89.47% | Kappa = 0.78 • Sparse PSL-DA prediction is conclusive on 92.31% (48/52)
Proprietary aspects	R environment (open source software with Package *spls*)
Proposed regulatory use	The Sparse PLS-DA model may be combined with other models (Gomes et al., 2012).
Potential role within the IATA	Not applicable

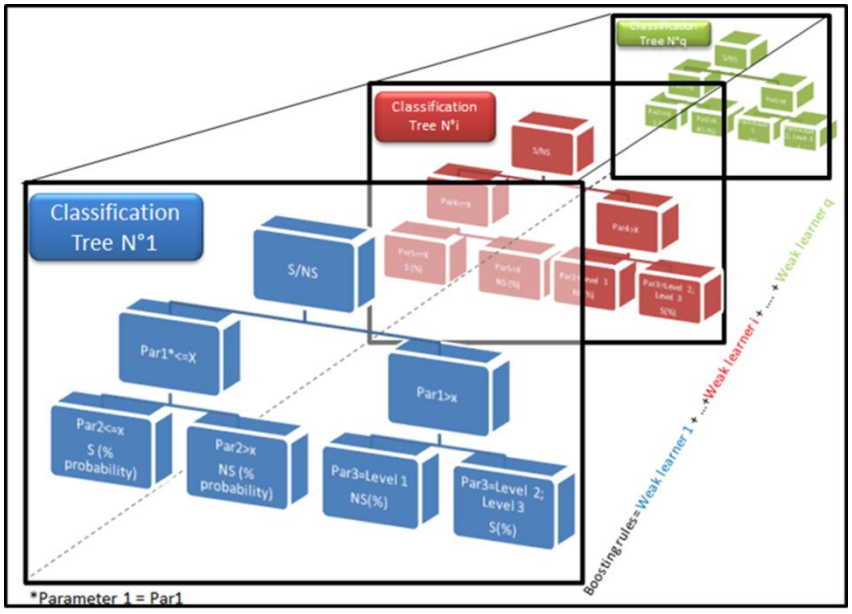
References

Chun H, Keleş S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72: 3–25.

Gomes C, Noçairi H, Thomas M, Ibanez F, Collin J and Saporta G. (2012). Stacking prediction for a binary outcome. Compstat 2012, August 2012, pp.271-282.

Noçairi H, Qannari EM, Vigneau E and Bertrand D. (2005). Discrimination on latent components with respect to patterns. Application to multicollinear data. Computational Statistics & Data Analysis, 48, 139-147.

Boosting Ensemble Learning

Name of the information source	Boosting Ensemble Learning
Mechanistic basis including AOP coverage	Not applicable
Description	Boosting is a supervised classification method based on an approach developed by Shapire (1990). This method combines weak learners which generate a single strong learner. The boosting is a combination of decision trees (weak learners): in the first step, a decision tree is calculated on all the data identically balanced. The predictions are re-calculated and a higher weighting is then assigned to the misclassified data. The final classifier is sum of the weights of all the trees. 
Response(s)	Prediction of 2 groups (Non-sensitizer/Sensitizer) with a probability.
Prediction model	Boosting provided a probability of being dangerous. This model was calibrated on the training set (113 chemicals) and validated on the test set (52 Chemicals).
Metabolic competence (if applicable)	Not applicable

Boosting Ensemble Learning

Status of information source development, standardisation, validation	To achieve this methodology, we have chosen the R environment (open source software) and we used the package *adabag* (Alfaro et al., 2013).
Technical limitations and limitations with regard to applicability	Not applicable
Weaknesses and Strengths	Strengths: This model has the advantage of being able to detect non-linear relationships and showing a good performance in presence of qualitative information. Limitations: a major downside is that we lose the simple interpretability of classification trees. The final classifier is a weighted sum of trees, which cannot necessarily be represented by a single tree.
Reliability	The method has been tested on a validation sample and the results were similar to those obtained with the training sample.
Predictive capacity (if applicable)	Training set 113 Chemicals (66 S / 47 NS) : • Sensitivity = 97.73% | Specificity = 86.96% | Kappa = 0.86 • Boosting prediction is conclusive on 79.29% (67/113) External validation Set 52 Chemicals (31 S/ 21 NS) : • Sensitivity = 91.67% | Specificity = 86.67% | Kappa = 0.78 • Boosting prediction is conclusive on 75% (39/52)
Proprietary aspects	R environment (open source software with package adabag)
Proposed regulatory use	The Boosting model may be combined with other models (Gomes et al., 2012).
Potential role within the IATA	Not applicable

References

Alfero E, Gamez M, Garcia N. (2013). Adabag: Na R Package for classification with Boosting and Bagging. Journal of statistical software. August 2013, Vol. 54, Issue 2.

Gomes C, Noçairi H, Thomas M, Ibanez F, Collin J and Saporta G. (2012). Stacking prediction for a binary outcome. Compstat 2012, August 2012, pp.271-282.

Shapire R.E. (1990). The strength of weak learnability. Machine Learning, 5(2), 197-227.Alfaro, E., Gamez, M. and Garcia, N. (2013): adabag: An R Package for Classification with Boosting and Bagging. Journal of Statistical Software, 54(2), 1–35.

Appendix IV.2: Description of the global chemical set, the informative source data and the DIP outcomes

N	INCI name or chemical name	CAS	SMILE	ingredient type	Chemical Class	MW	cLog P v5.2	Water Sol (mol/L) Epiwin v4.0	pH	Volatility	Protein binding class Toxtree v2.5.0	Toxtree v2.5.0	TIMES	DPRA	U-SENS	KeratinoSens	ITS (p)	ITS result	LLNA	Set
1	SALICYLIC ACID	69-72-7	c1ccc(c(c1)O)C(=O)O	mono-ingredient	PRESERVATIVE	138.12	2.19	2.76 E-02	2.5	Semi volatil	No binding	NR	NS	INC	NS	NS	3.5	NS	NS	Training
2	BENZOIC ACID	65-85-0	C1=CC=C(C=C1)C(=O)O	mono-ingredient	PRESERVATIVE	122.12	1.88	2.04 E-02	2.9	Semi volatil	No binding	NR	NS	INC	NS	NS	3.4	NS	NS	Test
3	BENZYL ALCOHOL	100-51-6	c1ccc(cc1)CO	mono-ingredient	PRESERVATIVE	108.14	1.10	3.80 E-01	8.2	Semi volatil	No binding	NR	NS	NR	NS	NS	2.2	NS	NS	Training
4	p-METHYLAMINOPENOL SULFATE	55-55-0	c1(ccc(cc1)O)NC.S(O)(=O)(O)=O	mono-ingredient	DYE/PIGMENT	221.23	0.97	1.67 E-01	3.9	Semi volatil	Michael Acceptor	R	S	R	S	S	97.9	S	S	Training
5	PROPYLPARABEN	94-13-3	c1(ccc(cc1)C(=O)OCCC)O	mono-ingredient	PRESERVATIVE	180.21	3.04	2.94 E-03	7.0	Semi volatil	No binding	NR	NS	NR	S	S	10.8	NS	NS	Training
6	BENZALKONIUM CHLORIDE	8001-54-5	NA	complex ingredient	PRESERVATIVE	NA	NA	NA	6.3	NA	NA	NA	NA1	NR	S	NS	17.8	NS	NS	Training
7	OA1	OA1	OA1	mono-ingredient	ACTIVE COMPOUND	166.22	2.58	4.50 E-03	7.5	Semi volatil	Michael Acceptor	R	S	R	S	S	95.6	S	S	Training
8	OA2	OA2	OA2	complex ingredient	POLYMER	NA	NA	NA	8.3	NA	NA	NA	NA1	NR	S	S	34.8	Equivocal	S	Training
9	OA3	OA3	OA3	complex ingredient	SURFACTANT	NA	NA	NA	3.6	NA	NA	NA	NA1	R	S	S	92.1	S	S	Test
10	OA4	OA4	OA4	mono-ingredient	DYE/PIGMENT	393.30	7.44	7.60 E-08	2.4	Non volatil	No binding	NR	INC	R	S	NS	35.0	Equivocal	NS	Training

11	OA5	OA5	OA5	mono-ingredient	FATTY COMPOUND	639.07	17.27	3.07 E-17	7.8	Non volatil	No binding	NR	NS	NR	NS	S	2.2	NS	NS	Test
12	OA6	OA6	OA6	complex ingredient	SURFACTANT	NA	NA	NA	7.8	NA	NA	NA	NA1	R	S	NS	77.3	S	S	Training
13	OA7	OA7	OA7	complex ingredient	SURFACTANT	NA	NA	NA	4.2	NA	NA	NA	NA1	R	S	NS	74.5	S	S	Training
14	OA8	OA8	OA8	complex ingredient	POLYMER	NA	NA	NA	8.0	NA	NA	NA	NA1	INC	NS	S	4.2	NS	NS	Test
15	OA9	OA9	OA9	mono-ingredient	PRESERVATIVE	146.23	1.59	2.01 E-02	3.7	Semi volatil	No binding	NR	NS	NR	S	NS	9.2	NS	NS	Test
16	OA10	OA10	OA10	complex ingredient	POLYMER	NA	NA	NA	6.9	NA	NA	NA	NA1	INC	NS	S	7.9	NS	NS	Training
17	OA11	OA11	OA11	complex ingredient	SURFACTANT	NA	NA	NA	8.1	NA	NA	NA	NA1	INC	S	S	51.2	Equivocal	NS	Training
18	OA12	OA12	OA12	complex ingredient	VEGETABLE EXTRACT	NA	NA	NA	8.2	NA	NA	NA	NA1	INC	NS	S	4.1	NS	NS	Training
19	OA13	OA13	OA13	complex ingredient	VEGETABLE EXTRACT	NA	NA	NA	4.2	NA	NA	NA	NA1	R	S	S	91.8	S	S	Training
20	OA14	OA14	OA14	mono-ingredient	SOLVENT	94.13	-1.50	1.06 E+01	8.0	Semi volatil	No binding	NR	NS	NR	NS	NS	2.2	NS	NS	Training
21	OA15	OA15	OA15	mono-ingredient	ACTIVE COMPOUND	147.15	-0.25	4.12 E+00	2.1	Semi volatil	Acyl Transfer agent	R	INC	NR	NS	NS	20.2	NS	NS	Test
22	OA16	OA16	OA16	complex ingredient	SURFACTANT	NA	NA	NA	5.7	NA	NA	NA	NA1	NR	NS	NS	3.6	NS	NS	Test
23	OA17	OA17	OA17	complex ingredient	POLYMER	NA	NA	NA	8.2	NA	NA	NA	NA1	INC	NS	INC	4.1	NS	NS	Training
24	OA18	OA18	OA18	complex ingredient	POLYMER	NA	NA	NA	8.2	NA	NA	NA	NA1	INC	NS	S	5.0	NS	NS	Training

25	OA19	OA19	OA19	mono-ingredient	SOLVENT	226.27	4.21	5.30 E-05	8.3	Semi volatil	No binding	NR	NS	NR	NS	S	3.1	NS	NS	Training
26	OA20	OA20	OA20	complex ingredient	POLYMER	NA	NA	NA	5.9	NA	NA	NA	NA1	INC	S	INC	25.4	NS	NS	Training
27	OA21	OA21	OA21	complex ingredient	POLYMER	NA	NA	NA	7.5	NA	NA	NA	NA	NR	NS	No data	3.9	NS	NS	Test
28	benzyl bromide	100-39-0	c1(ccccc1)CBr	mono-ingredient	NON COSMETIC	171.04	2.92	1.14 E-03	3.6	Very volatil	SN2	R	S	R	S	S	98.0	S	S	Training
29	2-methyl-3-phenyl-2-propenal	101-39-3	c1cccc(c1)/C=C(\C=O)/C	mono-ingredient	FRAGRANCE INGREDIENT	146.19	2.36	5.15 E-03	8.6	Semi volatil	Michael Acceptor	R	S	R	S	S	95.5	S	S	Training
30	p-PHENYLENEDIAMINE	106-50-3	c1cc(ccc1N)N	mono-ingredient	DYE/PIGMENT	108.14	-0.31	1.84 E+00	9.0	Semi volatil	Michael Acceptor	R	S	R	S	S	95.4	S	S	Training
31	1,4-benzoquinone	106-51-4	C\1(\C=C/C(\C=C1)=O)=O	mono-ingredient	NON COSMETIC	108.10	0.21	6.89 E-01	7.9	Semi volatil	Michael Acceptor	R	S	R	S	S	95.6	S	S	Test
32	ethylenediamine	107-15-3	C(N)CN	mono-ingredient	NON COSMETIC	60.10	-2.02	1.66 E+01	11.4	Very volatil	Schiff base formation	R	INC	NR	S	S	81.8	S	S	Training
33	GLYOXAL	107-22-2	O=CC=O	mono-ingredient	PRESERVATIVE	58.04	-1.80	1.72 E+01	5.0	Very volatil	No binding	NR	INC	R	S	S	94.8	S	S	Training
34	HYDROXYCITRONELLAL	107-75-5	C(CCCC(CC=O)C)(O)(C)C	mono-ingredient	FRAGRANCE INGREDIENT	172.27	1.54	1.77 E-02	3.4	Semi volatil	Schiff base formation	R	S	R	S	S	97.9	S	S	Test
35	m-PHENYLENEDIAMINE	108-45-2	c1(cccc(c1)N)N	mono-ingredient	DYE/PIGMENT	108.14	-0.31	1.95 E+00	9.2	Semi volatil	Michael Acceptor	R	INC	R	S	S	91.5	S	S	Training
36	RESORCINOL	108-46-3	c1cc(cc(c1)O)O	mono-ingredient	DYE/PIGMENT	110.11	0.81	7.78 E-01	6.3	Semi volatil	Michael Acceptor	R	S	R	S	NS	93.2	S	S	Training

37	3-dimethylaminopropylamine	109-55-7	C(CCN(C)C)N	mono-ingredient	NON COSMETIC	102.18	-0.03	9.79E+00	11.5	Very volatil	Schiff base formation	R	INC	NR	S	S	81.7	S	S	Training
38	1-bromobutane	109-65-9	BrCCCC	mono-ingredient	NON COSMETIC	137.02	2.66	2.82E-03	8.9	Very volatil	SN2	R	NS	NR	NS	NS	4.2	NS	NS	Test
39	ISOPROPYL MYRISTATE	110-27-0	C(CCCCCCCCCCCC(=O)OC(C)C)C	mono-ingredient	FATTY COMPOUND	270.46	7.37	5.01E-08	8.4	Semi volatil	No binding	NR	NS	NR	NS	NS	2.2	NS	S	Test
40	HEXANE	110-54-3	CCCCCC	mono-ingredient	SOLVENT	86.18	3.87	2.00E-04	8.9	Very volatil	No binding	NR	NS	NR	NS	NS	2.5	NS	NS	Training
41	OA22	OA22	OA22	mono-ingredient	FRAGRANCE INGREDIENT	154.21	2.71	2.81E-03	5.4	Very volatil	Michael Acceptor	R	S	R	S	S	97.5	S	S	Test
42	diethylenetriamine	111-40-0	NCCNCCN	mono-ingredient	NON COSMETIC	103.17	-2.28	9.69E+00	11.2	Very volatil	Schiff base formation	R	INC	NR	NS	NS	12.0	NS	S	Training
43	safranal	116-26-7	C/1=C/C(=C(\C(C1)(C)C)/C=O)/C	mono-ingredient	NON COSMETIC	150.22	2.93	8.93E-04	3.3	Very volatil	Schiff Base Formation Michael Acceptor	R	S	R	S	S	98.0	S	S	Training
44	METHYL SALICYLATE	119-36-8	c1(ccccc1O)C(=O)OC	mono-ingredient	FRAGRANCE INGREDIENT	152.15	2.34	1.23E-02	7.8	Semi volatil	No binding	NR	NS	NR	NS	NS	2.2	NS	NS	Test
45	OA23	OA23	OA23	mono-ingredient	FRAGRANCE INGREDIENT	148.16	1.63	7.79E-02	5.4	Semi volatil	Acyl Transfer agent	R	S	NR	NS	NS	43.4	Equivocal	S	Test
46	BENZYL BENZOATE	120-51-4	c1cc(ccc1)C(=O)OCc1ccccc1	mono-ingredient	FRAGRANCE INGREDIENT	212.25	3.94	7.25E-05	8.8	Semi volatil	Acyl Transfer agent	R	S	NR	S	S	91.9	S	S	Test
47	ETHYL VANILLIN	121-32-4	c1cc(cc(c1O)OCC)C=O	mono-ingredient	FRAGRANCE INGREDIENT	166.18	2.06	1.73E-02	5.6	Semi volatil	Schiff base formation	R	INC	NR	NS	S	22.9	NS	NS	Training

48	Sulfanilic acid	121-57-3	c1c(ccc(c1)S(O)(=O)=O)N	mono-ingredient	NON COSMETIC	173.19	-1.87	2.40 E-01	2.4	Non volatil	No binding	NR	NS	INC	NS	NS	3.3	NS	NS	Training
49	PROPYL GALLATE	121-79-9	c1(c(cc(cc1O)C(OCCC)=O)O)O	mono-ingredient	FRAGRANCE INGREDIENT	212.20	1.99	1.97 E-02	6.0	Semi volatil	Michael Acceptor	R	S	R	S	S	95.7	S	S	Test
50	4-phenyl-3-buten-2-one	122-57-6	c1ccccc1\C=C\C(=O)C	mono-ingredient	NON COSMETIC	146.19	2.15	9.20 E-03	8.3	Semi volatil	Michael Acceptor	R	S	R	S	S	95.5	S	S	Training
51	phenylacetaldehyde	122-78-1	c1c(cccc1)CC=O	mono-ingredient	NON COSMETIC	120.15	1.78	2.52 E-02	3.4	Very volatil	Schiff base formation	R	S	R	S	S	98.0	S	S	Training
52	HYDROQUINONE	123-31-9	c1(ccc(cc1)O)O	mono-ingredient	ACTIVE COMPOUND	110.11	0.81	1.18 E+00	7.2	Semi volatil	Michael Acceptor	R	S	R	S	S	95.6	S	S	Training
53	CAPRYLIC ACID	124-07-2	C(CCCCCC(O)=O)C	mono-ingredient	SURFACTANT	144.21	2.98	3.44 E-03	3.7	Semi volatil	No binding	NR	NS	NR	S	NS	9.2	NS	NS	Training
54	THIRAM	137-26-8	CN(C(SSC(=S)N(C)C)=S)C	mono-ingredient	PRESERVATIVE	240.44	1.76	3.77 E-02	8.7	Semi volatil	No binding	NR	S	R	S	S	92.2	S	S	Training
55	estragole	140-67-0	c1cc(ccc1OC)CC=C	mono-ingredient	NON COSMETIC	148.20	3.13	5.71 E-04	8.3	Very volatil	Michael Acceptor	R	S	R	S	NS	93.7	S	S	Test
56	diethyl maleate	141-05-9	C(\C=C/C(=O)OCC)(=O)OCC	mono-ingredient	NON COSMETIC	172.18	1.84	4.65 E-03	8.8	Very volatil	Michael Acceptor	R	S	R	S	S	95.8	S	S	Training
57	OA24	OA24	OA24	mono-ingredient	NON COSMETIC	167.24	2.96	3.25 E-03	5.4	Semi volatil	Acyl Transfer agent	R	S	R	S	S	97.4	S	S	Test
58	SODIUM LAURYL SULFATE	151-21-3	S(OCCCCCCCCCCCC)(=O)([O-])=O.[Na+]	mono-ingredient	SURFACTANT	288.40	5.00	6.14 E-04	8.3	Non volatil	SN2	R	NS	R	S	NS	13.0	NS	NS	Test
59	oxazolone	15646-46-5	O\1C(\C(\N=C1\c1ccccc1)=C/OCC)=O	mono-ingredient	NON COSMETIC	217.22	2.04	8.50 E-03	3.7	Semi volatil	Acyl Transfer agent	R	S	R	S	S	97.9	S	S	Training

60	farnesal	19317-11-4	C(\CC\C=C(\C)/C)(/C)=C/CC/C(=C/C=O)/C	mono-ingredient	FRAGRANCE INGREDIENT	220.36	4.98	1.94E-06	4.9	Semi volatil	Schiff base formation	R	S	R	S	S	97.8	S	S	Training
61	butyl glycidyl ether	2426-08-06	CCCCOCC1CO1	mono-ingredient	NON COSMETIC	130.19	1.24	2.04E-01	8.2	Very volatil	SN2	R	S	R	S	S	95.9	S	S	Test
62	METHYLCHLOROISOTHIAZOLINONE	26172-55-4	CN1SC(Cl)=CC1=O	mono-ingredient	PRESERVATIVE	149.60	0.60	9.96E-01	4.3	Semi volatil	SN2	R	S	R	S	S	97.9	S	S	Training
63	BENZISOTHIAZOLINONE	2634-33-5	c1cc2c(cc1)C(NS2)=O	mono-ingredient	PRESERVATIVE	151.19	0.61	1.42E-01	6.3	Semi volatil	Acyl Transfer agent SN2	R	S	R	S	S	95.7	S	S	Training
64	chloro octadécane	3386-33-2	C(CCCCCCCCCCCC)CCCCCCl	mono-ingredient	NON COSMETIC	288.95	9.93	4.31E-10	8.2	Semi volatil	SN2	R	S	INC	NS	INC	35.7	Equivocal	S	Training
65	3,5,5-trimethylhexanoyl chloride	36727-29-4	ClC(CC(CC(C)(C)C)C)=O	mono-ingredient	NON COSMETIC	176.69	3.44	1.36E-03	1.6	Very volatil	Acyl Transfer agent	R	S	R	S	NS	95.4	S	S	Training
66	diacetyl, 2,3 butanedione	431-03-8	C(C(=O)C)(=O)C	mono-ingredient	FRAGRANCE INGREDIENT	86.09	-1.37	1.16E+01	4.3	Very volatil	Schiff base formation	R	S	R	S	S	97.9	S	S	Training
67	furil	492-94-4	c1ccc(o1)C(=O)C(c1occc1)=O	mono-ingredient	NON COSMETIC	190.15	1.73	2.25E-02	8.5	Semi volatil	Schiff base formation	R	NS	R	S	S	42.6	Equivocal	NS	Training
68	LACTIC ACID	50-21-5	CC(C(O)=O)O	mono-ingredient	ACTIVE COMPOUND	90.08	-0.73	1.11E+01	2.4	Semi volatil	No binding	NR	NS	NR	NS	NS	2.9	NS	NS	Test
69	hydrocortisone	50-23-7	CC12CCC(=O)C=C1CCC3C2C(CC4(C3CCC4(C(=O)CO)O)C)O	mono-ingredient	NON COSMETIC	362.47	0.73	6.06E-04	8.3	Non volatil	Schiff base formation	R	INC	R	NS	S	14.2	NS	NS	Training
70	2-NITRO-p-PHENYLENEDIAMINE	5307-14-2	[N+](=O)([O-])c1c(ccc(c1)N)N	mono-ingredient	DYE/PIGMENT	153.14	0.75	1.69E-01	8.2	Semi volatil	Michael Acceptor	R	S	R	S	S	95.5	S	S	Training

71	trimellitic anhydride	552-30-7	c1cc(cc2c1C(OC2=O)=O)C(=O)O	mono-ingredient	NON COSMETIC	192.13	1.61	5.39 E-03	2.1	Semi volatil	Acyl Transfer agent	R	S	NR	NS	NS	58.9	Equivocal	S	Training
72	GLYCEROL	56-81-5	OCC(CO)O	mono-ingredient	SOLVENT	92.09	-1.54	1.09 E+01	8.3	Semi volatil	No binding	NR	NS	NR	NS	NS	2.2	NS	NS	Training
73	PROPYLENE GLYCOL	57-55-6	CC(CO)O	mono-ingredient	SOLVENT	76.09	-1.06	1.07 E+01	9.1	Very volatil	No binding	NR	NS	INC	NS	NS	3.0	NS	NS	Training
74	1-phenyl-1,2-propanedione	579-07-7	c1cccc(c1)C(=O)C(=O)C	mono-ingredient	NON COSMETIC	148.16	1.07	5.88 E-02	3.9	Very volatil	Schiff base formation	R	S	R	S	S	97.9	S	S	Test
75	sulfanilamide	63-74-1	c1c(ccc(c1)S(N)(=O)=O)N	mono-ingredient	NON COSMETIC	172.21	-0.57	1.18 E+00	8.2	Semi volatil	No binding	NR	NS	NR	NS	NS	2.2	NS	NS	Training
76	METHYLPARABEN	99-76-3	C(OC)(c1ccc(cc1)O)=O	mono-ingredient	PRESERVATIVE	152.20	1.99	3.93 E-02	6.4	Semi volatil	No binding	NR	NS	NR	S	S	11.3	NS	NS	Training
77	diethylsulfate	64-67-5	C(OS(OCC)(=O)=O)C	mono-ingredient	NON COSMETIC	154.18	1.09	5.01 E-02	2.2	Very volatil	SN2	R	S	R	S	S	98.0	S	S	Training
78	methyl methanesulfonate	66-27-3	S(C)(OC)(=O)=O	mono-ingredient	NON COSMETIC	110.13	-0.48	3.60 E+00	3.2	Very volatil	SN2	R	S	R	S	S	98.0	S	S	Training
79	ISOPROPYL ALCOHOL	67-63-0	OC(C)C	mono-ingredient	SOLVENT	60.10	0.07	6.70 E+00	8.8	Very volatil	No binding	NR	S	NR	NS	NS	44.0	Equivocal	NS	Training
80	trans-2-hexen-1-al	6728-26-3	O=C\C=C\CCC	mono-ingredient	NON COSMETIC	98.14	1.58	5.36 E-02	2.9	Very volatil	Michael Acceptor	R	S	R	S	S	98.0	S	S	Training
81	n-BUTYL ALCOHOL	71-36-3	CCCCO	mono-ingredient	SOLVENT	74.12	0.82	1.03 E+00	8.2	Very volatil	No binding	NR	NS	NR	NS	NS	2.5	NS	NS	Test

82	nonanoyl chloride	764-85-2	CCCCCCCCC(Cl)=O	mono-ingredient	NON COSMETIC	176.69	3.83	9.49 E-04	2.0	Very volatil	Acyl Transfer agent	R	S	R	S	NS	95.4	S	S	Training
83	DIETHYL PHTHALATE	84-66-2	c1(c(cccc1)C(OCC)=O)C(OCC)=O	mono-ingredient	SOLVENT	222.24	3.14	1.29 E-03	8.4	Semi volatil	No binding	NR	NS	NR	S	NS	6.3	NS	NS	Training
84	TARTARIC ACID	87-69-4	C(C(C(=O)O)O)(C(=O)O)O	mono-ingredient	ACTIVE COMPOUND	150.09	-3.22	6.66 E+00	2.2	Non volatil	No binding	NR	NS	NR	NS	NS	2.7	NS	NS	Test
85	COUMARIN	91-64-5	c1ccc\2c(c1)OC(\C=C2)=O	mono-ingredient	FRAGRANCE INGREDIENT	146.14	1.41	3.51 E-02	8.9	Semi volatil	Michael Acceptor	R	NS	NR	S	S	20.7	NS	S	Training
86	6-METHYL COUMARIN	92-48-8	c12c(cc(cc1)C)\C=C/C(O2)=O	mono-ingredient	FRAGRANCE INGREDIENT	160.17	1.91	7.42 E-03	8.9	Semi volatil	Michael Acceptor	R	NS	NR	S	S	20.7	NS	NS	Training
87	THIOGLYCERIN	96-27-5	OCC(CS)O	mono-ingredient	ACTIVE COMPOUND	108.16	-0.78	5.30 E+00	6.9	Semi volatil	No binding	NR	S	R	S	NS	83.4	S	S	Training
88	DNCB	97-00-7	O=[N+]([O-])c1ccc(c(c1)[N+](=O)[O-])Cl	mono-ingredient	NON COSMETIC	202.55	2.34	1.20 E-03	8.7	Semi volatil	SNAr	R	INC	R	S	S	91.8	S	S	Training
89	ISOEUGENOL	97-54-1	c1c(cc(c(c1)O)OC)/C=C/C	mono-ingredient	FRAGRANCE INGREDIENT	164.20	2.83	1.01 E-03	7.3	Semi volatil	Michael Acceptor	R	S	R	S	S	95.6	S	S	Training
90	GLYCOL DIMETHACRYLATE	97-90-5	CC(C(OCCOC(=O)C(=C)C)=O)=C	mono-ingredient	ACTIVE COMPOUND	198.22	1.98	2.93 E-03	8.3	Very volatil	Michael Acceptor	R	S	R	S	S	95.9	S	S	Test
91	diethyl acetaldehyde	97-96-1	C(C(C=O)CC)C	mono-ingredient	NON COSMETIC	100.16	1.89	3.92 E-02	3.0	Very volatil	Schiff base formation	R	S	R	NS	S	94.2	S	S	Training
92	4-HYDROXYBENZOIC ACID	99-96-7	c1cc(ccc1O)C(=O)O	mono-ingredient	PRESERVATIVE	138.12	1.56	1.05 E-01	3.1	Semi volatil	No binding	NR	NS	NR	NS	NS	2.8	NS	NS	Training
93	OA25	OA25	NA	complex ingredient	FATTY COMPOUND	NA	NA	NA	8.4	NA	NA	NA	NA	NR	S	INC	14.5	NS	NS	Training

94	OA26	OA26	OA26	mono-ingredient	ACTIVE COMPOUND	486.58	-6.18	NA	5.9	NA	NA	NA	NA	INC	NS	NS	4.4	NS	NS	Training
95	OA27	OA27	OA27	mono-ingredient	FRAGRANCE INGREDIENT	164.20	1.22	8.20 E-02	7.4	Semi volatil	No binding	NR	NS	INC	S	S	7.8	NS	NS	Test
96	METHYLDIBROMO GLUTARONITRILE	35691-65-7	N#CC(CCC#N)(Br)CBr	mono-ingredient	PRESERVATIVE	265.90	1.42	1.59 E-03	8.0	Semi volatil	Michael Acceptor SN2	R	S	R	S	S	95.5	S	S	Training
97	OA28	OA28	NA	complex ingredient	POLYMER	NA	NA	NA	4.3	NA	NA	NA	NA	R	NS	No data	27.3	NS	NS	Training
98	GLUTARAL	111-30-8	O=CCCCC=O	mono-ingredient	PRESERVATIVE	100.12	-0.17	1.67 E+00	4.1	Very volatil	Schiff base formation	R	S	R	S	S	97.9	S	S	Training
99	OA29	OA29	NA	complex ingredient	POLYMER	NA	NA	NA	4.4	NA	NA	NA	NA	R	NS	NS	12.3	NS	NS	Test
100	OA30	OA30	NA	complex ingredient	FATTY COMPOUND	NA	NA	NA	6.5	NA	NA	NA	NA	NR	S	NS	17.4	NS	NS	Training
101	OA31	OA31	NA	complex ingredient	FATTY COMPOUND	NA	NA	NA	8.3	NA	NA	NA	NA	NR	NS	No data	3.9	NS	NS	Test
102	HC RED No. 7	24905-87-1	c1cc(c(cc1NCCO)[N+](=O)[O-])N	mono-ingredient	COLOR ADDITIVE	197.19	0.76	1.79 E-01	7.4	Non volatil	Michael Acceptor	R	S	R	S	S	89.8	S	S	Training
103	HC BLUE No.2	33229-34-4	c1(N(CCO)CCO)cc(c(cc1)NCCO)[N+](=O)[O-]	mono-ingredient	COLOR ADDITIVE	285.30	0.21	9.79 E-02	6.4	Non volatil	Michael Acceptor	R	S	R	NS	NS	21.4	NS	S	Training
104	HC VIOLET No.2	104226-19-9	c1(N(CCO)CCO)cc(c(cc1)NCCCO)[N+](=O)[O-]	mono-ingredient	COLOR ADDITIVE	299.33	0.58	2.95 E-02	8.7	Non volatil	Michael Acceptor	R	S	R	S	S	88.6	S	S	Training
105	HC VIOLET No.1	82576-75-8	c1(c(cc(c(c1)NCCO)C)N)[N+](=O)[O-]	mono-ingredient	COLOR ADDITIVE	211.22	1.26	4.83 E-02	7.8	Non volatil	Michael Acceptor	R	S	R	NS	S	43.3	Equivocal	S	Test
106	4-AMINO-3-NITROPHENOL	610-81-1	c1cc(c(cc1O)[N+](=O)[O-])N	mono-ingredient	DYE/PIGMENT	154.12	1.35	4.57 E-02	7.2	Semi volatil	Michael Acceptor	R	S	R	S	S	95.6	S	S	Training

107	3-NITRO-p-HYDROXYETHYLAMINO PHENOL	65235-31-6	c1(c(cc(cc1)O)[N+](=O)[O-])NCCO	mono-ingredient	DYE/PIGMENT	198.18	0.94	1.48E-01	7.3	Semi volatil	Michael Acceptor	R	S	R	S	S	95.6	S	S	Test
108	HC ORANGE No.2	85765-48-6	c1(c(cc(cc1)OCCO)[N+](=O)[O-])NCCN	mono-ingredient	COLOR ADDITIVE	241.25	0.69	8.38E-01	9.3	Non volatil	Schiff base formation	R	S	R	S	S	75.4	S	S	Training
109	3-METHYLAMINO-4-NITROPHENOXYETHANOL	59820-63-2	c1cc(c(cc1OCCO)NC)[N+]([O-])=O	mono-ingredient	DYE/PIGMENT	212.20	1.38	1.95E-02	8.4	Semi volatil	No binding	NR	INC	NR	NS	S	8.3	NS	NS	Training
110	2-NITRO-5-GLYCERYL METHYLANALINE	80062-31-3	c1cc(c(cc1OCC(CO)O)NC)[N+]([O-])=O	mono-ingredient	DYE/PIGMENT	242.23	0.56	1.89E-02	8.6	Non volatil	No binding	NR	INC	R	NS	S	6.0	NS	NS	Training
111	HC YELLOW No.7	104226-21-3	c1(N(CCO)CCO)cc(c(cc1)\N=N\c1ccc(cc1)N)C	mono-ingredient	DYE/PIGMENT	314.39	2.67	1.11E-04	7.9	Non volatil	Michael Acceptor	R	S	R	S	NS	82.4	S	NS	Training
112	DIHYDROXYINDOLE	3131-52-0	c12c(cc(c(c1)O)O)[nH]cc2	mono-ingredient	DYE/PIGMENT	149.15	0.87	2.27E-01	6.7	Semi volatil	Michael Acceptor	R	INC	R	S	NS	83.3	S	S	Test
113	2-HYDROXYETHYLAMINO-5-NITROANISOLE	66095-81-6	c1cc(cc(c1NCCO)OC)[N+](=O)[O-]	mono-ingredient	DYE/PIGMENT	212.20	1.34	1.95E-02	8.2	Semi volatil	No binding	NR	INC	NR	NS	NS	5.5	NS	NS	Test
114	OA32	OA32	OA32	mono-ingredient	DYE/PIGMENT	266.73	0.42	4.75E-02	3.2	Non volatil	Michael Acceptor	R	S	INC	S	S	94.9	S	S	Training
115	HC YELLOW No.10	109023-83-8	c1(cc(c(cc1NCCO)NCCO)[N+](=O)[O-])Cl	mono-ingredient	COLOR ADDITIVE	275.69	1.20	4.13E-02	8.0	Non volatil	Michael Acceptor	R	INC	INC	S	NS	26.0	NS	NS	Training
116	OA33	OA33	OA33	mono-ingredient	DYE/PIGMENT	181.14	-0.31	1.84E+00	2.1	Semi volatil	Michael Acceptor	R	S	R	S	S	97.9	S	S	Training
117	TOLUENE-2,5-DIAMINE SULFATE	615-50-9	S(O)(O)(=O)=O.c1(c(ccc(c1)N)N)C	mono-ingredient	DYE/PIGMENT	220.25	0.19	5.93E-01	2.5	Semi volatil	Michael Acceptor	R	S	R	S	S	97.9	S	S	Training
118	OA34	OA34	OA34	mono-ingredient	DYE/PIGMENT	282.17	-0.90	1.30E-01	2.1	Non volatil	Michael Acceptor	R	INC	INC	S	S	89.6	S	S	Training

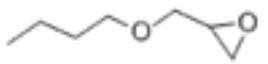
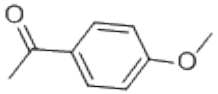
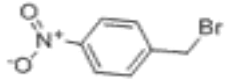
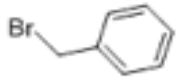
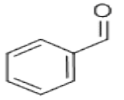
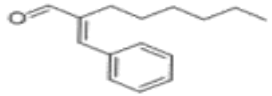
119	OA35	OA35	OA35	mono-ingredient	DYE/PIGMENT	269.17	-0.40	4.10E-01	2.0	Non volatil	Michael Acceptor	R	S	R	S	S	95.9	S	S	Training
120	N,N-Bis(2-HYDROXYETHYL)-p-PHENYLENEDIAMINE SULFATE	54381-16-7	c1(N(CCO)CCO)ccc(cc1)N.S(O)(O)(=O)=O.O	mono-ingredient	DYE/PIGMENT	312.35	-0.40	4.10E-01	2.2	Non volatil	Michael Acceptor	R	S	R	S	S	95.8	S	S	Training
121	OA36	OA36	OA36	mono-ingredient	DYE/PIGMENT	235.16	1.19	6.33E-03	2.0	Semi volatil	Michael Acceptor	R	S	R	S	S	97.9	S	S	Test
122	m-AMINOPHENOL	591-27-5	c1ccc(cc1O)N	mono-ingredient	DYE/PIGMENT	109.13	0.25	6.64E-01	7.3	Semi volatil	Michael Acceptor	R	INC	NR	S	NS	49.6	Equivocal	S	Training
123	2-METHYL-5-HYDROXYETHYLAMINOPHENOL	55302-96-0	c1(cc(c(cc1)C)O)NCCO	mono-ingredient	DYE/PIGMENT	167.21	0.71	1.85E-01	7.3	Semi volatil	Michael Acceptor	R	S	R	S	S	95.6	S	NS	Training
124	2,4-DIAMINOPHENOXYETHANOL HCl	66422-95-5	c1(c(cc(cc1)N)N)OCCO.Cl.Cl	mono-ingredient	DYE/PIGMENT	241.12	-0.87	4.65E+00	2.2	Semi volatil	Michael Acceptor	R	INC	R	S	S	96.0	S	S	Training
125	HC BLUE No.14	99788-75-7	c12c(c(ccc1NCC(CO)O)NCC(CO)O)C(c1c(C2=O)cccc1)=O	mono-ingredient	COLOR ADDITIVE	386.40	2.26	2.21E-04	6.6	Non volatil	Michael Acceptor	R	NS	R	S	S	28.7	NS	NS	Training
126	PHENYL METHYL PYRAZOLONE	89-25-8	N1(c2ccccc2)/N=C(\CC1=O)/C	mono-ingredient	DYE/PIGMENT	174.20	1.33	2.19E-03	5.3	Semi volatil	acyl Transfer agent	R	S	NR	S	S	96.0	S	S	Training
127	ACID VIOLET 43	4430-18-6	c12c(C(c3c(C1=O)cccc3)=O)c(ccc2Nc1c(cc(cc1)C)S(=O)(=O)[O-])O.[Na+]	mono-ingredient	COLOR ADDITIVE	432.42	3.60	1.19E-08	11.2	Non volatil	Michael Acceptor	R	INC	INC	S	NS	20.7	NS	NS	Training
128	1,2,4-TRIHYDROXYBENZENE	533-73-3	c1(c(ccc(c1)O)O)O	mono-ingredient	DYE/PIGMENT	126.11	0.21	9.69E-01	4.0	Semi volatil	Michael Acceptor	R	INC	R	S	S	97.0	S	S	Test
129	OA37	OA37	OA37	mono-ingredient	SURFACTANT	468.50	-0.98	5.74E-03	4.7	Non volatil	No binding	NR	INC	NR	NS	NS	3.4	NS	NS	Training
130	ISATIN	91-56-5	c12c(C(C(N1)=O)=O)cccc2	mono-ingredient	DYE/PIGMENT	147.13	0.83	1.04E-01	8.0	Semi volatil	acyl Transfer agent	R	INC	R	S	S	92.2	S	S	Test

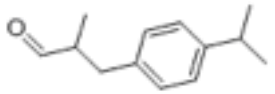
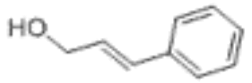
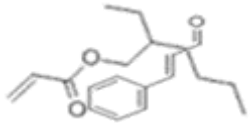
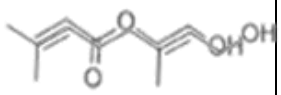
131	OA38	OA38	OA38	mono-ingredient	ACTIVE COMPOUND	181.19	1.47	1.87 E-02	7.4	Semi volatil	Michael Acceptor Acyl Transfert agent	R	INC	NR	S	S	78.0	S	NS	Training
132	OA39	OA39	OA39	mono-ingredient	ACTIVE COMPOUND	440.58	6.08	1.37 E-05	8.1	Non volatil	No binding	NR	INC	R	S	S	32.3	Equivocal	NS	Training
133	OA40	OA40	OA40	mono-ingredient	ACTIVE COMPOUND	741.12	13.47	4.98 E-18	8.3	Non volatil	No binding	NR	INC	NR	NS	NS	2.7	NS	NS	Test
134	OA41	OA41	OA41	mono-ingredient	DYE/PIGMENT	372.81	-3.11	(approx.) 6,09 E-03	2.0	Non volatil	Michael Acceptor	R	S	R	S	S	95.9	S	S	Training
135	OA42	OA42	OA42	mono-ingredient	DYE/PIGMENT	266.13	-1.97	4.03 E-01	3.1	Semi volatil	Michael Acceptor	R	INC	INC	S	S	95.7	S	S	Training
136	OA43	OA43	OA43	mono-ingredient	DYE/PIGMENT	288.82	-0.81	(approx.) 1,62 E-03	7.4	Semi volatil	Michael Acceptor	R	INC	INC	INC	S	83.1	S	S	Training
137	OA44	OA44	OA44	mono-ingredient	DYE/PIGMENT	378.49	-0.81	(approx.) 1,62 E-03	8.1	Semi volatil	Michael Acceptor	R	INC	INC	S	S	85.5	S	NS	Test
138	OA45	OA45	OA45	mono-ingredient	DYE/PIGMENT	229.71	1.99	3.84 E-03	3.3	Semi volatil	No binding	NR	INC	R	S	No data	90.5	S	S	Test
139	OA46	OA46	OA46	mono-ingredient	DYE/PIGMENT	336.03	-1.20	6.40 E+00	1.7	Semi volatil	No binding	NR	INC	R	S	S	95.1	S	S	Training
140	OA47	OA47	OA47	mono-ingredient	DYE/PIGMENT	247.13	-1.20	6.40 E+00	1.7	Semi volatil	No binding	NR	INC	R	S	S	95.1	S	S	Test
141	OA48	OA48	OA48	mono-ingredient	DYE/PIGMENT	227.74	0.44	2.82 E-01	1.6	Semi volatil	Michael Acceptor	R	S	R	S	S	97.9	S	S	Test
142	OA49	OA49	OA49	mono-ingredient	DYE/PIGMENT	227.65	1.20	2.02 E-02	1.7	Semi volatil	No binding	NR	INC	R	S	S	95.1	S	S	Training

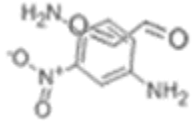
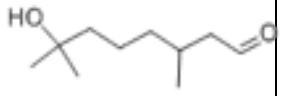
143	FORMALDEHYDE	50-00-0	C=O	mono-ingredient	PRESERVATIVE	30.00	0.35	1.90 E+00	8.0	Very volatil	Schiff base formation	R	S	R	S	S	95.9	S	S	Test
144	OA50	OA50	OA50	mono-ingredient	DYE/PIGMENT	386.33	-3.18	(approx.) 1,06 E-02	2.4	Non volatil	Michael Acceptor	R	S	R	S	S	95.8	S	S	Test
145	OA51	OA51	OA51	mono-ingredient	ACTIVE COMPOUND	223.23	1.85	1.57 E-02	6.9	Semi volatil	No binding	NR	INC	NR	S	NS	24.1	NS	NS	Training
146	OA52	OA52	OA52	mono-ingredient	DYE/PIGMENT	315.25	-3.58	(approx.) 2,09 E-01	3.1	Non volatil	Michael Acceptor	R	S	R	S	S	95.7	S	S	Training
147	OA53	OA53	OA53	mono-ingredient	DYE/PIGMENT	264.20	0.44	2.82 E-01	1.5	Semi volatil	Michael Acceptor	R	S	R	S	S	97.9	S	S	Training
148	OA54	OA54	OA54	mono-ingredient	ACTIVE COMPOUND	265.70	2.73	3.23 E-04	8.2	Semi volatil	Acyl Transfer agent	R	INC	NR	S	S	76.4	S	S	Test
149	OA55	OA55	NA	complex ingredient	POLYMER	NA	NA	NA	9.0	NA	NA	NA	NA	NR	NS	NS	3.3	NS	NS	Training
150	OA56	OA56	OA56	mono-ingredient	DYE/PIGMENT	315.25	-3.58	(approx.) 2,10 E-01	3.6	Non volatil	Michael Acceptor	R	INC	R	S	S	90.5	S	S	Test
151	OA57	OA57	OA57	mono-ingredient	ACTIVE COMPOUND	233.36	2.66	2.04 E-03	7.2	Semi volatil	Acyl Transfer agent	R	INC	INC	NS	NS	12.1	NS	S	Test
152	2,3-DIAMINODIHYDROPYRAZOLO PYRAZOLONE DIMETHOSULFONATE	857035-95-1	C/1(=C(/C(N2N1CCC2)=O)\N)\N.S(O[H])(=O)(=O)C	mono-ingredient	DYE/PIGMENT	346.39	-1.53	6.49 E+00	1.8	Semi volatil	SN2	R	INC	R	NS	S	84.5	S	NS	Test
153	OA58	OA58	OA58	mono-ingredient	PRESERVATIVE	351.96	5.85	3.07 E-05	4.7	Semi volatil	No binding	NR	INC	NR	S	S	76.1	S	S	Test
154	OA59	OA59	OA59	mono-ingredient	ACTIVE COMPOUND	323.24	3.48	1.53 E-04	8.8	Non volatil	Acyl Transfer agent	R	INC	R	S	S	64.2	Equivocal	NS	Training

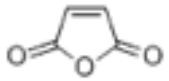
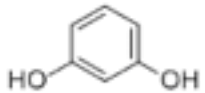
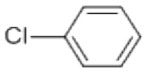
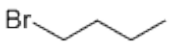
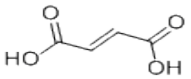
155	OA60	OA60	OA60	mono-ingredient	DYE/PIGMENT	322.80	-2.52	(approx.) 5,64 E-04	8.0	Semi volatil	No binding	NR	INC	INC	S	NS	53.4	Equivocal	NS	Test
156	OA61	OA61	OA61	mono-ingredient	DYE/PIGMENT	196.25	-0.05	3.19 E-01	1.6	Semi volatil	SN2	R	INC	R	NS	S	84.9	S	NS	Training
157	OA62	OA62	OA62	mono-ingredient	DYE/PIGMENT	241.72	2.36	1.22 E-04	3.1	Semi volatil	No binding	NR	INC	R	S	S	94.7	S	S	Training
158	OA63	OA63	OA63	mono-ingredient	DYE/PIGMENT	342.70	-3.49	(approx.) 1,21 E-01	1.8	Non volatil	Schiff base formation	R	INC	R	S	S	92.3	S	S	Training
159	OA64	OA64	OA64	mono-ingredient	DYE/PIGMENT	264.30	-0.97	2.90 E+00	2.2	Semi volatil	SN2	R	INC	R	NS	S	83.8	S	S	Training
160	OA65	OA65	OA65	mono-ingredient	DYE/PIGMENT	168.16	-0.98	2.66 E+00	7.5	Semi volatil	SN2 Schiff base formation	R	INC	R	NS	S	42.2	Equivocal	NS	Training
161	OA66	OA66	OA66	mono-ingredient	ACTIVE COMPOUND	349.35	-0.06	4.38 E-03	8.3	Non volatil	No binding	NR	INC	INC	S	NS	9.6	NS	NS	Training
162	OA67	OA67	OA67	mono-ingredient	DYE/PIGMENT	356.72	-3.46	(approx.) 3,65 E-02	1.9	Non volatil	Schiff base formation	R	INC	R	S	S	92.2	S	S	Test
163	OA68	OA68	OA68	mono-ingredient	DYE/PIGMENT	227.69	1.83	4.05 E-04	3.1	Semi volatil	No binding	NR	INC	R	S	S	94.7	S	S	Test
164	OA69	OA69	OA69	mono-ingredient	DYE/PIGMENT	329.23	-3.79	2.90 E-01	4.0	Non volatil	No binding	NR	INC	R	S	S	71.6	S	S	Training
165	OA70	OA70	OA70	mono-ingredient	DYE/PIGMENT	360.33	-1.57	(approx.) 1,28 E-03	4.8	Non volatil	No binding	NR	INC	INC	S	S	53.2	Equivocal	S	Test

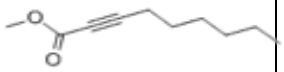
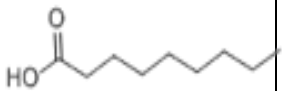
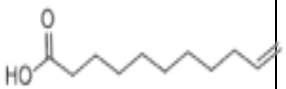
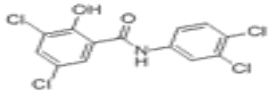
Appendix V.1: ICCVAM IDS Database

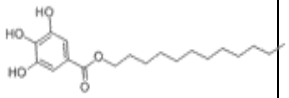
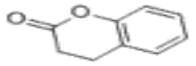
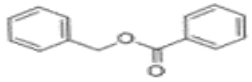
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	CASRN	LLNA Result (Majority)	h-CLAT Result (Majority)	OECD QSAR Toolbox Read-across Result	MW (g/mol)	Log P	Log S mol/L	Log VP mm Hg	MP °C	BP °C	DPRA Result (Majority)	KeratinoSens Result (Majority)	Avg % Depletion Lys (0.5mM:2.5mM [1:50])	Avg % Depletion Cys (0.5mM:5mM [1:10])	Avg % Depletion Lys & Cys	Pre/Pro-hapten?	Pre/Pro-hapten Reference	Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)	Product Class	Structural Group (for Training and Test Set Construction)	Structure	Test or Training Set
Butyl glycidyl ether	2426-08-6	POS	NEG	POS	130.185	0.63	-0.814	0.5051	-31	165	POS	POS	11.8	67.3	39.6	N	Ashikaga 2010	SN2 >> Ring opening SN2 reaction >> Epoxides, Aziridines and Sulfuranes	Intermediate in chemical synthesis; Manufacturing	Other		Training
4-Methoxyacetophenone	100-06-1	NEG	NEG	POS	150.17	1.74	0.393	-2.1911	38.5	258	NEG	POS	-0.1	1.6	0.7	NA	Ashikaga 2010; Natsch et al. 2013	No parent alert; Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes for product	Intermediate in chemical synthesis	Other		Training
4-Nitrobenzyl bromide	100-11-8	POS	POS	POS	216.032	2.7	-1.139	-3.0101	99	281.7	POS	POS	24.2	100	62.1	N	Ashikaga 2010	SN2 >> Nucleophilic substitution on benzilyc carbon atom >> alpha-Activated benzyls	Intermediate in chemical synthesis	Nitrobenzenes		Test
Benzyl bromide	100-39-0	POS	POS	POS	171.034	2.92	-0.415	-0.3468	-3	201	POS	POS	22.1	100	62	N	Ashikaga 2010	SN2 >> Nucleophilic substitution on benzilyc carbon atom >> alpha-Activated benzyls	Intermediate in chemical synthesis; Manufacturing	Other		Training
Benzaldehyde	100-52-7	NEG	POS	NEG	106.12	1.48	-1.208	0.1038	-26	179	NEG	POS	0	7.2	3.6	NA	Ashikaga 2010; Natsch et al. 2013	No alert	Intermediate in chemical synthesis; Manufacturing; Pesticides; Pharmaceuticals; Solvent	Other		Training
Hexyl cinnamic aldehyde	101-86-0	POS	NEG	POS	216.319	4.82	-2.561	-3.2725	39.2	318.7	NEG	POS	-0.45	6.2	2.9	N	Troutman et al. 2011	Michael Addition >> alpha,beta-Unsaturated carbonyl compounds >> alpha,beta-Aldehydes	Food additive; Fragrance agent	Benzaldehydes with unsaturated aliphatic chains		Training
2-Ethylhexylacrylate	103-11-7	POS	POS	POS	184.28	4.09	-1	-0.7496	-90	213.5	POS	POS	34.1	98.8	66.45	U	NA	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Manufacturing; Plastics	Acrylates		Test

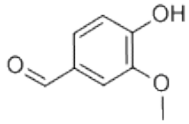
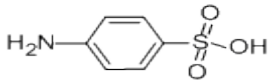
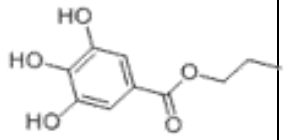
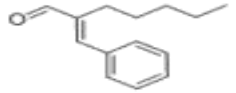
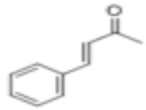
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Cyclamen aldehyde	103-95-7	POS	NEG	POS	190.29	3.4	0	-2.6478	16.77	234	POS	POS	1	18.9	9.95	N	Ashikaga 2010	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Food additive; Fragrance agent	Benzyl aldehydes with longer unsaturated aliphatic chains		Training
Cinnamyl alcohol	104-54-1	POS	POS	POS	134.18	1.95	0.0792	-1.6198	33	250	POS	POS	6.4	10.7	9.6	Pro	Natsch and Haupt 2013; Gerberick et al. 2009	No parent alert; Michael Addition >> alpha,beta-Unsaturated carbonyl compounds >> alpha,beta-Aldehydes for product AND Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes for product	Cosmetics; Food additive; Fragrance agent; Intermediate in chemical synthesis; Personal care products	Other		Training
Cinnamic aldehyde	104-55-2	POS	POS	POS	132.159	1.9	-1.969	-1.539	-8	246	POS	POS	46.6	68.8	57.8	N	Ball et al. 2011	Michael Addition >> alpha,beta-Unsaturated carbonyl compounds >> alpha,beta-Aldehydes	Cosmetics; Food additive; Fragrance agent; Intermediate in chemical synthesis; Personal care products; Pesticides	Benzaldehydes with unsaturated aliphatic chains		Training
Geraniol	106-24-1	POS	POS	POS	154.249	3.56	-1	-1.5228	-15	230	NEG	POS	10	0	2.5	Pro	Natsch and Haupt 2013; Gerberick et al. 2009	No parent alert; Schiff base formation >> Direct acting Schiff base formers >> Di-substituted alpha,beta-unsaturated aldehydes AND Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes AND SN2 >> Ring opening SN2 reaction >> Epoxides, Aziridines and Sulfuranes for product	Cosmetics; Fragrance agent; Personal care products; Pesticides	Unsaturated aliphatic alcohols		Training

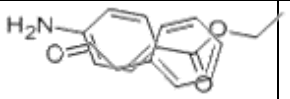
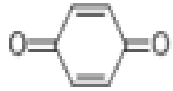
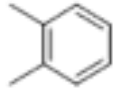
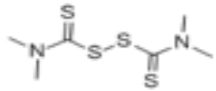
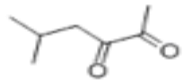
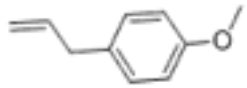
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
4-Phenylenediamine	106-50-3	POS	POS	POS	108.141	-0.3	-0.466	-2.301	146	267	POS	POS	26.5	93	58.3	Pre	Natsch and Haupt 2013; Gerberick et al. 2009	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines for product	Manufacturing; Personal care products	Benzyl diamines		Training
Benzoquinone	106-51-4	POS	POS	POS	108.095	0.2	-0.989	-1.0458	116	180	POS	POS	91	99	95	N	Troutman et al. 2011	Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines	Manufacturing; Pesticides; Pharmaceuticals	Other		Training
Ethylenediamine	107-15-3	POS	POS	POS	60.0984	-2.04	3	1.0792	11.1	117	POS	POS	0.7	18.6	9.65	Pro	Natsch and Haupt 2013	No parent alert; Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes for product	Manufacturing	Aliphatic amines		Test
Glyoxal	107-22-2	POS	POS	POS	58.0361	-1.66	3	2.4065	15	50.4	POS	POS	67.8	56.5	62.15	N	Troutman et al. 2011	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Intermediate in chemical synthesis; Manufacturing; Pharmaceuticals	Ketones with unsaturated carbon chains		Training
Hydroxycitronellal	107-75-5	POS	POS	POS	172.26	2.11	0.481	-2.2366	23	241	POS	POS	14.6	32.9	23.8	N	Troutman et al. 2011	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Food additive; Fragrance agent; Personal care products	Other		Training

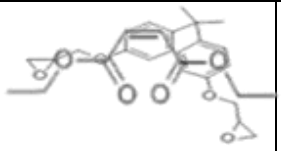
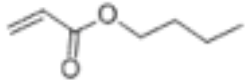
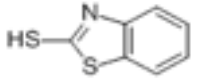
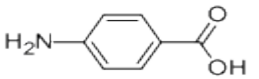
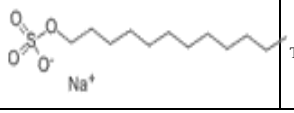
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Maleic anhydride	108-31-6	POS	POS	POS	98.06	1.62	0.691	-0.6021	52.8	202	POS	POS	41	100	59.4	N	Lalko et al. 2013	Acylation >> Direct acylation involving a leaving group >> Anhydrides (sulphur analogues of anhydrides)	Intermediate in chemical synthesis; Manufacturing; Pesticides; Pharmaceuticals	Other		Training
Resorcinol	108-46-3	POS	POS	POS	110.111	0.8	0.814	-3.3107	111	280	NEG	NEG	0	1.6	0.8	Pro	Natsch and Haupt 2013	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/ imines for product	Cosmetics; Manufacturing; Personal care products; Pharmaceuticals	Dihydro quinones		Test
Chlorobenzene	108-90-7	NEG	POS	NEG	112.557	2.84	-2.354	1.0781	-45.2	131.7	NEG	NEG	1.3	0.4	0.85	NA	Troutman et al. 2011	No alert	Manufacturing; Pesticides; Solvent	Other		Training
3-Dimethylaminopropylamine	109-55-7	POS	POS	POS	102.18	-0.45	3	1	-60	133	NEG	POS	0	10.2	4.5	Pro	Natsch and Haupt 2013	No alert	Intermediate in chemical synthesis	Aliphatic amines		Training
1-Bromobutane	109-65-9	NEG	POS	NEG	137.018	2.75	-2.198	1.6229	-112.4	101.3	POS	NEG	1.2	13.8	7.5	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Manufacturing; Pharmaceuticals	Other		Training
Fumaric acid	110-17-8	NEG	NEG	NEG	146.141	0.46	0.845	-3.8125	287	522	POS	NEG	4.6	10.8	7.7	NA	Bauch et al. 2012	No alert	Food additive; Household products; Manufacturing	Other		Training
Hexane	110-54-3	NEG	NEG	NEG	86.1754	3.9	-3.958	2.1798	-95.3	68.7	NEG	POS	1.3	-0.5	0.43	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Manufacturing; Solvent	Other		Training

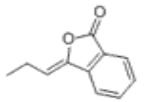
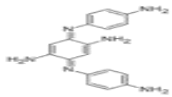
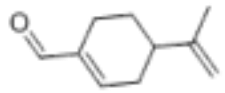
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Pyridine	110-86-1	POS	POS	NEG	79.0999	0.65	1.102	1.3181	-41.6	115.2	NEG	NEG	-1.1	0.7	-0.2	Pro	Jaworska et al. 2011	No alert	Food additive; Intermediate in chemical sysnthesis	Other		Training
1-Bromohexane	111-25-1	POS	NEG	POS	165.08	3.8	-3.808	0.5906	-84.7	155.3	POS	POS	68.2	14.1	41.15	N	Ashikaga 2010	SN2 >> Nucleophilic SN2 reaction at sp3 carbon atom >> Alkyl halides	Solvent	Other		Training
Glutaraldehyde	111-30-8	POS	POS	POS	100.12	-0.18	2.223	-0.2218	-14	188	POS	POS	85.4	30.2	57.8	N	Lalko et al. 2013	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Cosmetics; Manufacturing; Personal care products; Pesticides (antimicrobial)	Other		Training
Methyl 2-nonynoate	111-80-8	POS	POS	POS	168.24	3.1	-0.846	-1.1255	38	224.5	POS	POS	3.2	100	51.6	N	Ashikaga 2010	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized triple bond	Food additive; Fragrance agent	Other		Training
Nonanoic acid	112-05-0	POS	POS	NEG	158.24	3.42	-2.746	-2.7825	12.3	254.5	NEG	NEG	0	0	0	N	Troutman et al. 2011; Natsch et al. 2013	No alert	Intermediate in chemical synthesis; Manufacturing; Pesticides (antimicrobial)	Aliphatic carboxylic acids		Training
Undecylenic acid	112-38-9	POS	NEG	NEG	184.277	3.86	-3.398	-3.0283	24.5	275	POS	POS	-0.1	11.4	5.75	N	Bauch et al. 2012	No alert	Cosmetics; Pharmaceuticals; Pesticides (antimicrobial)	Aliphatic carboxylic acids		Test
3,3,4,5-Tetrachlorosalicylanilide	1154-59-2	POS	POS	POS	351.02	5.87	-3.796	-9.5229	161	488.1	POS	POS	9	36.8	22.9	N	Nishijima et al. 1999	Acylation >> Ester aminolysis >> Amides	Pesticide (antifungal)	Other		Training

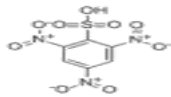
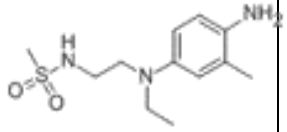
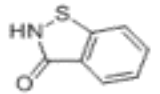
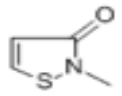
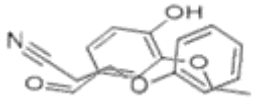
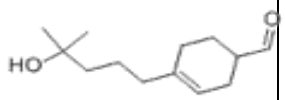
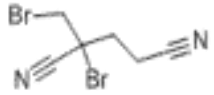
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Lauryl gallate	1166-52-5	POS	POS	NEG	338.44	6.21	-3.863	-8.9914	96.5	468	POS	POS	8.7	90.9	49.8	Pre/ Pro	Troutman et al. 2011	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines for product	Antioxidant	Trihydroxy lated benzyl esters		Test
Methyl salicylate	119-36-8	NEG	NEG	NEG	152.147	2.55	-1.313	-1.4647	-8	222.9	NEG	NEG	1.6	0.4	3.15	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Cosmetics; Food additive; Fragrance agent; Personal care products; Pharmaceuticals; Solvent; Pesticides (antimicrobial)	Phenolic esters		Training
3,4-Dihydrocoumarin	119-84-6	POS	POS	POS	148.16	0.97	0.477	-2.0825	25	272	POS	NEG	39.7	0	19.85	Pro	Gerberick et al. 2004	Acylation >> Ester aminolysis or thiolysis >> Activated aryl esters	Food additive; Fragrance agent; Pharmaceuticals	Benzopyran [2-ring heterocyclics with oxygen]		Training
Benzyl benzoate	120-51-4	POS	NEG	POS	212.25	3.97	-1.812	-3.6498	21	323.5	NEG	POS	3	0.2	1.6	U		SN2 >> SN2 Reaction at a sp3 carbon atom >> Activated alkyl esters and thioesters	Food additive; Fragrance agent; Manufacturing; Pharmaceuticals; Pesticides	Benzyl esters		Training
Ethyl vanillin	121-32-4	NEG	NEG	POS	166.18	1.58	0.45	-4.9842	77.5	294	NEG	POS	9.7	1.1	5.4	NA	Ashikaga 2010; Natsch et al. 2013	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines AND Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes for product	Food additive; Fragrance agent; Manufacturing	Vanillin		Training

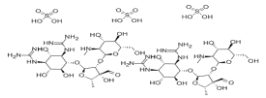
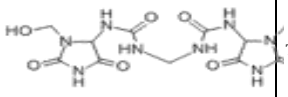
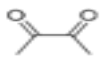
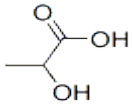
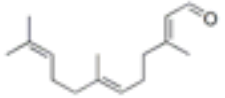
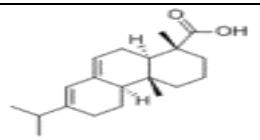
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Vanillin	121-33-5	NEG	NEG	POS	152.15	1.21	-1.141	-3.9274	81.5	285	POS	POS	9.9	0.9	3.9	NA	Ashikaga 2010; Natsch et al. 2013	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines AND Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes for product	Food additive; Fragrance agent	Vanillin		Test
Sulfanilic acid	121-57-3	NEG	NEG	NEG	173.191	-2.16	1.033	-6.6946	288	300	NEG	NEG	0.5	5.3	2.9	NA	Natsch et al. 2013	No alert	Pharmaceuticals	Amino benzoic acid or ester		Training
Propyl gallate	121-79-9	POS	POS	NEG	212.2	1.8	-1.783	-6.3893	130	363.6	POS	POS	41.1	59.5	51.5	N	Troutman et al. 2011	No alert	Food additive	Trihydroxy lated benzyl esters		Training
α-Amyl cinnamic aldehyde	122-40-7	POS	POS	POS	202.3	4.33	0.544	-3.3449	80	304.8	NEG	NEG	3.9	0.6	2.25	N	Troutman et al. 2011	Michael Addition >> alpha,beta-Unsaturated carbonyl compounds >> alpha,beta-Aldehydes	Food additive	Benzaldehydes with unsaturated aliphatic chains		Test
Benzylidene acetone	122-57-6	POS	POS	NEG	146.19	2.07	0.13	-1.9066	42	261	POS	POS	0.1	93.2	46.7	N	Troutman et al. 2011	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Food additive; Manufacturing	Other		Training

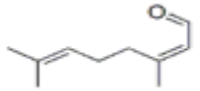
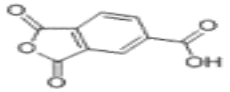
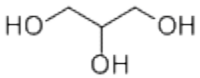
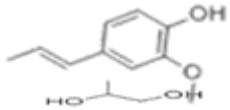
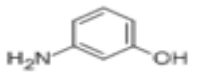
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Phenylacetaldehyde	122-78-1	POS	POS	POS	120.15	1.78	0.481	-0.4067	33.5	195	POS	POS	22.6	60.7	41.65	N	Ashikaga et al. 2010	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Food additive; Fragrance agent	Benzaldehydes with short unsaturated aliphatic chains		Test
1,4-Dihydroquinone	123-31-9	POS	POS	POS	110.11	0.59	-0.184	-4.6198	172.3	287	POS	POS	51.1	83.3	67.2	Pre	Troutman et al. 2011; Gerberick et al. 2009	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines for product	Manufacturing; Personal care products	Dihydro quinones		Training
Octanoic acid	124-07-2	NEG	POS	NEG	144.21	3.05	-2.262	-2.4306	16.3	239	NEG	NEG	0.9	0	0.45	N	Troutman et al. 2011; Natsch et al. 2013	No alert	Cosmetics; Food additive; Pesticides (antimicrobial)	Other		Training
Xylene	1330-20-7	POS	NEG	NEG	106.078	3.16	-0.975	0.9025 46779	-47	138.5	NEG	NEG	0.2	0.4	0.3	N	Bauch et al. 2012	No alert	Pesticide (herbicide); Intermediate in chemical sysnthesis; Manufacturing	Other		Training
Tetramethylthiuram disulfide	137-26-8	POS	POS	POS	240.437	1.73	-1.523	-4.7632	155.6	129	POS	POS	6.9	99.5	53.2	N	Ashikaga 2010	Acylation >> Ester aminolysis >> Dithiocarbamates AND SN2 >> interchange with sulfur containing compounds >> Thiols and disulfide compounds	Pesticides (antimicrobial)	Other		Training
5-Methyl-2,3-hexanedione (Acetyl isovaleryl)	13706-86-0	POS	POS	POS	128.17	0.06	1.917	0.939	-16.7	138	POS	POS	7.5	25.8	16.65	U	NA	Schiff base formation >> Direct acting Schiff base formers >> 1,2-Dicarbonyls and 1,3-Dicarbonyls	Food additive	Diones		Training
4-Allylanisole	140-67-0	POS	POS	POS	148.202	3.47	-2.92	-0.78251 6056	-1.19	215.5	POS	POS	-1.3	45	21.8	Pro	Natsch et al. 2013	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime	Food additive; Fragrance agent	Other		Training

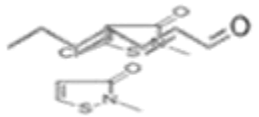
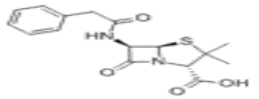
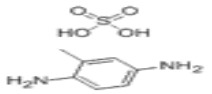
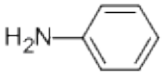
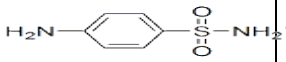
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
																		structure; Nitroquinones, Naphthoquinone(s)/imines for product				
Diethyl maleate	141-05-9	POS	POS	POS	172.18	2.2	1.146	-0.9788	-9	223	POS	POS	82.1	99.9	91	N	Troutman et al. 2011	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Food additive; Intermediate in chemical synthesis	Other		Training
Butyl acrylate	141-32-2	POS	POS	POS	128.17	2.36	-1.807	0.7366	-64.6	145	POS	POS	85.1	100	92.55	U	NA	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Intermediate in chemical synthesis; Manufacturing	Acrylates		Training
Oxalic acid	144-62-7	POS	POS	NEG	90.05	-2.22	2.342	-3.6308	190	246.9	NEG	POS	0	0.9	0.22	N	Ashikaga 2010	No alert	Household products; Intermediate in chemical synthesis; Manufacturing	Other		Training
2-Mercaptobenzothiazole	149-30-4	POS	POS	POS	167.253	2.42	-3.182	-3.3335	181	338.8	POS	POS	3.2	99.9	51.55	Pro	Bauch et al. 2011	Acylation >> Ester aminolysis >> Dithiocarbamates AND SN2 >> interchange with sulfur containing compounds >> Thiols and disulfide compounds	Manufacturing; Pesticides	Thiazoles [sulfur and nitrogen-containing heterocyclics]		Training
p-Aminobenzoic acid	150-13-0	NEG	NEG	NEG	137.14	0.83	-1.351	-3.556	188.5	307.7	NEG	NEG	0.25	4.2	2.3	NA	Bauch et al. 2012	No alert	Cosmetics; Food additive; Pharmaceuticals	Amino benzoic acid or ester		Test
Sodium lauryl sulfate	151-21-3	POS	NEG	NEG	288.38	1.6	2	-12.3279	205.5	216	POS	NEG	48.5	-0.1	24.5	N	Ashikaga 2010	No alert	Cosmetics; Food additive; Manufacturing; Personal care products; Pesticides	Other		Training

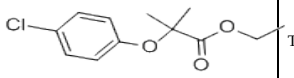
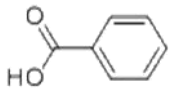
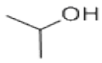
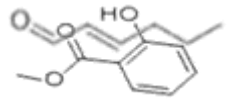
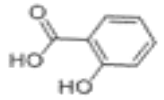
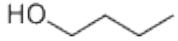
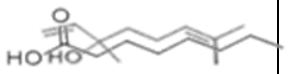
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
																			(antimicrobial); Pharmaceuticals			
Oxazolone	15646-46-5	POS	POS	POS	217.221	1.51	0.266	-5.1945	120	365.7	POS	POS	47	75.5	61.3	N	Troutman et al. 2011	Acylation >> Direct acylation involving a leaving group >> Azlactones and unsaturated lactone derivatives	Intermediate in chemical synthesis; Manufacturing	Other		Training
Bisphenol A diglycidyl ether	1675-54-3	POS	POS	POS	340.42	3.84	-3.155	-6.9586	12	421	POS	POS	1.1	42.5	21.8	U	NA	SN2 >> Ring opening SN2 reaction >> Epoxides, Aziridines and Sulfuranes	Manufacturing	Other		Training
3-Propylidenephthalide	17369-59-4	POS	POS	POS	174.2	2.03	0.036	-3.5243	66.8	318.2	POS	NEG	30.6	14.3	11.9	N	Ashikaga 2010	Acylation >> Direct acylation involving a leaving group >> Azlactones and unsaturated lactone derivatives AND Acylation >> Ring opening acylation >> Active cyclic agents	Food additive	Other		Training
Bandrowski's base	20048-27-5	POS	POS	POS	318.38	0.74	1.359	-10.3645	224.8	526.1	POS	POS	4.2	87.5	45.85	Pre/Pro	Troutman et al. 2011	Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines	Personal care products	Other		Training
Perillaldehyde	2111-75-3	POS	POS	POS	150.22	3.34	-0.794	-1.3344	218.8	240	POS	POS	13.8	31.9	22.85	U	NA	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Fragrance agent	Other		Training

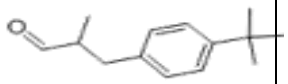
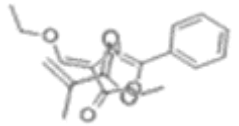
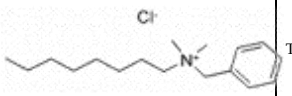
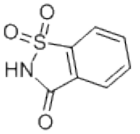
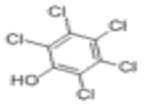
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
2,4,6-Trinitrobenzenesulfonic acid	2508-19-2	POS	POS	NEG	293.167694	-1.71	2.199	-11.284	206.9	487.7	POS	POS	22.6	99.7	61.15	N	Bauch et al. 2012	No alert	Intermediate in chemical synthesis	Other		Training
4-(N-Ethyl-N-2-methan-sulfamido-ethyl)-2-methyl-1,4-phenylenediamine	25646-71-3	POS	POS	POS	274.34	1.2	-3.25	-5.76	121.03	318.05	POS	POS	13.6	90.1	51.85	Pre/Pro	Troutman et al. 2011	No alert	Manufacturing	Benzyl diamines		Training
1,2-Benzisothiazolin-3-one	2634-33-5	POS	POS	POS	151.187	0.64	1.331	-4.5901	121.6	339.5	POS	POS	9.7	97.7	53.7	N	Ashikaga 2010	SN2 >> SN2 reaction at a sulphur atom >> Isothiazolidin-3-ones (sulphur) and Isothiazolone derivatives	Manufacturing; Pesticides (antimicrobial)	Thiazoles [sulfur and nitrogen-containing heterocyclics]		Training
Methylisothiazolinone	2682-20-4	POS	POS	NEG	115.15	-0.83	2.73	2.4914	47.5	237.8	POS	POS	-5.6	97.9	46.15	N	Ashikaga 2010	SN2 >> ring opening SN2 reaction >> Isothiazolone derivatives	Pesticides (antimicrobial)	Thiazoles [sulfur and nitrogen-containing heterocyclics]		Test
3-Phenoxypropiononitrile	3055-86-5	NEG	POS	POS	147.18	1.61	0.253	-2.0311	36.6	265.7	POS	NEG	29.7	0	14.85	NA	Troutman et al. 2011; Natsch et al. 2013	No parent alert; Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes for product	Manufacturing	Other		Training
3 and 4-(4-Hydroxy-4-methylpentyl)-3-cyclohexane-1-carbox-aldehyde [Lyral HMPCC]	31906-04-4	POS	POS	POS	210.32	3.32	-0.734	-4.067	-30	307.1	POS	POS	3.4	39.4	21.4	U	NA	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Fragrance agent	Other		Training
1,2-Dibromo-2,4-dicyanobutane	35691-65-7	POS	POS	NEG	265.933	1.63	0.114	-4.2989	52	327.8	POS	POS	28.5	99.5	52.1	Pro	Natsch and Haupt 2013; Gerberick et al. 2009	No alert	Manufacturing; Personal care products; Pesticides (antimicrobial)	Other		Training

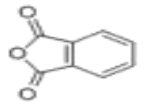
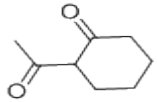
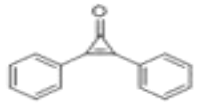
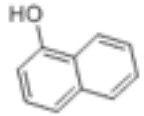
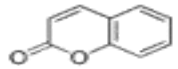
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Streptomycin sulfate	3810-74-0	NEG	NEG	POS	581.57	-3.61	-1.48	-11.48	123.57	326.63	NEG	NEG	4.5	0	2.25	N	Troutman et al. 2011; Natsch et al. 2013	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Pharmaceuticals; Pesticides	Other		Training
Imidazolidinyl urea	39236-46-9	POS	POS	NEG	388.294	-8.28	3	-28.4698003	240.5	932.2	POS	POS	15.2	54.1	29.7	N	Ashikaga 2010	No alert	Cosmetics; Personal care products; Pesticides	Other		Training
2,3-Butanedione	431-03-8	POS	POS	POS	86.0892	-1.34	0.366	1.7545	-1.2	88	POS	POS	30.7	85.6	58.2	N	Troutman et al. 2011	Schiff base formation >> Direct acting Schiff base formers >> 1,2-Dicarbonyls, 1,3-Dicarbonyls	Food additive	Diones		Test
Formaldehyde	50-00-0	POS	POS	POS	30.026	0.35	1.121	5.8899	-148.5	-19.1	POS	POS	2.4	44.7	24.45	N	Ashikaga 2010	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Manufacturing; Pesticides (antimicrobial)	Other		Training
Lactic acid	50-21-5	NEG	NEG	NEG	90.0779	-0.72	3	-1.0894	18	204.2	NEG	NEG	-0.2	1.5	0.6	NA	Gerberick et al. 2009; Natsch et al. 2013	No alert	Food additive; Pesticides (antimicrobial)	Other		Training
Farnesal	502-67-0	POS	POS	POS	220.36	5.74	-3.369	-2.7595	16.7	302.2	POS	POS	2.2	24.7	14.4	N	Troutman et al. 2011	Schiff base formation >> Direct acting Schiff base formers >> Di-substituted alpha,beta-unsaturated aldehydes	Food additive; Fragrance agent; Manufacturing	Ketones with unsaturated carbon chains		Training
Abietic acid	514-10-3	POS	NEG	NEG	302.46	6.46	-1.315	-6.4962	173.5	394.9	POS	POS	16.3	99.9	58.1	Pro	Natsch and Haupt 2013	No parent alert; Radical reactions >> Free radical formation >> Hydroperoxides for product	Manufacturing	Other		Training
2-Nitro-1,4-phenylenediamine	5307-14-2	POS	POS	POS	153.14	0.53	1.413	-4.2518	140	134	POS	POS	50	50	50	Pre/Pro	Troutman et al. 2011	No alert	Manufacturing	Benzyl diamines		Training

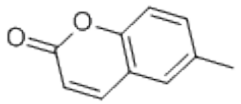
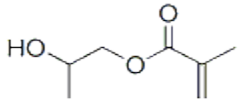
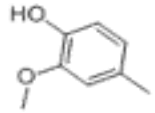
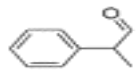
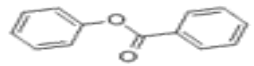
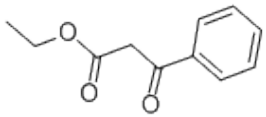
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Citral	5392-40-5	POS	POS	POS	152.233	3.45	-2.055	-1.0395	-10	227	POS	POS	12.1	82.1	47.3	N	Troutman et al. 2011	Schiff base formation >> Direct acting Schiff base formers >> Di-substituted alpha,beta-unsaturated aldehydes	Fragrance agent; Pesticides	Ketones with unsaturated carbon chains		Test
Trimellitic anhydride	552-30-7	POS	POS	POS	192.13	1.95	0.017	-5.0061	162	390	POS	NEG	43.7	0	21.85	N	Lalko et al. 2013	Acylation >> Direct acylation involving a leaving group >> Anhydrides (sulphur analogues of anhydrides)	Intermediate in chemical synthesis; Manufacturing	Anhydrides		Test
Kathon CG	55965-84-9	POS	POS	POS	132.3	1.38	-1.61	-1.2218 4875	91.85	229	POS	POS	4.5	99.1	52.1	N	Bauch et al. 2012	SN2 >> Ring opening SN2 reaction >> Isothiazolone derivatives AND SNVinyl >> Nucleophilic vinylic substitution on activated halogens >> Halogenated isothiazolones	Pesticides (antimicrobial)	Thiazoles [sulfur and nitrogen-containing heterocyclics]		Training
Glycerol	56-81-5	NEG	NEG	NEG	92.09	-1.76	1.036	-3.7747	18.2	290	NEG	NEG	0.8	1.2	0.95	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Cosmetics; Food additive; Intermediate in chemical synthesis; Manufacturing; Personal care products; Pharmaceuticals; Solvent	Alcohols		Training
Propylene glycol	57-55-6	NEG	NEG	NEG	76.0944	-0.92	1.119	-0.8894	-60	187.6	NEG	NEG	0.6	-0.9	-0.3	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Cosmetics; Food additive; Intermediate in chemical synthesis; Personal care products; Pharmaceuticals; Solvent; Pesticides (antimicrobial)	Alcohols		Test
3-Aminophenol	591-27-5	POS	POS	POS	109.13	0.21	-0.607	-2.02	123	164	NEG	NEG	1.2	8.2	3.9	Pro	Kern et al. 2010	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime	Cosmetics; Pharmaceuticals	Amino phenols		Test

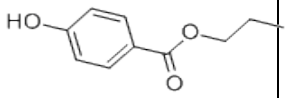
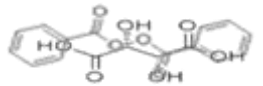
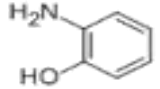
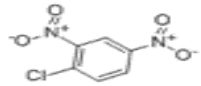
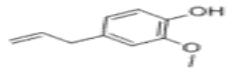
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
																		structure; Nitroquinones, Naphthoquinone(s)/imines for product				
2,4-Heptadienal	5910-85-0	POS	POS	POS	110.16	1.86	0.448	1.6274	-44.2	167.8	POS	POS	23.9	97.3	60.6	U	NA	Michael Addition >> alpha,beta-Unsaturated carbonyl compounds >> alpha,beta-Aldehydes	Food additive	Ketones with unsaturated carbon chains		Training
Penicillin G	61-33-6	POS	POS	NEG	334.391	1.83	-0.699	-11.593 45982	110	565.3	POS	NEG	0	14.3	7.15	N	Ashikaga 2010	Acylation >> Ring opening acylation >> beta-Lactams	Pharmaceuticals	Other		Training
2, 5-Toluenediamine sulfate	615-50-9	POS	POS	POS	156.252	0.74	0.702	-10.064	71.44	122	POS	POS	15	78.4	46.7	U	NA	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines AND Nucleophilic addition >> Addition to carbon-hetero double bonds >> Ketones for product	Intermediate in chemical synthesis; Manufacturing	Benzyl diamines		Test
Aniline	62-53-3	NEG	POS	POS	93.1265	0.9	-0.413	-0.3098	-6	184.1	NEG	POS	2.4	0.3	0.7	NA	Gerberick et al. 2009	No parent alert; Nucleophilic addition >> Nucleophilic addition reaction at polarized N-functional double bond >> C-Nitroso compounds for product	Food additive; Manufacturing; Personal care products; Pesticides; Pharmaceuticals	Other		Training
Sulfanilamide	63-74-1	NEG	NEG	NEG	172.206	-0.62	-1.361	-5.1367	165.5	342	NEG	NEG	-4.8	1.9	-1.9	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Pharmaceuticals	Other		Training

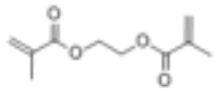
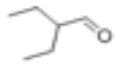
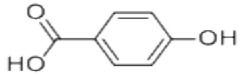
		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Ethyl (2-(4-chlorophenoxy)-2-methylpropanoate [Clofibrate]	637-07-0	NEG	POS	NEG	242.7	3.62	-1.527	-2.5952	15.74	150	NEG	NEG	1.9	2.3	2.1	NA	Natsch et al. 2013	No alert	Pharmaceuticals	Other		Training
Benzoic acid	65-85-0	NEG	NEG	NEG	212.2	1.87	-1.655	-3.1549	122.4	249.2	POS	NEG	39.9	19.1	31.2	NA	Ashikaga 2010; Natsch et al. 2013	No alert	Food additive; Manufacturing; Pharmaceuticals; Pesticides (antimicrobial)	Benzoic acids		Test
Isopropanol	67-63-0	NEG	NEG	NEG	60.095	0.05	1.221	1.6572	-89.5	82.3	NEG	NEG	0.5	1.2	0.82	NA	Gerberick et al. 2009	No alert	Cosmetics; Food additive; Intermediate in chemical synthesis; Manufacturing; Personal care products; Pharmaceuticals; Solvent; Pesticides (antimicrobial)	Alcohols		Training
trans-2-Hexenal	6728-26-3	POS	POS	POS	98.15	1.58	0.721	0.8195	-55.6	146.5	POS	POS	3.6	97.9	50.75	U	NA	Michael Addition >> alpha,beta-Unsaturated carbonyl compounds >> alpha,beta-Aldehydes	Fragrance agent	Ketones with unsaturated carbon chains		Training
Salicylic acid	69-72-7	POS	POS	NEG	138.121	2.26	-1.79	-4.0862	158	211	NEG	NEG	1	4.3	4.12	N	Troutman et al. 2011	No alert	Food additive; Manufacturing; Pharmaceuticals; Pesticides	Other		Training
1-Butanol	71-36-3	NEG	POS	NEG	74.12	0.88	-0.069	0.8261	-89.8	117.7	NEG	NEG	-0.1	0.2	0.1	NA	Gerberick et al. 2009; Natsch et al. 2013	No alert	Intermediate in chemical synthesis; Household products; Manufacturing; Pharmaceuticals	Alcohols		Training
Linalool	78-70-6	POS	POS	POS	154.25	2.97	-1.987	-0.7959	7.38	197	NEG	NEG	7.9	0	3.95	N	Ashikaga 2010	No parent alert; SN2 >> Ring opening SN2 reaction >> Epoxides, Aziridines and Sulfuranes for product	Cosmetics; Food additive; Manufacturing; Personal care products; Pharmaceuticals;	Unsaturated aliphatic alcohols		Test

		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
																			Pesticides			
Lilial	80-54-6	POS	POS	POS	204.3	4.36	-2.105	-2.4461	10.54	279.5	POS	NEG	0.7	14	7.35	N	Ashikaga 2010	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Fragrance agent	Benzyl aldehydes with longer unsaturated aliphatic chains		Test
Methyl methacrylate	80-62-6	POS	POS	POS	100.12	1.38	-0.824	1.5855	-48	100.5	POS	POS	5.3	42.1	23.7	N	Bauch et al. 2012	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Manufacturing	Acrylates		Training
Benzalkonium chloride	8001-54-5	POS	NEG	NEG	339.26	1.68	-5.94	-11.4522	241	384.54	NEG	NEG	0	2.8	1.4	N	Ashikaga 2010; Natsch et al. 2013	No alert	Cosmetics; Manufacturing; Household products; Pharmaceuticals; Pesticides (antimicrobial)	Other		Training
Saccharin	81-07-2	NEG	NEG	POS	183.18	0.91	-1.661	-6.1904	228	320.11	POS	NEG	0	2.4	12	NA	Ashikaga 2010; Natsch et al. 2013	Acylation >> Direct acylation involving a leaving group >> N-acylsulfonamides	Food additive	Other		Training
2-Hydroxyethyl acrylate	818-61-1	POS	POS	NEG	116.115	-0.21	0.935	-1.2815	-15.9	191	POS	POS	88.9	92.6	90.75	N	Troutman et al. 2011	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Intermediate in chemical synthesis; Manufacturing	Acrylates		Training
Diethyl phthalate	84-66-2	NEG	POS	NEG	222.24	2.42	-2.313	-2.6778	-40.5	295	NEG	NEG	0	0.8	0.4	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Household products; Manufacturing	Benzyl esters		Training

		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Phthalic anhydride	85-44-9	POS	NEG	POS	148.117	1.6	-1.378	-3.2865	130.8	295	POS	NEG	75	1.9	37.5	N	Ashikaga 2010	Acylation >> Direct acylation involving a leaving group >> Anhydrides (sulphur analogues of anhydrides)	Intermediate in chemical synthesis; Manufacturing	Anhydrides		Training
Tartaric acid	87-69-4	NEG	NEG	NEG	150.087	-1	2.765	-6.8297	169	179.1	NEG	NEG	-0.4	5.2	2.4	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Food additive	Dicarboxylic acids		Test
Pentachlorophenol	87-86-5	POS	POS	POS	226.34	5.12	-4.279	-3.9586	174	309.5	POS	NEG	14.5	0	7.25	Pre/Pro	Troutman et al. 2011	No alert	Pesticides (antimicrobial)	Other		Training
2-Acetylcyclohexanone	874-23-7	NEG	POS	NEG	140.18	1.34	0.778	-0.9318	26	229	POS	POS	12.5	18.2	15.35	NA	Natsch et al. 2013	No alert	Intermediate in chemical synthesis	Other		Training
Diphenylcyclopropenone	886-38-4	POS	POS	POS	206.25	3.25	-0.398	-2.6253	119	296	POS	POS	0	98.8	50.7	N	Ashikaga 2010	Acylation >>Ring opening acylation >> cyclopropenones	Pharmaceuticals; Personal care products	Other		Training
1-Naphthol	90-15-3	POS	POS	POS	144.17	2.85	-2.221	-3.5622	95	288	POS	POS	12.4	68.2	40.3	Pre/Pro	Troutman et al. 2011	No alert	Manufacturing; Pesticides	Other		Training
Coumarin	91-64-5	POS	NEG	NEG	146.14	1.39	-1.886	-3.0088	71	301.7	NEG	POS	0	1	0.5	N	Troutman et al. 2011; Natsch et al. 2013	No alert	Fragrance agent; Manufacturing; Pharmaceuticals	Benzopyran [2-ring heterocyclics with oxygen]		Test

		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
6-Methylcoumarin	92-48-8	NEG	NEG	NEG	160.169	2.06	0.075	-3.2907	76.5	304	NEG	POS	1	-0.2	0.4	NA	Ashikaga 2010; Natsch et al. 2013	No alert	Food additive; Fragrance agent	Other		Training
2-Hydroxypropyl methacrylate	923-26-2	NEG	NEG	POS	144.168	0.97	1.587	-1.1427	70	205	POS	POS	0	58.4	25.3	NA	Ashikaga 2010; Natsch et al. 2013	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Manufacturing; Plastics; Rubber	Other		Training
2-Methoxy-4-methylphenol	93-51-6	POS	POS	POS	138.164	1.88	0.32	-1.6198	5.5	221	POS	NEG	11.5	0	7.1	Pro	Natsch et al. 2013; Gerberick et al. 2009	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines for product	Pharmaceuticals	Other		Training
Phenylpropionaldehyde [α-Methylphenylacetaldehyde]	93-53-8	POS	POS	POS	134.07	1.96	0.414	-0.5686	-10	203.5	POS	POS	10.4	35.8	23.5	NA	Troutman et al. 2011	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Food additive; Fragrance agent	Benzaldehydes with short unsaturated aliphatic chains		Training
Phenyl benzoate	93-99-2	POS	POS	POS	198.217	3.59	-1.416	-2.7696	71	314	POS	NEG	2.2	64.9	33.55	N	Ashikaga 2010	Acylation >> Ester aminolysis or thiolysis >> Activated aryl esters	Manufacturing; Pesticides	Benzyl esters		Training
Ethyl benzoylacetate	94-02-0	NEG	NEG	NEG	192.21	1.87	0.083	-3.3893	0	284	NEG	POS	0.9	11.2	6.05	NA	Ashikaga 2010; Natsch et al. 2013	Nucleophilic addition >> Addition to carbon-hetero double bonds >> Ketones	Food additive	Benzyl esters		Test

		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20						
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Benzocaine	94-09-7	NEG	POS	NEG	165.189	1.86	-2.101	-3.585	92	310	POS	POS	6.1	14.6	9.4	NA	Ashikaga 2010; Natsch et al. 2013	No alert	Pharmaceuticals	Amino benzoic acid or ester		Training
Propylparaben	94-13-3	NEG	POS	NEG	180.2	3.04	-2.557	-3.2557	97	301	NEG	POS	-0.7	3.3	1.3	NA	Troutman et al. 2011; Natsch et al. 2013	No alert	Cosmetics; Food additive; Household products; Pharmaceuticals	Phenolic esters		Test
Benzoyl peroxide	94-36-0	POS	NEG	NEG	242.23	3.46	-4.425	-4.1512	105	328.6	POS	NEG	81.3	100	90.65	Pro	Ashikaga 2010	Acylation >> Direct acylation involving a leaving group >> Diacyl peroxides, anhydrides (sulphur analogues of diacyl peroxides)	Manufacturing; Household products	Benzyl esters		Test
2-Aminophenol	95-55-6	POS	POS	POS	109.13	0.62	-0.737	-2.02	174	153	POS	POS	18.1	96.2	57.15	Pro	Natsch and Haupt 2013; Gerberick et al. 2009	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines for product	Intermediate in chemical synthesis; Manufacturing	Amino phenols		Training
2,4-Dinitrochlorobenzene	97-00-7	POS	POS	POS	202.55	2.17	-2.097	-4.0711	53	315	POS	POS	29.3	100	67.4	N	Ashikaga 2010	SNAr >> Nucleophilic aromatic substitution on activated aryl and heteroaryl compounds >> Activated aryl and heteroaryl compounds	Manufacturing; Pesticides	Nitrobenzenes		Training
Eugenol	97-53-0	POS	POS	POS	164.201	2.27	-1.562	-1.6455	-9.1	253.2	POS	NEG	9	18.8	17.4	Pro	Natsch and Haupt 2013; Gerberick et al. 2009	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines for product	Cosmetics; Food additive; Intermediate in chemical synthesis; Manufacturing; Personal care products; Pharmaceuticals; Pesticides	Ether phenols with carbon side chains		Training

		Model Inputs for Best SVM Model									Other Model Inputs Tested					Additional Chemical Information20							
Chemical Name	**CASRN**	**LLNA Result (Majority)**	**h-CLAT Result (Majority)**	**OECD QSAR Toolbox Read-across Result**	**MW (g/mol)**	**Log P**	**Log S mol/L**	**Log VP mm Hg**	**MP °C**	**BP °C**	**DPRA Result (Majority)**	**KeratinoSens Result (Majority)**	**Avg % Depletion Lys (0.5mM:2.5mM [1:50])**	**Avg % Depletion Cys (0.5mM:5mM [1:10])**	**Avg % Depletion Lys & Cys**	**Pre/Pro-hapten?**	**Pre/Pro-hapten Reference**	**Protein Binding Mechanism (Protein binding alerts for skin sensitization by OASISv1.2 in QSAR Toolboxv3.2)**	**Product Class**	**Structural Group (for Training and Test Set Construction)**	**Structure**	**Test or Training Set**
Isoeugenol	97-54-1	POS	NEG	POS	164.201	3.04	-0.449	-1.9223	-10	266	POS	POS	15.3	94.1	55	Pre	Emter et al. 2010; Gerberick et al. 2009	No parent alert; Michael Addition >> Quinoide type compounds >> Quinone methide(s)/imines; Quinoide oxime structure; Nitroquinones, Naphthoquinone(s)/imines for product	Food additive; Fragrance agent	Ether phenols with carbon side chains		Test
Ethylene glycol dimethacrylate	97-90-5	POS	POS	POS	198.216	2.21	-0.236	-0.7258	-75	80.7	POS	POS	20.5	90.3	56.4	N	Ashikaga 2010	Michael Addition >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-Carbonyl compounds with polarized double bonds	Manufacturing	Acrylates		Training
Diethylacetaldehyde	97-96-1	POS	POS	POS	100.16	1.73	0.594	1.2279	116	118	POS	POS	60.2	54.5	57.35	N	Troutman et al. 2011	Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes	Food additive	Other		Training
4-Hydroxybenzoic acid	99-96-7	NEG	NEG	NEG	138.121	1.58	-1.441	-6.7167	214.5	298	NEG	NEG	1.2	1.5	1.4	NA	Ashikaga 2010; Natsch et al. 2013	No alert	Intermediate in chemical synthesis; Manufacturing; Pesticides	Benzoic acids		Training

Appendix V.1.1: SVM Model Results

Chemical Name	CASRN	LLNA Result (Majority)	h-CLAT + Toolbox + 6 properties	DPRA + KeratinoSens + h-CLAT + Toolbox + Lys + Cys + Avg.Lys.Cys + 6 properties	KeratinoSens + h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties	KeratinoSens + Toolbox + Avg.Lys.Cys + 6 properties	KeratinoSens + h-CLAT + Avg.Lys.Cys + 6 properties	h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties	KeratinoSens + h-CLAT + Toolbox + 6 properties
1-Bromobutane	109-65-9	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
1-Bromohexane	111-25-1	POS	POS	POS	POS	POS	POS	POS	POS
1-Butanol	71-36-3	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
1-Naphthol	90-15-3	POS	POS	POS	POS	POS	POS	POS	POS
1,2-Benzisothiazolin-3-one	2634-33-5	POS	POS	POS	POS	POS	POS	POS	POS
1,2-Dibromo-2,4-dicyanobutane	35691-65-7	POS	POS	POS	POS	POS	POS	POS	POS
1,4-Dihydroquinone	123-31-9	POS	POS	POS	POS	POS	POS	POS	POS
2-Acetylcyclohexanone	874-23-7	NEG	NEG	NEG	NEG	NEG	POS	NEG	NEG
2-Aminophenol	95-55-6	POS	POS	POS	POS	POS	POS	POS	POS
2-Ethylhexylacrylate	103-11-7	POS	POS	POS	POS	POS	POS	POS	POS
2-Hydroxyethyl acrylate	818-61-1	POS	POS	POS	POS	POS	POS	POS	NEG
2-Hydroxypropyl methacrylate	923-26-2	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
2-Mercaptobenzothiazole	149-30-4	POS	POS	POS	POS	POS	POS	POS	POS
2-Methoxy-4-methylphenol	93-51-6	POS	POS	POS	POS	POS	POS	POS	POS
2-Nitro-1,4-phenylenediamine	5307-14-2	POS	POS	POS	POS	POS	POS	POS	POS

2, 5-Toluenediamine sulfate	615-50-9	POS	POS	POS	POS	POS	POS	POS	POS
2,3-Butanedione	431-03-8	POS	POS	POS	POS	POS	POS	POS	POS
2,4-Dinitrochlorobenzene	97-00-7	POS	POS	POS	POS	POS	POS	POS	POS
2,4-Heptadienal	5910-85-0	POS	POS	POS	POS	POS	POS	POS	POS
2,4,6-Trinitrobenzenesulfonic acid	2508-19-2	POS	POS	POS	POS	POS	POS	POS	POS
3 and 4-(4-Hydroxy-4-methylpentyl)-3-cyclohexane-1-carbox-aldehyde [Lyral HMPCC]	31906-04-4	POS	POS	POS	POS	POS	POS	POS	POS
3-Aminophenol	591-27-5	POS	POS	POS	POS	NEG	POS	NEG	POS
3-Dimethylaminopropylamine	109-55-7	POS	POS	POS	POS	POS	POS	POS	POS
3-Phenoxypropiononitrile	3055-86-5	NEG	POS	POS	POS	NEG	POS	POS	POS
3-Propylidenephthalide	17369-59-4	POS	POS	POS	POS	POS	POS	POS	POS
3,3,4,5-Tetrachlorosalicylanilide	1154-59-2	POS	POS	POS	POS	POS	POS	POS	POS
3,4-Dihydrocoumarin	119-84-6	POS	POS	POS	POS	NEG	POS	POS	POS
4-(N-Ethyl-N-2-methan-sulfamido-ethyl)-2-methyl-1,4-phenylenediamine	25646-71-3	POS	POS	POS	POS	POS	POS	POS	POS
4-Allylanisole	140-67-0	POS	POS	POS	POS	POS	POS	POS	POS
4-Hydroxybenzoic acid	99-96-7	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
4-Methoxyacetophenone	100-06-1	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
4-Nitrobenzyl bromide	100-11-8	POS	POS	POS	POS	POS	POS	POS	POS
4-Phenylenediamine	106-50-3	POS	POS	POS	POS	POS	POS	POS	POS
5-Methyl-2,3-hexanedione (Acetyl isovaleryl)	13706-86-0	POS	POS	POS	POS	POS	POS	POS	POS
6-Methylcoumarin	92-48-8	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Abietic acid	514-10-3	POS	POS	POS	POS	POS	POS	POS	POS
Aniline	62-53-3	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG

Bandrowski's base	20048-27-5	POS	POS	POS	POS	POS	POS	POS	POS
Benzaldehyde	100-52-7	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Benzalkonium chloride	8001-54-5	POS	POS	POS	POS	POS	POS	POS	POS
Benzocaine	94-09-7	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Benzoic acid	65-85-0	NEG	NEG	NEG	NEG	NEG	POS	NEG	NEG
Benzoquinone	106-51-4	POS	POS	POS	POS	POS	POS	POS	POS
Benzoyl peroxide	94-36-0	POS	POS	POS	POS	POS	POS	POS	POS
Benzyl benzoate	120-51-4	POS	POS	POS	POS	POS	POS	POS	POS
Benzyl bromide	100-39-0	POS	POS	POS	POS	POS	POS	POS	POS
Benzylidene acetone	122-57-6	POS	NEG	POS	POS	POS	POS	POS	NEG
Bisphenol A diglycidyl ether	1675-54-3	POS	POS	POS	POS	POS	POS	POS	POS
Butyl acrylate	141-32-2	POS	POS	POS	POS	POS	POS	POS	POS
Butyl glycidyl ether	06-08-2426	POS	POS	POS	POS	POS	POS	POS	POS
Chlorobenzene	108-90-7	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Cinnamic aldehyde	104-55-2	POS	POS	POS	POS	POS	POS	POS	POS
Cinnamyl alcohol	104-54-1	POS	POS	POS	POS	POS	POS	POS	POS
Citral	5392-40-5	POS	POS	POS	POS	POS	POS	POS	POS
Coumarin	91-64-5	POS	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Cyclamen aldehyde	103-95-7	POS	POS	POS	POS	POS	POS	POS	POS
Diethyl maleate	141-05-9	POS	POS	POS	POS	POS	POS	POS	POS
Diethyl phthalate	84-66-2	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Diethylacetaldehyde	97-96-1	POS	POS	POS	POS	POS	POS	POS	POS
Diphenylcyclopropenone	886-38-4	POS	POS	POS	POS	POS	POS	POS	POS
Ethyl (2-(4-chlorophenoxy)-2-methylpropanoate [Clofibrate]	637-07-0	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG

Ethyl benzoylacetate	94-02-0	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Ethyl vanillin	121-32-4	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Ethylene glycol dimethacrylate	97-90-5	POS	POS	POS	POS	POS	POS	POS	POS
Ethylenediamine	107-15-3	POS	POS	POS	POS	POS	POS	POS	POS
Eugenol	97-53-0	POS	POS	POS	POS	POS	NEG	POS	POS
Farnesal	502-67-0	POS	POS	POS	POS	POS	POS	POS	POS
Formaldehyde	50-00-0	POS	POS	POS	POS	POS	POS	POS	POS
Fumaric acid	110-17-8	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Geraniol	106-24-1	POS	POS	POS	POS	POS	POS	POS	POS
Glutaraldehyde	111-30-8	POS	POS	POS	POS	POS	POS	POS	POS
Glycerol	56-81-5	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Glyoxal	107-22-2	POS	POS	POS	POS	POS	POS	POS	POS
Hexane	110-54-3	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Hexyl cinnamic aldehyde	101-86-0	POS	POS	POS	POS	POS	POS	POS	POS
Hydroxycitronellal	107-75-5	POS	POS	POS	POS	POS	POS	POS	POS
Imidazolidinyl urea	39236-46-9	POS	POS	POS	POS	POS	POS	POS	POS
Isoeugenol	97-54-1	POS	POS	POS	POS	POS	POS	POS	POS
Isopropanol	67-63-0	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Kathon CG	55965-84-9	POS	POS	POS	POS	POS	POS	POS	POS
Lactic acid	50-21-5	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Lauryl gallate	1166-52-5	POS	POS	POS	POS	POS	POS	POS	POS
Lilial	80-54-6	POS	POS	POS	POS	POS	POS	POS	POS
Linalool	78-70-6	POS	POS	POS	POS	POS	NEG	POS	POS
Maleic anhydride	108-31-6	POS	POS	POS	POS	POS	POS	POS	POS
Methyl 2-nonynoate	111-80-8	POS	POS	POS	POS	POS	POS	POS	POS

Methyl methacrylate	80-62-6	POS	POS	POS	POS	POS	POS	POS	POS
Methyl salicylate	119-36-8	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Methylisothiazolinone	2682-20-4	POS	POS	POS	POS	POS	POS	POS	POS
Nonanoic acid	112-05-0	POS	NEG	NEG	NEG	POS	NEG	NEG	NEG
Octanoic acid	124-07-2	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Oxalic acid	144-62-7	POS	POS	POS	POS	POS	POS	POS	POS
Oxazolone	15646-46-5	POS	POS	POS	POS	POS	POS	POS	POS
p-Aminobenzoic acid	150-13-0	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Penicillin G	61-33-6	POS	POS	POS	POS	POS	POS	POS	POS
Pentachlorophenol	87-86-5	POS	POS	POS	POS	POS	POS	POS	POS
Perillaldehyde	2111-75-3	POS	POS	POS	POS	POS	POS	POS	POS
Phenyl benzoate	93-99-2	POS	POS	POS	POS	POS	POS	POS	POS
Phenylacetaldehyde	122-78-1	POS	POS	POS	POS	POS	POS	POS	POS
Phenylpropionaldehyde [α-Methylphenylacetaldehyde]	93-53-8	POS	POS	POS	POS	POS	POS	POS	POS
Phthalic anhydride	85-44-9	POS	POS	POS	POS	POS	POS	POS	POS
Propyl gallate	121-79-9	POS	POS	POS	POS	POS	POS	POS	POS
Propylene glycol	57-55-6	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Propylparaben	94-13-3	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Pyridine	110-86-1	POS	POS	POS	POS	POS	POS	NEG	POS
Resorcinol	108-46-3	POS	POS	POS	POS	NEG	POS	NEG	POS
Saccharin	81-07-2	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Salicylic acid	69-72-7	POS	POS	POS	POS	POS	POS	POS	POS
Sodium lauryl sulfate	151-21-3	POS	POS	POS	POS	POS	POS	POS	POS
Streptomycin sulfate	3810-74-0	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG

Sulfanilamide	63-74-1	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Sulfanilic acid	121-57-3	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Tartaric acid	87-69-4	NEG	NEG	NEG	NEG	NEG	NEG	NEG	POS
Tetramethylthiuram disulfide	137-26-8	POS	POS	POS	POS	POS	POS	POS	POS
trans-2-Hexenal	6728-26-3	POS	POS	POS	POS	POS	POS	POS	POS
Trimellitic anhydride	552-30-7	POS	POS	POS	POS	POS	POS	POS	POS
Undecylenic acid	112-38-9	POS	POS	NEG	NEG	POS	POS	POS	NEG
Vanillin	121-33-5	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Xylene	1330-20-7	POS	POS	POS	POS	POS	POS	NEG	POS
α-Amyl cinnamic aldehyde	122-40-7	POS	POS	POS	POS	POS	POS	POS	POS

Gray shading indicates misclassified compounds (discordant with the LLNA outcome).

Appendix V.1.2: SVM Model Performance

Model	Dataset	Sensitivity	Specificity	Accuracy
h-CLAT + Toolbox + 6 properties	Training	97	97	97
	Test	**95**	**100**	**96**
DPRA + KeratinoSens + h-CLAT + Toolbox + Lys + Cys + Avg.Lys.Cys + 6 properties	Training	99	96	98
	Test	**90**	**100**	**92**
KeratinoSens + h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties	Training	99	96	98
	Test	**90**	**100**	**92**
KeratinoSens + Toolbox + Avg.Lys.Cys + 6 properties	Training	99	100	99
	Test	**84**	**100**	**89**
KeratinoSens + h-CLAT + Avg.Lys.Cys + 6 properties	Training	97	92	96
	Test	**90**	**86**	**89**
h-CLAT + Toolbox + Avg.Lys.Cys + 6 properties	Training	96	96	96
	Test	**84**	**100**	**89**
KeratinoSens + h-CLAT + Toolbox + 6 properties	Training	96	96	96
	Test	**90**	**86**	**89**

Appendix V.2: Protocol for Generating Read-across Prediction for Skin Sensitization Hazard Using QSAR Toolboxv3.2

1. Input the target compound into QSAR Toolboxv3.2 by CASRN.

2. Profile the target compound for protein binding by OASISv1.2, protein binding by OECD, protein binding potency, and protein binding alerts for skin sensitization by OASIS v1.2.

3. If there are no protein binding alerts go back to input and generate auto-oxidation products and skin metabolites and then profile the products/metabolites for protein binding alerts. To generate auto-oxidation products and metabolites (which cannot be done simultaneously):

    a. Right click on SMILES structure and then right click to select "Multiplication," then "Metabolism/Transformation," and then "Auto-oxidation Products." Generating auto-oxidation products may take a little while, but afterwards, they appear in a tree form after clicking "[set]Auto-oxidation simulator."

    b. Select single component mode on the right of top banner for profiling the products (repeat #2).

    c. Generate skin metabolites by repeating #3a and #3b for metabolites: right click on SMILES structure and then right click to select "Multiplication," then "Metabolism/Transformation," and then "Skin metabolism simulator". The metabolites appear in a tree form after clicking "[set]Skin Metabolism simulator."

    d. If any auto-oxidation products or metabolites have protein binding alerts, select the one with more alerts for the skin sensitization hazard analysis. If an auto-oxidation product and a metabolite have an equal number of protein binding alerts, select the auto-oxidation product by right clicking on it in the data matrix and then select "Focus," which allows the product/metabolite to represent the target chemical. The selected product/metabolite appears in a new data matrix. Go to #4.

    e. **If neither the parent nor the auto-oxidation products nor the metabolites have protein binding alerts, the substance is negative**. No report can be generated for these.

4. For substances with protein binding alerts, retrieve data on *in vivo* skin sensitization endpoints for the target compound in the endpoint module. Use "Skin sensitization" and "Skin sensitization ECETOC" databases.

5. Go to the category definition module and use a structural profiler in this order of preference: US EPA new chemicals categories, OECD categories, organic functional groups, or structural similarity (use the default option, which is the Dice method) to look for analogs. If the substance is not classified by the first preference, go to the next. If the chemical is categorized in multiple categories, click "OR" at "combine profiles logically" in the dialog box to provide the largest possible group of analogs. Make sure "Skin sensitization ECETOC" and "Skin sensitization" are selected so that analogs are sought among substances with skin sensitization data. If there are multiple outcomes for a chemical, choose select one at the dialog box. It will use 1 representative of each different outcome (if there are 3 positives and 4 negatives for a substance it will use only 1 positive and one negative).

6. Fill data gap by read across. Select the in vivo skin sensitization cell under toxicity for the target substance. Select apply. Select the scale option for Skin sensitization II ECETOC (the lowest common denominator – positive and negative categories).

7. Use the default descriptor option of log Kow, which identifies the analogs closest to the target chemical.

8. Subcategorize: if all analogs are sensitizers or if all are nonsensitizers, there is no need to continue with subcategories. Go to #9. Otherwise, verify that analogs have the same mechanism/mode of action by opening the Select/Filter data menu on the right of the graph and re-profiling the identified analogs. Click on subcategorize and choose "Protein binding alerts for skin sensitization by OASISv1.2". [Note: This profiler is in the endpoint-specific section of profilers and is different from the general mechanistic profiler called "Protein binding by OASISv1.2."] This compares the mechanistic properties of the analogs with the target chemical. Eliminate the dissimilar chemicals by clicking on the "Remove" button. Those will be highlighted in green on the graph and in blue on the right of the subcategorization menu.

9. Accept the read across prediction.

    a. If an "unreliable" message is received (usually because the log Kow for the target chemical is outside the range for the analogs) or if a message indicates that there are too few data points to make a prediction, go back to #4 to add the ECHA CHEM database and then repeat the subsequent steps. **If adding the ECHA CHEM database fails to achieve a reliable prediction, make a "qualified" prediction based only on the structural category (i.e., without using the subcategorization step).** If a qualified prediction cannot be made, a prediction cannot be made with this protocol. (NOTE: Some, but not all, predictions made with the ECHA CHEM database yield a message that so much proprietary data was used for the prediction that a report cannot be made. Thus, if subcategorization yields an unreliable prediction with or without the ECHA CHEM database, the prediction should be made without the ECHA CHEM database (and without subcategorization) so as to get a report.

    b. If the prediction was made for a product/metabolite of the parent chemical of interest, assign prediction to parent by going back to data matrix of the target chemical (go to the input module). Click on "[set]Skin Metabolism simulator" or "[set]Auto-oxidation simulator," as appropriate. Then select the in vivo skin sensitization cell of the target chemical and then "Data Gap Filling." Click "Independent MOA" and then "Apply." Accept prediction and return to matrix. Final prediction for the parent is CI for component based independent mode. Select prediction in data matrix; right click on the prediction and then select report to generate a report.

    c. If the prediction was not made for a product/metabolite, accept the prediction, return to matrix, and go to the report module. Double click the prediction of interest on the left panel and then click the create button to generate a report.

Appendix V.3: Skin Data Training Updated, Skin Data Test Updated, and Skin Data New

Skin Data New																
Chemical Name	CASRN	MW	LogP	LogS	LogVP	MP	BP	Lys	Cys	avg-Lys-Cys	hCLAT	DPRA	Keratino	OECD	LLNA	Training/Test
NewChem	xxx-xx-x	175.4	1.33	-1.2	-2.9	80	280	9.9	0.9	3.9	1	1	1	1	POS	

Skin Data Test Updated														
ChemicalName	CASRN	MW	LogP	LogS	LogVP	MP	BP	Lys	Cys	avg-Lys-Cys	hCLAT	DPRA	Keratino	OECD
Vanillin	121-33-5	152.15	1.21	-1.14	-3.93	81.5	285	9.9	0.9	3.9	0	1	1	1
p-Aminobenzoic acid	150-13-0	137.14	0.83	-1.35	-3.56	188.5	307.7	0	0	0	0	0	0	0
Propylene glycol	57-55-6	76.09	-0.92	1.12	-0.89	-60	187.6	0	0	0	0	0	0	0
Benzoic acid	65-85-0	212.2	1.87	-1.66	-3.15	122.4	249.2	39.9	19.1	31.2	0	1	0	0
Tartaric acid	87-69-4	150.09	-1	0.59	-6.83	169	179.1	0	0	0	0	0	0	0
Ethyl benzoylacetate	94-02-0	192.21	1.87	-2.2	-3.39	0	284	0	0	0	0	0	1	0
Propylparaben	94-13-3	180.2	3.04	-2.56	-3.26	97	301	0	0	0	1	0	1	0

Skin Data Training Updated																
Chemical Name	CASRN	MW	LogP	LogS	LogVP	MP	BP	Lys	Cys	avg-Lys-Cys	hCLAT	DPRA	Keratino	OECD	LLNA	Training/Test
4-Methoxyacetophenone	100-06-1	150.17	1.74	-1.78	-2.19	38.5	258	0	0	0	0	0	1	1	NEG	Training
Benzaldehyde	100-52-7	106.12	1.48	-1.21	0.1	-26	179	0	0	0	1	0	1	0	NEG	Training
Chlorobenzene	108-90-7	112.56	2.84	-2.35	1.08	-45.2	131.7	0	0	0	1	0	0	0	NEG	Training
1-Bromobutane	109-65-9	137.02	2.75	-2.2	1.62	-112.4	101.3	1.2	13.8	7.5	1	1	0	0	NEG	Training
Fumaric acid	110-17-8	146.14	0.46	-1.32	-3.81	287	522	4.6	10.8	7.7	0	1	0	0	NEG	Training
Hexane	110-54-3	86.18	3.9	-3.96	2.18	-95.3	68.7	0	0	0	0	0	1	0	NEG	Training
Methyl salicylate	119-36-8	152.15	2.55	-1.31	-1.46	-8	222.9	0	0	0	0	0	0	0	NEG	Training
Ethyl vanillin	121-32-4	166.18	1.58	-1.77	-4.98	77.5	294	9.7	1.1	5.4	0	1	1	1	NEG	Training
Sulfanilic acid	121-57-3	173.19	-2.16	-1.21	-6.69	288	300	0	0	0	0	0	0	0	NEG	Training
Octanoic acid	124-07-2	144.21	3.05	-2.26	-2.43	16.3	239	0	0	0	1	0	0	0	NEG	Training
3-Phenoxypropiononitrile	3055-86-5	147.18	1.61	-1.92	-2.03	36.6	265.7	29.7	0	14.85	0	1	0	1	NEG	Training
Streptomycin sulfate	3810-74-0	581.57	-3.61	-1.48	-11.48	123.57	326.63	0	0	0	0	0	0	1	NEG	Training

Lactic acid	50-21-5	90.08	-0.72	1.05	-1.09	18	204.2	0	0	0	0	0	0	0	NEG	Training
Glycerol	56-81-5	92.09	-1.76	1.04	-3.77	18.2	290	0	0	0	0	0	0	0	NEG	Training
Aniline	62-53-3	93.13	0.9	-0.41	-0.31	-6	184.1	0	0	0	1	0	1	1	NEG	Training
Ethyl (2-(4-chlorophenoxy)-2-methylpropanoate [Clofibrate]	637-07-0	242.7	3.62	-3.91	-2.6	15.74	150	0	0	0	1	0	0	0	NEG	Training
Sulfanilamide	63-74-1	172.21	-0.62	-1.36	-5.14	165.5	342	0	0	0	0	0	0	0	NEG	Training
Isopropanol	67-63-0	60.1	0.05	1.22	1.66	-89.5	82.3	0	0	0	0	0	0	0	NEG	Training
1-Butanol	71-36-3	74.12	0.88	-0.07	0.83	-89.8	117.7	0	0	0	1	0	0	0	NEG	Training
Saccharin	81-07-2	183.18	0.91	-1.66	-6.19	228	320.11	0	2.4	12	0	1	0	1	NEG	Training
Diethyl phthalate	84-66-2	222.24	2.42	-2.31	-2.68	-40.5	295	0	0	0	1	0	0	0	NEG	Training
2-Acetylcyclohexanone	874-23-7	140.18	1.34	-1.37	-0.93	26	229	12.5	18.2	15.35	1	1	1	0	NEG	Training
2-Hydroxypropyl methacrylate	923-26-2	144.17	0.97	-0.57	-1.14	70	205	0	58.4	25.3	0	1	1	1	NEG	Training
6-Methylcoumarin	92-48-8	160.17	2.06	-2.13	-3.29	76.5	304	0	0	0	0	0	1	0	NEG	Training
Benzocaine	94-09-7	165.19	1.86	-2.1	-3.59	92	310	6.1	14.6	9.4	1	1	1	0	NEG	Training

4-Hydroxybenzoic acid	99-96-7	138.12	1.58	-1.44	-6.72	214.5	298	0	0	0	0	0	0	0	NEG	Training
Benzyl bromide	100-39-0	171.03	2.92	-2.65	-0.35	-3	201	22.1	100	62	1	1	1	1	POS	Training
Hexyl cinnamic aldehyde	101-86-0	216.32	4.82	-4.9	-3.27	39.2	318.7	0	0	0	0	0	1	1	POS	Training
Cyclamen aldehyde	103-95-7	190.29	3.4	-3.46	-2.65	16.77	234	1	18.9	9.95	0	1	1	1	POS	Training
Cinnamyl alcohol	104-54-1	134.18	1.95	-1.34	-1.62	33	250	6.4	10.7	9.6	1	1	1	1	POS	Training
Cinnamic aldehyde	104-55-2	132.16	1.9	-1.97	-1.54	-8	246	46.6	68.8	57.8	1	1	1	1	POS	Training
Geraniol	106-24-1	154.25	3.56	-3.19	-1.52	-15	230	0	0	0	1	0	1	1	POS	Training
4-Phenylenediamine	106-50-3	108.14	-0.3	-0.47	-2.3	146	267	26.5	93	58.3	1	1	1	1	POS	Training
Benzoquinone	106-51-4	108.1	0.2	-0.99	-1.05	116	180	91	99	95	1	1	1	1	POS	Training
Glyoxal	107-22-2	58.04	-1.66	1.24	2.41	15	50.4	67.8	56.5	62.15	1	1	1	1	POS	Training
Hydroxycitronellal	107-75-5	172.26	2.11	-1.75	-2.24	23	241	14.6	32.9	23.8	1	1	1	1	POS	Training
Maleic anhydride	108-31-6	98.06	1.62	-1.3	-0.6	52.8	202	41	100	59.4	1	1	1	1	POS	Training
3-Dimethylaminopropylamine	109-55-7	102.18	-0.45	0.99	1	-60	133	0	0	0	1	0	1	1	POS	Training
Pyridine	110-86-1	79.1	0.65	1.1	1.32	-41.6	115.2	0	0	0	1	0	0	0	POS	Training

1-Bromohexane	111-25-1	165.08	3.8	-3.81	0.59	-84.7	155.3	68.2	14.1	41.15	0	1	1	1	POS	Training
Glutaraldehyde	111-30-8	100.12	-0.18	0.22	-0.22	-14	188	85.4	30.2	57.8	1	1	1	1	POS	Training
Methyl 2-nonynoate	111-80-8	168.24	3.1	-3.07	-1.13	38	224.5	3.2	100	51.6	1	1	1	1	POS	Training
Nonanoic acid	112-05-0	158.24	3.42	-2.75	-2.78	12.3	254.5	0	0	0	1	0	0	0	POS	Training
3,3, 4,5-Tetrachlorosalicyl anilide	1154-59-2	351.02	5.87	-6.34	-9.52	161	488.1	9	36.8	22.9	1	1	1	1	POS	Training
3,4-dihydrocoumarin	119-84-6	148.16	0.97	-1.69	-2.08	25	272	39.7	0	19.85	1	1	0	1	POS	Training
Benzyl benzoate	120-51-4	212.25	3.97	-4.14	-3.65	21	323.5	0	0	0	0	0	1	1	POS	Training
Propyl gallate	121-79-9	212.2	1.8	-1.78	-6.39	130	363.6	41.1	59.5	51.5	1	1	1	0	POS	Training
Benzylidene acetone	122-57-6	146.19	2.07	-2.04	-1.91	42	261	0.1	93.2	46.7	1	1	1	1	POS	Training
1,4-Dihydroquinone	123-31-9	110.11	0.59	-0.18	-4.62	172.3	287	51.1	83.3	67.2	1	1	1	1	POS	Training
Xylene	1330-20-7	106.08	3.16	-3	0.9	-47	138.5	0	0	0	0	0	0	0	POS	Training
5-Methyl-2,3-hexanedione (Acetyl isovaleryl)	13706-86-0	128.17	0.06	1.92	0.94	-16.7	138	7.5	25.8	16.65	1	1	1	1	POS	Training
Tetramethylthiuram disulfide	137-26-8	240.44	1.73	-3.9	-4.76	155.6	129	6.9	99.5	53.2	1	1	1	1	POS	Training
4-Allylanisole	140-67-0	148.2	3.47	-0.19	-0.78	-1.19	215.5	0	45	21.8	1	1	1	1	POS	Training

Diethyl maleate	141-05-9	172.18	2.2	-1.09	-0.98	-9	223	82.1	99.9	91	1	1	1	1	POS	Training
Butyl acrylate	141-32-2	128.17	2.36	-1.81	0.74	-64.6	145	85.1	100	92.55	1	1	1	1	POS	Training
Oxalic acid	144-62-7	90.05	-2.22	0.39	-3.63	190	246.9	0	0	0	1	0	1	0	POS	Training
2-Mercaptobenzothiazole	149-30-4	167.25	2.42	-3.18	-3.33	181	338.8	3.2	99.9	51.55	1	1	1	1	POS	Training
Sodium lauryl sulfate	151-21-3	288.38	1.6	-0.46	-12.33	205.5	216	48.5	0	24.5	0	1	0	0	POS	Training
Oxazolone	15646-46-5	217.22	1.51	-2.07	-5.19	120	365.7	47	75.5	61.3	1	1	1	1	POS	Training
Bisphenol A-diglycidyl ether	1675-54-3	340.42	3.84	-5.69	-6.96	12	421	1.1	42.5	21.8	1	1	1	1	POS	Training
3-Propylidenephthalide	17369-59-4	174.2	2.03	-2.21	-3.52	66.8	318.2	30.6	14.3	11.9	1	1	0	1	POS	Training
Bandrowski? base	20048-27-5	318.38	0.74	-1.14	-10.36	224.8	526.1	4.2	87.5	45.85	1	1	1	1	POS	Training
Perillaldehyde	2111-75-3	150.22	3.34	-2.97	-1.33	218.8	240	13.8	31.9	22.85	1	1	1	1	POS	Training
Butyl glycidyl ether	06-08-2426	130.19	0.63	-0.81	0.51	-31	165	11.8	67.3	39.6	0	1	1	1	POS	Training
2,4,6-trinitrobenzenesulfonic acid	2508-19-2	293.17	-1.71	-0.27	-11.28	206.9	487.7	22.6	99.7	61.15	1	1	1	0	POS	Training
4-(N-Ethyl-N-2-methan-sulfamido-ethyl)-2-methyl-1,4,-phenylenediamine	25646-71-3	274.34	1.2	-3.25	-5.76	121.03	318.05	13.6	90.1	51.85	1	1	1	1	POS	Training

1,2-Benzisothiazolin-3-one	2634-33-5	151.19	0.64	-0.85	-4.59	121.6	339.5	9.7	97.7	53.7	1	1	1	1	POS	Training
3 and 4-(4-Hydroxy-4-methylpentyl)-3-cyclohexane-1-carboxaldehyde [Lyral HMPCC]	31906-04-4	210.32	3.32	-3.06	-4.07	-30	307.1	3.4	39.4	21.4	1	1	1	1	POS	Training
1,2-Dibromo-2,4-dicyanobutane	35691-65-7	265.93	1.63	-2.31	-4.3	52	327.8	28.5	99.5	52.1	1	1	1	0	POS	Training
Imidazolidinyl urea	39236-46-9	388.29	-8.28	0.41	-28.47	240.5	932.2	15.2	54.1	29.7	0	1	1	0	POS	Training
Formaldehyde	50-00-0	30.03	0.35	1.12	5.89	-148.5	-19.1	2.4	44.7	24.45	1	1	1	1	POS	Training
Farnesal	502-67-0	220.36	5.74	-5.71	-2.76	16.7	302.2	2.2	24.7	14.4	1	1	1	1	POS	Training
Abietic acid	514-10-3	302.46	6.46	-3.8	-6.5	173.5	394.9	16.3	99.9	58.1	0	1	1	0	POS	Training
2-Nitro-1,4-phenylenediamine	5307-14-2	153.14	0.53	-0.77	-4.25	140	134	0	0	50	1	1	1	1	POS	Training
Kathon CG	55965-84-9	132.3	1.38	-1.61	-1.22	91.85	229	4.5	99.1	52.1	1	1	1	1	POS	Training
2,4-Heptadienal	5910-85-0	110.16	1.86	-1.59	1.63	-44.2	167.8	23.9	97.3	60.6	1	1	1	1	POS	Training
Penicillin G	61-33-6	334.39	1.83	-3.22	-11.59	110	565.3	0	14.3	7.15	1	1	0	0	POS	Training
trans-2-Hexenal	6728-26-3	98.15	1.58	-1.27	0.82	-55.6	146.5	3.6	97.9	50.75	1	1	1	1	POS	Training
Salicylic acid	69-72-7	138.12	2.26	-1.79	-4.09	158	211	0	0	0	1	0	0	0	POS	Training

Benzalkonium chloride	8001-54-5	339.26	1.68	-5.94	-11.45	241	384.54	0	0	0	0	0	0	0	POS	Training
Methyl methacrylate	80-62-6	100.12	1.38	-0.82	1.59	-48	100.5	5.3	42.1	23.7	1	1	1	1	POS	Training
2-Hydroxyethyl acrylate	818-61-1	116.12	-0.21	0.94	-1.28	-15.9	191	88.9	92.6	90.75	1	1	1	0	POS	Training
Phthalic anhydride	85-44-9	148.12	1.6	-1.38	-3.29	130.8	295	75	1.9	37.5	0	1	0	1	POS	Training
Pentachlorophenol	87-86-5	226.34	5.12	-4.28	-3.96	174	309.5	14.5	0	7.25	1	1	0	1	POS	Training
Diphenylcyclopropenone	886-38-4	206.25	3.25	-3.49	-2.63	119	296	0	98.8	50.7	1	1	1	1	POS	Training
1-Naphthol	90-15-3	144.17	2.85	-2.22	-3.56	95	288	12.4	68.2	40.3	1	1	1	1	POS	Training
2-Methoxy-4-methylphenol	93-51-6	138.16	1.88	-1.82	-1.62	5.5	221	11.5	0	7.1	1	1	0	1	POS	Training
Phenylpropionaldehyde [alpha-Methylphenylacetaldehyde]	93-53-8	134.07	1.96	-1.71	-0.57	-10	203.5	10.4	35.8	23.5	1	1	1	1	POS	Training
Phenyl benzoate	93-99-2	198.22	3.59	-3.71	-2.77	71	314	2.2	64.9	33.55	1	1	0	1	POS	Training
2-Aminophenol	95-55-6	109.13	0.62	-0.74	-2.02	174	153	18.1	96.2	57.15	1	1	1	1	POS	Training
2,4-Dinitrochlorobenzene	97-00-7	202.55	2.17	-4.4	-4.07	53	315	29.3	100	67.4	1	1	1	1	POS	Training
Eugenol	97-53-0	164.2	2.27	-1.56	-1.65	-9.1	253.2	9	18.8	17.4	1	1	0	1	POS	Training

Ethylene glycol dimethacrylate	97-90-5	198.22	2.21	-2.53	-0.73	-75	80.7	20.5	90.3	56.4	1	1	1	1	POS	Training
Diethylacetaldehyde	97-96-1	100.16	1.73	-1.41	1.23	116	118	60.2	54.5	57.35	1	1	1	1	POS	Training

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

Complete document available on OLIS in its original format
This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

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

© **OECD** Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

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.