Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2022)32

Unclassified

English - Or. English

1 September 2022

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Case Study on the Use of Integrated Approaches for Testing and Assessment for skin sensitisation: Demonstrating the Next Generation Risk Assessment Framework using Geraniol

Series on Testing and Assessment
No. 368

JT03501681

OECD Environment, Health and Safety Publications

SERIES ON TESTING AND ASSESSMENT

NO. 368

Case Study on the Use of Integrated Approaches for Testing and Assessment for skin sensitisation: Demonstrating the Next Generation Risk Assessment Framework using Geraniol

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 COOPERATION AND DEVELOPMENT

Paris 2022

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 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 twelve 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; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** 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 Organizations.

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.

- **Also published in the Series on Testing and Assessment: link**

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

E-mail: ehscont@oecd.org

Foreword

OECD member countries have been making efforts to expand the use of alternative methods in assessing chemicals. The OECD has been developing guidance documents and tools for the use of alternative methods such as (Q)SAR, chemical categories and Adverse Outcome Pathways (AOPs) as a part of Integrated Approaches for Testing and Assessment (IATA). There is a need for the investigation of the practical applicability of these methods/tools for different aspects of regulatory decision-making, and to build upon case studies and assessment experience across jurisdictions.

The objective of the IATA Case Studies Project is to increase experience with the use of IATA by developing case studies, which constitute examples of predictions that are fit for regulatory use. The aim is to create common understanding of using novel methodologies and the generation of considerations/guidance stemming from these case studies.

This case study was developed by Cosmetic Europe and submitted through BIAC for illustrating practical use of IATA and submitted to the 2021 review cycle of the IATA Case Studies Project.

The case study was reviewed by the project team, and endorsed at the 6th meeting of the Working Party on Hazard Assessment in June 2022.

The case study is illustrative examples, and their publication as OECD monographs does not translate into direct acceptance of the methodologies for regulatory purposes across OECD countries. In addition, the cases study should not be interpreted as official regulatory decisions made by the authoring member countries.

This document is published under the responsibility of the Chemicals and Biotechnology Committee of the OECD.

Acknowledgements

This case study was performed as part of the Cosmetics Europe Long Range Science Strategy (www.lrsscosmeticseurope.eu). The scientists that were involved in many different ways and at different times during the course of the case study in alphabetical order were:

Nathalie Alépée[1], Fanny Boisleve[2], Dagmar Bury[1], Nicola Gilmour[3], Sebastian Hoffmann[4], Petra Kern[5], Jochen Kühnl[6], Masaaki Miyazawa[7], Hayato Nishida[8], Erwin Van Vliet[9], Mustafa Varçin[10]

Affiliation at time of contribution:

[1] L'Oréal, Research & Innovation, Aulnay-sous-Bois, France; L'Oréal, Research & Innovation, Clichy, France
[2] Chanel, 135 avenue Charles de Gaulle, 92521 Neuilly, France
[3] Unilever, Colworth Science Park, Bedford, MK44 1LQ United Kingdom
[4] seh consulting + services, Stembergring 15, 33106 Paderborn, Germany
[5] Procter & Gamble Services NV/SA, Temselaan/ Boechoutlaan, 1853 Strombeek-Bever, Belgium
[6] Beiersdorf AG, Unnastraße 48, 20245 Hamburg, Germany
[7] Kao Corporation, 2606 Akabane, Ichikai, Haga, Tochigi, 321-3497, Japan
[8] Shiseido Global Innovation Center, 1-2-11, Takashima, Nishi-ku, Yokohama-shi, Kanagawa 220-0011, Japan
[9] Innovitox Consulting & Services, Regentenland 35, 3994TZ, Houten, The Netherlands
[10] Cosmetics Europe, Avenue Herrmann Debroux 40, 1160 Brussels, Belgium

Table of contents

FIGURES

TABLES

Abbreviations and acronyms

ACD	Allergic contact dermatitis
ANN	Artificial Neural Network
AOP	Adverse outcome pathway
BF	Bayes Factors
BN	Bayesian network
CE	Cosmetics Europe
DA	Defined approach
DC	Dendritic cells
DIP	Data integration procedure
DPRA	Direct Peptide Reactivity Assay
ECHA	European Chemicals Agency
GHS	Globally Harmonised System
HRIPT	Human Repeat Insult Patch Test
h-CLAT	Human Cell Line Activation Test
IATA	Integrated approaches to testing and assessment
ITS	Integrated Testing Strategy
KE	Key event
LLNA	Local Lymph Node Assay
MOE	Margin of exposure
MDBGN	Methyldibromo glutaronitrile
MIE	Molecular initiating event
MW	Molecular weight
NA	Not applicable
NAM	New approach methodology
NGRA	Next generation risk assessment
NR	Non-reactive
NS	Non-sensitiser
OECD	Organisation for Economic Cooperation and Development
POD	Point of departure
QRA	Quantitative risk assessment
SARA	Skin Allergy Risk Assessment
SCCS	Scientific Committee on Consumer Safety
UF	Uncertainty factor
UN	United Nations
WoE	Weight of evidence.

Executive Summary

The induction of skin sensitisation is a key adverse health effect to be addressed in the safety assessment of cosmetic ingredients. Regulatory and ethical demands have driven the development of Next Generation Risk Assessment (NGRA) based New Approach Methodologies (NAM) and Defined Approaches (DAs) to replace animal models. The revised SCCS Notes of Guidance contain a NGRA framework providing guidance for the skin sensitisation risk assessment of cosmetic ingredients. This illustrative IATA aims to demonstrate the applicability of this tiered NGRA framework to assess the potential risk from consumer exposure to geraniol at 0.1% via a face cream. In Tier 0 the consumer exposure was determined to be 2.73 µg/cm^2 based on the face cream use scenario. For this illustrative case study, it was assumed that geraniol is a new chemical for which no *in vivo* toxicity data and no historic human evidence were available. In addition, read across was not considered. The collection of existing NAM information for geraniol included structural formula, physicochemical properties, *in silico* predictions *in chemico* / *in vitro* data. Based on this information, geraniol was hypothesised to be a skin sensitiser (Tier 1) thus to progress the risk assessment potency information was required. The information collected in Tier 0 was used as input data in five DAs to demonstrate the integration of their hazard and potency predictions in a weight of evidence-based point of departure (POD) or risk determination. The DAs used here predicted geraniol to be a skin sensitiser, with a weak/moderate potency, or a GHS Cat. 1B classification. In Tier 2 the predictions from the DAs were converted into POD values, which ranged from >250 µg/cm^2 to 4600 µg/cm^2. A comparison of the POD values with the consumer exposure showed a margin of exposure (MOE) ranging from 91.6 to 1700. Based on these MOEs, risk predictions within this IATA, geraniol at the outlined exposure scenario was considered as safe or borderline safe based upon the individual DAs in terms of a potential risk of induction of skin sensitisation.

1 Introduction

Cosmetic products must undergo a risk assessment for critical health effects to make sure they are safe for consumers before placed onto the market. Since the main route of exposure to cosmetics is via dermal application, skin sensitisation and the subsequent development of allergic contact dermatitis (ACD) is one of the key adverse health effects to be addressed (Zirwas, 2019) (Alinaghi et al. 2019) (SCCS, 2021).

ACD is a T cell-mediated hypersensitivity reaction which consists of an induction and an elicitation phase. The induction of skin sensitisation occurs if a susceptible individual is exposed to a quantity of a contact allergen sufficient to induce a chemical specific T cell activation. In the elicitation phase, ACD is triggered when a previously sensitised individual is re-exposed to the chemical, which leads to the characteristic skin reactions of ACD (Martin, 2015).

Risk assessment for ingredients in cosmetic products ensures that consumer exposure does not lead to the induction of skin sensitisation (Api et al. 2008) (Goebel et al. 2012) (SCCS, 2018). A quantitative risk assessment (QRA) for skin sensitisation follows the same elements of risk assessment as for other adverse health effects, i.e., determination of consumer exposure; hazard identification; hazard characterisation and establishment of a dose response or potency and finally a risk characterisation.

Driven by advances in the mechanistic understanding of ACD development and political, regulatory and ethical demands, New Approach Methodologies (NAMs), which are non-animal-based approaches, tests or assays (e.g., *in vitro* testing in cell lines or 3D tissues, *in chemico* reactivity measurements and *in silico* predictions) are currently the preferred choice when generating new data for skin sensitisation hazard identification and potency characterisation (ECHA, 2008).

Investigation of skin sensitisation induction resulted in the identification of four mechanistic key events (KEs) as outlined in the “Adverse Outcome Pathway (AOP) for Skin Sensitisation Initiated by Covalent Binding to Proteins” (OECD 2014), respectively, (KE1) the binding of haptens to proteins of the skin; (KE2) the activation of keratinocytes; (KE3) the activation of dendritic cells and (KE4) the proliferation of antigen-specific T-cells. These four KEs have become fundamental for the application of NAMs to assess skin sensitisation (Ezendam et al. 2016) (Reisinger et al. 2015).

The current consensus is that more than one NAM should be applied to cover the skin sensitisation induction mechanism and obtain a similar level of information for hazard identification or potency characterisation as from historical animal models (Jowsey et al. 2006). Therefore, NAMs have been combined via Defined Approaches (DAs) to derive skin sensitisation hazard or potency predictions (OECD, 2016; 2017) (Kleinstreuer et al. 2018). This combination is done following a specific data integration procedure (DIP), i.e., an algorithm, applied to data generated in a defined set of NAMs to derive a prediction. A defined approach (DA) prediction can be used on its own or in integrated approaches to testing and assessment (IATA) to support regulatory decision making (Tollefsen et al. 2014) (OECD, 2016; 2017). Recently the OECD approved the first DA-based guideline for skin sensitisation (Guideline 497) (OECD, 2021b).

Cosmetics Europe (CE) has implemented a scientific research program to foster the development, assessment and application of NAMs in human health risk assessments and to support their regulatory acceptance (Desprez et al. 2018). For skin sensitisation, the experience obtained from several case

studies exploring the use of NAMs and DAs in the risk assessment of cosmetic ingredients, allowed to identify overarching principles and develop a next generation risk assessment (NGRA) framework (Gilmour and Kern et al. 2020). The tiered framework is based upon principles outlined by the International Cooperation on Cosmetics Regulation (ICCR) for NGRA approaches (Dent et al. 2018) and provides guidance on how NAMs and DAs can be used alongside historical *in vivo* data aligned to the four elements of risk assessment.

The purpose of this illustrative NGRA case study was to demonstrate the applicability of the developed NGRA framework (Figure 2) to assess the potential risk from consumer exposure to geraniol at 0.1% via a face cream.

2 Purpose

2.1. Purpose of the case study

The revised Scientific Committee on Consumer Safety (SCCS) Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation (SCCS, 2021) contain a NGRA framework by (Gilmour and Kern et al. 2020) to provide guidance for the skin sensitisation safety assessment of cosmetic ingredients. The SCCS accepts NGRA submissions for ingredients to be evaluated on a case-by-case basis. The purpose of this illustrative case study is to demonstrate the applicability of the NGRA framework by following its tiered workflow to assess the potential risk from consumer exposure to the hypothetic case study; use of geraniol at 0.1% via a face cream. The IATA does not define a maximum safe use level for geraniol.

2.2. Case study chemical selection

The fragrance chemical geraniol (CAS# 106-24-1) was selected for this illustrative case study because it is commonly used in cosmetic and cleaning products and it is known that it may induce sensitisation at certain dose levels (SCCS, 2012; ECHA, 2017). The availability of existing NAM information for geraniol including a defined chemical structure, physicochemical properties, bioavailability *in silico* and *in chemico/ in vitro* data (OECD test guideline 442C, 442D and 442E) (OECD, 2018a, 2018b, 2021a) allowed application of the NGRA framework, incorporating multiple defined approaches (DAs) to assess the potential risk from consumer exposure to the hypothetic case study.

Although studies have shown that geraniol is a pre- pro-hapten (Hagvall et al. 2007; 2008), these properties played no role in the chemical selection process. Geraniol was not chosen for its fragrance properties, but solely as a case study chemical with sufficient NAM data to conduct an NGRA case study.

For this case study we purposely decided to restrict decision making solely on available NAM data for geraniol and not to use any read across approach. The aim was to show a “simple” case study to illustrate the principles of the NGRA, without considering read across. CE is currently performing other case studies that will specifically address the application of read across in skin sensitisation NGRA.

2.3. Endpoint

The endpoint of interest is skin sensitisation.

3 Reasoning for performing IATA

3.1. Reasoning

An NGRA case study example for use of Methyldibromo glutaronitrile (MDBGN) in a cosmetic product was provided in the NGRA framework publication by Gilmour and Kern et al. 2020. This provides a basis for this case study to illustrate how the NGRA framework can be used as a weight of evidence IATA to assess the potential risk from consumer exposure to geraniol at 0.1% via a face cream.

4 Background information

4.1 Skin sensitisation adverse outcome pathway (AOP)

The mechanism behind skin sensitisation and the elicitation of Allergic Contact Dermatitis (ACD) has been documented by the OECD as an Adverse Outcome Pathway (AOP) (OECD, 2014). The AOP for skin sensitisation Initiated by Covalent Binding to Proteins captures the impact of skin exposure to sensitising chemicals as a series of biological and chemical key events (KEs).

In the induction phase of skin sensitisation, the chemical or allergen penetrates the outer epidermis of the skin. During this passage, chemicals are potentially subject to biotransformation processes which can binds covalently to skin proteins of the viable cells (key event 1) to form hapten-protein conjugates, which can be immunogenic. In parallel, keratinocytes become activated and release danger signals e.g. pro-inflammatory cytokines as a response to trauma (key event 2). Next, the phenotype of dendritic cells (DC) changes by the concerted recognition of hapten-protein conjugates by MHC (major histocompatibility complex) molecules and of danger signals (key event 3). The activated DCs mobilise and migrate, after maturational changes, from the skin to the draining lymph node. In the lymph nodes, the dendritic cells display major histocompatibility complex molecules, which include part of the hapten-protein complex to naive T-lymphocytes (T-cells). This induces differentiation and proliferation of allergen chemical- specific memory T-cells, some of which re-circulate throughout the body (key event 4).

The elicitation or challenge phase occurs following a subsequent contact with the same allergen. Again, the hapten-protein conjugate is formed and subsequently taken up by epidermal dendritic cells, as well as other antigen-presenting cells. The circulating allergen-specific, activated memory T-cells are triggered to secrete specific cytokines, which induce the release of inflammatory cytokines and mobilization of cytotoxic T-cells, as well as other inflammatory cells leading to the eventual adverse outcome ACD.

The mechanistic understanding of skin sensitisation and description of the AOP has enabled the development and regulatory acceptance of a multitude of NAMs that each aim to measure the impact of chemical on one or more of the AOP KEs to distinguish sensitising from non-sensitising chemicals or to generate information on skin sensitisation potency (Figure 1).

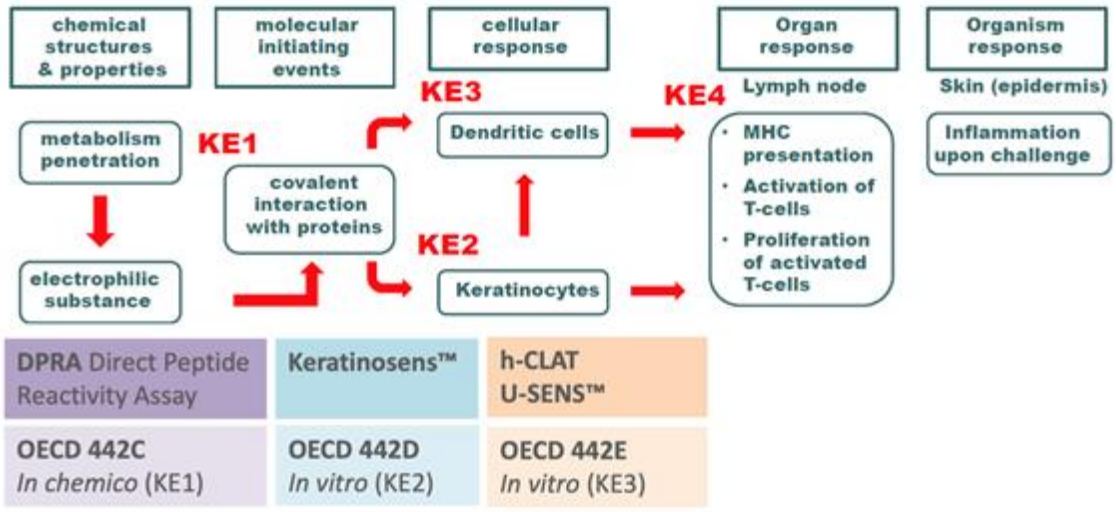


Figure 1. Overview of biological and chemical key events (KEs) as described in the skin sensitisation AOP and used NAMs in this case study covering these events.

Please note that the NAMs listed in the figure are not exhaustive and OECD test guidelines 442C, D and E include additional NAMs.

4.2. NGRA framework

The experience gained how to conduct risk assessments based upon NAMs has allowed cosmetic industries to develop a non-animal, next generation risk assessment (NGRA) framework for the assessment of skin sensitisers (Gilmour and Kern et al. 2020). The framework is based upon the principles published by the International Cooperation on Cosmetic Regulation (ICCR) and is human relevant, exposure led, hypothesis driven and designed to prevent harm. It is structured into three tiers, integrating all relevant information using a weight of evidence approach that can be iterated when new information becomes available (Figure 2). The initial tier (Tier 0) involves a thorough review of the existing information including; identification of the use scenario/consumer exposure; characterisation of the chemical purity and structure; *in silico* predictions; existing data pertaining to skin sensitisation hazard (historical or non-animal); the identification of suitable read-across candidates with supporting hazard identification/characterisation information and application of exposure-based waiving. Considering all information identified in TIER 0, the next step is the generation of a hypothesis (Tier 1). All data are considered in an exposure-led weight of evidence (WoE) approach, taking an initial view on whether a chemical is likely to be a skin sensitiser or not, choice of defined approach and availability of read-across candidates. If existing information is insufficient for concluding the risk assessment, the generation of additional information may be required to proceed (Tier 2). Such targeted testing could involve refinement of the exposure estimation or generation of data from *in vitro* or *in chemico* NAMs. Once sufficient information is available, the final stage of the NGRA framework is the determination of a point of departure (POD), characterising uncertainty and comparing to the consumer exposure in a WoE. Thorough evaluation of the sources of uncertainty is essential to ensure transparency and build trust in NGRA approaches. A more detailed description of each Tier and step in the NGRA framework can be found in the publication by Gilmour and Kern et al. 2020.

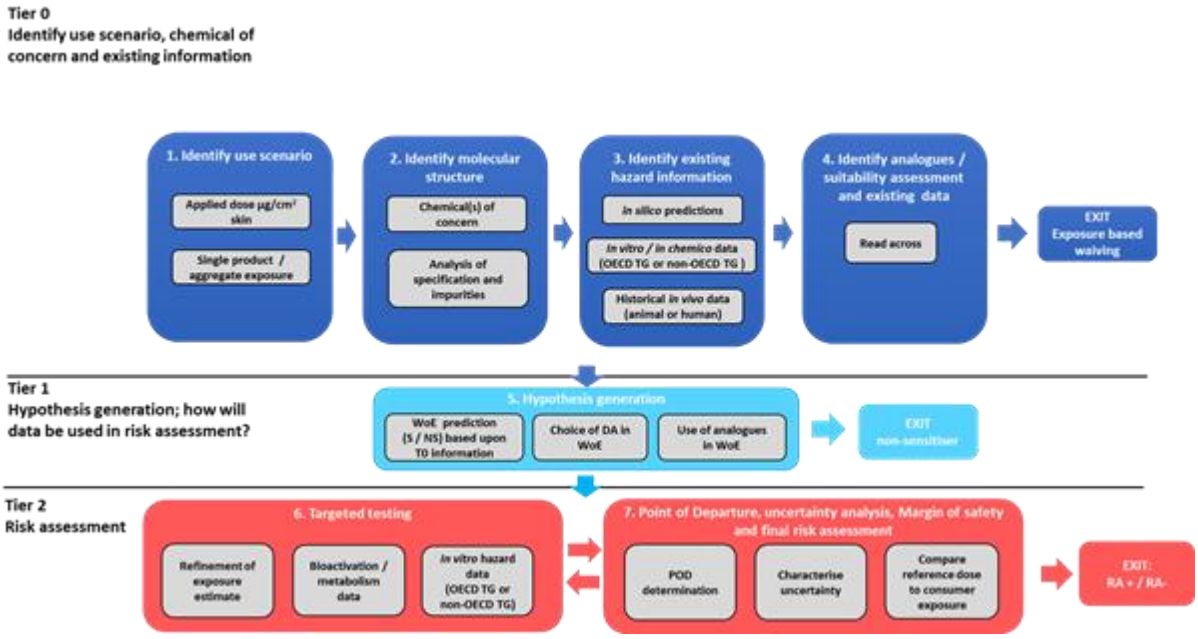


Figure 2. Next generation risk assessment (NGRA) framework for skin sensitisation.
Adopted from Gilmour and Kern et al. 2020.

5 Application of the NGRA framework

TIER 0

Identify use scenario and consumer exposure

The hypothetical use scenario for this illustrative case study was identified as 0.1% geraniol in a face cream. On the basis of an estimated daily face cream amount of 1.54 g, a concentration level of 0.1%, a skin retention factor of 1 and skin surface area of 565 cm^2, the consumer exposure was determined to be 2.73 µg/cm^2 (SCCS, 2021).

Identify chemical of interest and molecular structure

Molecular structure and physicochemical properties for geraniol were identified (Table 1***)***.

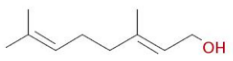
Identify existing hazard information.

The existing *in silico*, *in vitro* and *in chemico* information regarding the skin sensitisation hazard and potency for geraniol are presented in Table 1.

For this illustrative case study, it was assumed that geraniol is a new chemical for which no *in vivo* toxicity data and no historic human evidence are available, and for the purpose of this case study no read across analogues were considered.

Table 1. Existing information collected for geraniol in TIER 0 of the NGRA framework.

Information was obtained from the Cosmetics Europe database (Hoffmann et al. 2018), 1 Volatility class was calculated using a method by Spicer, 2002, *Data from Hewitt et al. 2020, For more information on the individual in silico and in vitro information sources see Annex I**: Individual information sources used**.

Names	Geraniol, (2E)-3,7-dimethylocta-2,6-dien-1-ol
CAS number	106-24-1
SMILES	CC(='CCC'/C(='C'/CO)/C)C
Structural formula	
Physicochemical properties	Molecular weight: 154.25 Da LogP: 3.0 LogS: -3.19

	LogVP: -1.52 Boiling pt. [°C]: 230 Melting pt. [°C]: -15 Volatility[1]: semi-volatile pH: 8.3 LogD @ pH 7: 3.3 H2O solubility @ pH 7: 4.4 mg/L Plasma protein binding (% bound): 95.5 Fraction ionized: 0
Mechanism	Pre-Michael acceptor Nucleophilic substitution (SN2)
TIMES-SS	Parent: Non-sensitiser Metabolite: Strong sensitiser (active after auto-oxidation)
TOXTREE	Schiff base
DEREK Nexus	Positive
DPRA	Negative (Cys depl: 0% and Lys depl: 10%)
KeratinoSens™	Positive (EC1.5: 110 µM, EC3: >2000 µM, IC50%: 875 µM)
U-SENS™	Positive (CD86 EC150: 53.6 µg/ml, CV70: 133.90 µg/ml)
h-CLAT	Positive (CD86 EC150: 123 µg/ml, CD54 EC200 - µg/ml, CV75: 139.2 µg/ml)
SENS-IS	Moderate sensitiser
Bioavailability	*Good skin penetration in 24h leave-on frozen skin study: bioavailable

Identify analogues/suitability assessment and existing data

Not applicable. We purposely decided to restrict decision making solely on available NAM data and read across was beyond the scope for this IATA case study. CE is currently performing other case studies that will specifically address the application of read across in skin sensitisation NGRA.

TIER 1

Hypothesis generation

Considering the positive skin sensitisation predictions from NAM information: *in silico* (TIMES-SS, DEREK, TOXTREE*), in vitro* (KeratinoSens™, U-SENS™, h-CLAT, and SENS-IS) and good bioavailability, the weight of evidence suggests that geraniol is most likely a skin sensitiser. Whilst the DPRA result for geraniol is negative, this result is likely due to the fact that the chemical is a pre-hapten (as predicted by TIMES), which are often not directly chemically reactive. Lack of sufficient oxidation or lack of metabolic capacity of this *in chemico* assay does not allow the chemical to be oxidized to a reactive chemical (e.g., an aldehyde). Altogether geraniol is a suspected pre-hapten requiring oxidation to become a reactive species. Confidence in the hypothesis that geraniol is a skin sensitiser is medium to high due to the majority of the NAM data indicating a sensitiser hazard. The existing information was considered sufficient to continue with the NGRA workflow without generation of additional information at this stage.

Choice of DAs in WoE

Our working hypothesis is that geraniol is a skin sensitiser, so in order to complete a risk assessment for the given exposure scenario using the available NAM data, it is necessary to derive potency information and subsequently a point of departure. Through the process of designing the NGRA framework (Gilmour and Kern et al. 2020), five DAs were considered as suitable, useful and accessible for industry risk assessors.

The DAs have been submitted as case studies to OECD (OECD, 2016; 2017) and supported a previous case study on MDBGN (Gilmour and Kern et al. 2020) demonstrating the application of individual DA predictions in individual WoE-based POD derivations for the risk assessment. The choice for selecting these five DAs was based on a number of factors, as outlined in the case study document and in Gilmour and Kern, et al. 2020 and below for this specific case study:

- Coverage of various KEs of the skin sensitisation AOP (see DA descriptions below).
- DAs provide information on skin sensitisation potency to support the derivation of a POD.
- All input data for the DAs were available.
- Geraniol falls within the applicability domains of the selected DAs.
- DAs were all accessible to CE. Including licences for *in silico* NAM inputs.

DAs have been used for a previous case study on MDBGN (Gilmour and Kern et al. 2020).The intent of this case study is to show that all of these DAs can be used in an IATA and to evaluate if the final risk assessment decision is the same. There is no need to use more than one DA. The individual DAs are described in the sections below, more detailed information about their development and construction can be found in the OECD case study document (OECD, 2017) and cited references. The individual data sources used in the DAs have been described in detail in Annex I: Individual information sources used.

6 Defined Approaches

6.1: Sensitiser potency categorization based on test methods addressing KE 1+3 and *in silico* prediction (ITSv1 DA)

Summary

The DA is constructed as ITSv1 DA for prediction of the skin sensitisation potential and potency of a substance. The ITSv1 DA which was recently adopted as an OECD guideline No. 497 (OECD, 2021b) includes an *in silico* prediction (Derek Nexus) and uses test methods that address two key events (KEs) 1 and 3 as defined in OECD Adverse Outcome Pathway (AOP) of skin sensitisation: KE1 of protein binding is evaluated using the Direct Peptide Reactivity Assay (DPRA; OECD TG 442C); KE3 of dendritic cell activation is evaluated using the human cell line activation test (h-CLAT; OECD TG 442E). This ITSv1 DA described here is originally based on 139 chemicals classified as Sensitisers / Non-Sensitisers (S/NS) in the LLNA (Takenouchi et al. 2015). 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 information sources can be used to predict the skin sensitising potential (hazard identification; S/NS) and potency of a substance. The potency prediction is given as sub-categorisation according to the UN GHS: UN GHS Cat. 1A (EC3<=2% in LLNA), UN GHS Cat. 1B (EC3>2%), Not Classified (Non-Sensitisers (NS)).

Rationale underlying the construction of the defined approach

Based on the adverse outcome pathway of skin sensitisation defined by OECD, the molecular initiating event (KE1) and the cellular response of dendritic cells (KE3) are taken into account in this ITS. KE1 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 DPRA. 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 KE3 (dendritic cell activation) of the skin sensitisation AOP and is OECD TG 442E. In the ITSv1 DA, the assay related to KE2 is not included, but DPRA cysteine depletion (KE1) and KeratinoSens™ covering KE2 are mechanistically relevant (Jaworska 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™ (KE2).

In the ITSv1 DA, 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<=2% in LLNA (Strong), EC3>2% (Weak to moderate), Non-Sensitisers (NS)) of a substance. The underlying rationale of this ITSv1 DA 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.

Description of the individual information sources used

1. Derek Nexus: *in silico* knowledge-based toxicity alerting software comprising alerts on skin sensitisation (version 6.1.0 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 likelihood level is associated. Substances with causative structural alert(s) (i.e., certain, probable, plausible, or equivocal) are conservatively considered to be positive.

2. DPRA 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, 2021a). Substances that induced mean peptide depletion of cysteine- and lysine-containing peptide above 6.38% (or in the case of co-elution, cysteine-only depletion above 13.89%) are considered to have peptide reactivity of sensitiser.

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).

Data interpretation procedure 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 in Table 2. 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 cut-offs 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 guideline 497. 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. If co-elution occurs with cysteine or both peptides, the result is inconclusive. 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: UN GHS Cat. 1A, Cat. 1B, 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 6 or 7 is defined as a strong (UN GHS Cat. 1A) sensitiser; score of 5, 4, 3, or 2 as a weak to moderate (UN GHS Cat. 1B) sensitiser; score of 1 or 0 as not-classified (non-sensitiser). The summed score yields a qualitative result (positive/negative and three rank classes).

Table 2. Conversion of the outcome in h-CLAT, DPRA, and DEREK for the ITSv1 DA.

Score	h-CLAT MIT (µg/mL)	DPRA Depletion (%) (Cysteine-only)	Derek Nexus
3	≤10	≥42.47 (≥98.24)	
2	>10, ≤150	≥22.62, <42.47 (≥23.09, <98.24)	
1	>150, ≤5000	≥6.38, <22.62 (≥13.89, <23.09)	Positive
0	Not calculated	<6.38 (<13.89)	Negative
	Potency	**Total Battery Score**	
	UN GHS Cat. 1A	6-7	
	UN GHS Cat. 1B	2-5	
	Not classified	0-1	

Limitations in the application of the defined approach

The limitations on the individual *in chemico* and *in vitro* test methods are described in the individual test guidelines (TG 442C, TG 442E). Chemicals that fall outside the applicability domains of the DPRA and h-CLAT cannot be applicable to the ITSv1 DA.

For *in silico* information source predictions from Derek Nexus, negative predictions that contain misclassified and/or unclassified features are considered to be outside applicability domain.

The level of confidence of the ITSv1 DA is dependent on the total battery score and applicability domain of the individual information sources. The DA predictions with low confidence are considered inconclusive for prediction of the hazard potential and/or potency sub-categorisation.

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 ITSv1 DA, the individual limitation can be minimized and the ITSv1 DA can lead to correct classification of pre- /pro-haptens.

Consideration of uncertainties associated with the application of the defined approach

DIP structure

- KE4 is not included due to lack of available tests.
- The ITSv1 DA covers KE1 and KE3 of AOP and is originally based on a dataset of 139 chemicals.
- In some cases, the confidence is lower for chemicals with log P > 3.5. Negative results for chemicals with LogP >3.5 should not be considered (OECD, 2021b).
- The confidence might be lower for pre-haptens and pro-haptens due to limited metabolic capacities of test methods.

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 high.

Benchmark data

- The ITSv1 DA 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.

Impact of uncertainty on the DIP's prediction

Some uncertainty might cause under- or over-estimation of hazard identification and potency classification for the ITSv1 DA. For a new test chemical, similar chemicals with *in vitro* or *in vivo* data should be checked.

Prediction for Geraniol

Geraniol was predicted to be a GHS Cat. 1B skin sensitiser based on an ITS score of 3. Because the score of 3 was well within the 2-5 GHS Cat. 1B range, and all in chemico/in vitro NAM outcomes were applicable there was a high confidence in this DA prediction.

6.2 Sensitiser potency prediction based on KE 1+2+3: The artificial neural network model for predicting LLNA EC3 (ANN)

Summary

The DA describes an integrated testing strategy for prediction of the skin sensitisation potential and potency of a substance (Hirota et al. 2018). The line of evidence predicted LLNA (primary target). The combination of test methods used covers the first three KEs of the AOP leading to skin sensitisation as formally described by the OECD: KE 1: protein binding (e.g. via the DPRA; OECD TG 442 C); KE 2: keratinocyte activation (e.g. via the KeratinoSens™; OECD TG 442D); and KE3: dendritic cell activation (e.g. via the h-CLAT; OECD TG 442E). TIMES-SS prediction of the target molecule is also used as a descriptor. The DA described here is based on 134 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 0.91, and RMS error was 0.47. 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 four classifications (extreme and strong, moderate, weak, and non-sensitiser), accuracy is 67.9%. Therefore, this approach may be used to subcategorise skin sensitisers in cat 1A and 1B for GHS/CLP.

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 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 KeratinoSens™ (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). TIMES-SS prediction is used as *in silico* parameter. 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. 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).

Description of the individual information sources used

A) TIMES-SS: *in silico* parameter.

B) AOP key event 1: i.e. DPRA is addressing a protein binding potential.

C) AOP Key event 2: i.e. KeratinoSens™ is addressing an activation of the keratinocyte (Natsch and Emter, 2008).

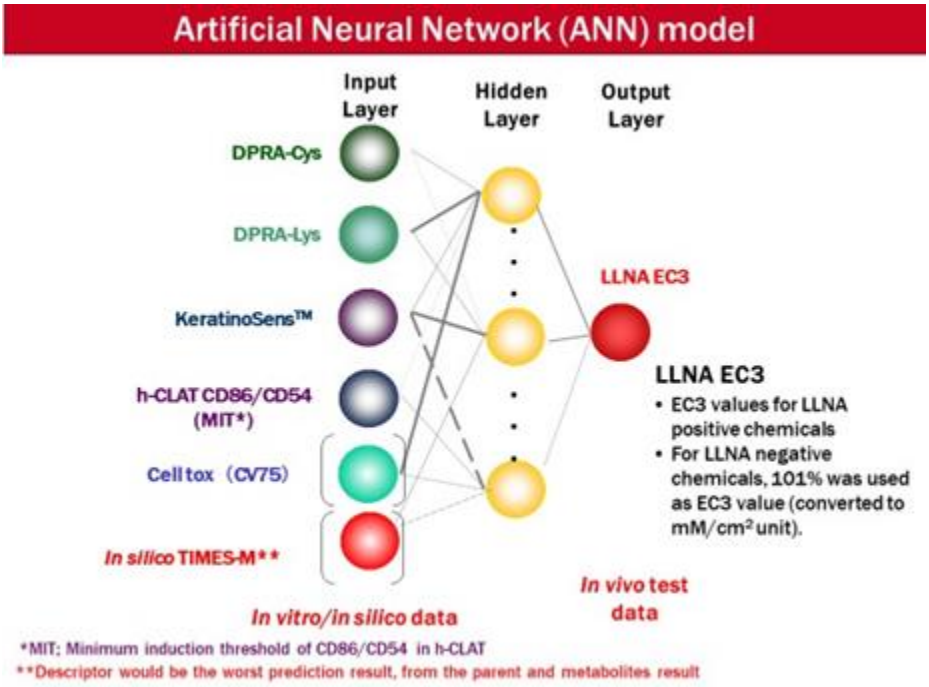
D) AOP Key event 3: i.e. h-CLAT: human Cell Line Activation Test (*in vitro*) h-CLAT is addressing a dendritic cell activation.

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).

Data interpretation procedure 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. Parameters resulting from the three *in vitro* tests and *in silico* data (TIMES-SS) 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.

Figure 3. Conceptual diagram of the artificial neural network consisting of input layer, hidden layer and output layer.
The LLNA Threshold is EC3 value (%).



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.

- The strengths and technical limitations on the individual test inputs are detailed in the respective individual data sources.
- 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.

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™ use cysteine reactivity as a read-out.

In silico limitations:

- It is impossible to see how much each information source contributed to the decision.
- The number of the dataset is limited (134 chemicals). Increasing the size of the dataset might be effective to improve the predictive capacity.

Consideration of uncertainties associated with the application of the defined approach

DIP 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 P > 3.5).
- The confidence is lower for pre-haptens and pro-haptens due to limited metabolic capacities of test methods.
- The predictive capacity can vary depending on the dataset used.
- 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.

The information sources used within the defined approach

- The DPRA, KeratinoSens™ and h-CLAT are approved within OECD test guidelines and have been validated by ECVAM. Reproducibility of peptide reactivity and CD86/CD54 measurements are high.

Benchmark data

- The Integrated Testing Strategy (ITS) for hazard identification is based on LLNA data. 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.

Impact of uncertainty on the DIP's prediction

The different sources of uncertainty could cause to under- or over-estimate skin sensitisation potential.

Prediction for Geraniol

Geraniol was predicted to be a skin sensitiser with an EC3 value: 10.1%. Because all in chemico/in vitro/in silico parameters were applicable, there was a high confidence in this DA prediction.

6.3: Sequential testing strategy for hazard identification (Tier 1) and potency (UN GHS cat. 1A / 1B) categorization (Tier 2) of skin sensitisation

Summary

In this sequential testing strategy DA, hazard identification and potency (UN GHS cat. 1A/1B/no Cat.) categorization is based on the combination of multiple *in vitro* and *in silico* parameters covering the Adverse Outcome Pathway (AOP's) key events 1 to 3 leading to skin sensitisation. The DA is constructed as a tiered approach with a decision point at the end of each tier, allowing stepwise and efficient information gathering (Figure 4). The first tier, which combines protein reactivity and skin sensitisation predictions (TIMES-SS, Toxtree), Direct Peptide Reactivity Assay (DPRA), U-SENS™ and KeratinoSens™ data as well as physicochemical parameters (pH, volatility), was built on 219 chemicals having a LLNA-based No Cat./Cat.1 classification (Del Bufalo et al. 2018; Tourneix et al. 2019). The

second tier, which combines physicochemical parameters (Molecular Weight, volatility and clogP) as well as DPRA, U-SENS™, KeratinoSens™ and optionally SENS-IS data, was built on 100 chemicals having a LLNA-based UN GHS Cat. 1A/1B classification. For each of those tiers, the combination of the different input parameters was achieved using a meta-model stacking five different statistical methods (Boosting, Naïve Bayes, Support Vector Machine (SVM), Sparse PLS-DA and Expert Scoring), providing a probability to belong to the group of interest ("to be a sensitiser" Tier 1, "to be a cat. 1A" Tier 2).

Rationale underlying the construction of the defined approach

The sequential testing strategy was constructed to allow a stepwise gathering of information using a tiered approach on skin sensitisation hazard (Tier 1) and potency (Tier 2) (Gautier et al. 2020; Assaf Vandecasteele et al. 2021). Both tiers integrate information covering the skin sensitisation AOP KE1 (MIE), KE2 and KE3. As such, the sequential testing strategy is based on both intrinsic physicochemical properties of the chemical and descriptors of early innate immune cell responses key events, as described below.

Intrinsic physicochemical properties of the chemical

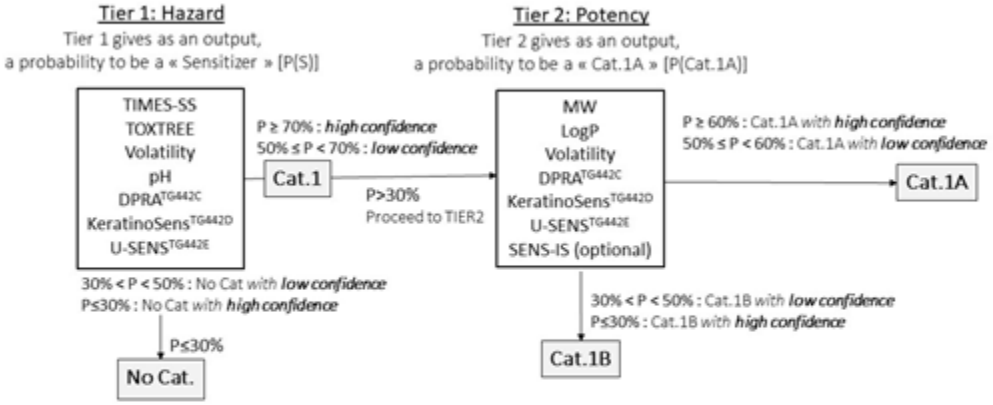
- Descriptors allowing to integrate stability and/or bioavailability characteristics. As such, the measured pH and the calculated volatility, cLogP and MW were considered as relevant variables to combine with *in silico*, *in chemico* and *in vitro* NAM, as defined in a splitting statistical analysis (Gomes et al. 2012). See individual information sources for rationale description ("Mechanistic basis including AOP coverage" section).
- Chemical reactivity (which is directly linked to the initial key event: haptenation of skin proteins): the Toxtree protein binding alerts. The Times-SS predictions also mainly take into account electrophilic binding to skin proteins either directly or following metabolism. Finally, the *in chemico* DPRA (OECD Test Guidelines 442C), related to AOP key event 1, is a method giving a measurement of MIEs as cysteine and lysine peptides modifications by the chemical.

Descriptors of early innate immune cell responses

- AOP key event 2: i.e. keratinocytes activation, with the KeratinoSens™ assay assessing the induction of the Nrf-2 pathway (OECD Test Guidelines 442D), and optionally the SENS-IS™ assay assessing the transcriptomic activation of keratinocytes within a reconstructed human epidermis (Cottrez et al. 2015; 2016).
- AOP 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; Alépée et al. 2017).

In both tiers, the combination of the different input parameters was achieved using a stacking meta-model. 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; Nocairi et al. 2016). These methods have strong differences, but they all produce posterior probability of belonging to the group of interest. Therefore, two stacking models (Tier 1 "to be a sensitiser" and Tier 2 "to be a UN GHS cat. 1A") were built independently on a proper training set (based on LLNA data). Instead of trying to choose a specific method, these methods were combined by the stacking methodology of Wolpert (1992) and Breiman (1996), in order to obtain a specific stacking meta-model for each tier.

Figure 4. Schematic representation of sequential testing strategy for hazard identification (Tier 1) and potency (UN GHS cat. 1A / 1B) categorization (Tier 2) of skin sensitisation.
There is the option to also use of SENS-IS in Tier 2.



Description of the individual information sources used

The individual data sources have been described in Annex I: Individual information sources used. The respective input parameters from the NAM that were combined for the construction of the two tiers are given in Table 3 below.

Table 3. Input parameters of each tier of the sequential testing strategy.
INC: Inconclusive; R: reactive; NR: No or minimal Reactivity; LR: Low Reactivity; MR: Moderate Reactivity; HR: High Reactivity; P: Positive; N: Negative. *SENS-IS used in case of borderline Tier 2 predictions.

Individual sources	Tier 1 (Hazard)	Tier 2 (Potency)
Physicochemical properties	-Volatility class (Very-volatile, Semi-volatile, Non-volatile, INC (Qualitative) - pH (Quantitative)	- Volatility class: Very-volatile, Semi-volatile, Non-volatile, INC (Qualitative) - cLogP (Quantitative) - MW (Quantitative)
TIMES-SS	- P/N/INC outcome (Qualitative)	
ToxTree	- R/NR outcome (Qualitative)	
DPRA	- R/NR/INC (Qualitative)	- Reactivity classes: HR, MR, LR (Qualitative)
KeratinoSens™	- P/N/INC outcome (Qualitative)	- $EC_{1.5}$ in µM (Quantitative)
U-SENS™	- P/N outcome (Qualitative)	- EC 150 in µg/mL (Quantitative)
SENS-IS™ (optional*)		- Potency class: weak, moderate, strong (Qualitative)

Data interpretation procedure applied

The respective input parameters of each tier (Table 3) are entered into the model where they are run in five different supervised classification models (Boosting, Naïve Bayes, SVM, Sparse PLS-DA and Expert Scoring), each of it providing a probability of being a skin sensitiser (Tier 1) or a probability of being a UN GHS cat. 1A skin sensitiser (Tier 2). These intermediate probabilities that are evidently highly positively correlated (Gomes et al. 2012; Nocairi et al. 2016) are then used in a stacking meta-model that provides a final probability to be a skin sensitiser (Tier 1) or to be a UN GHS Cat. 1A skin sensitiser (tier 2) (primary outcomes of the meta-models). Optimal predictive capacities based on LLNA reference data were obtained by setting the following probability thresholds:

Tier 1 (Hazard identification):

- Chemicals with probability to be sensitiser ≥ 70% are predicted “Sensitiser” with high confidence;
- Chemicals with probability to be sensitiser comprised between ≥50% and <70% are predicted “Sensitiser” with low confidence;
- Chemicals with probability to be sensitiser comprised between (>30 and < 50%) are predicted “Non-Sensitiser” with low confidence;
- Chemicals with probability to be sensitiser ≤ 30% are predicted “Non-Sensitiser” with high confidence.

Tier 2 (Potency prediction):

- Chemicals with probability to be sensitiser ≥ 60% are predicted “UN GHS cat.1A” with high confidence;
- Chemicals with probability to be sensitiser comprised between ≥50% and <60% are predicted “UN GHS cat.1A” with low confidence;
- Chemicals with probability to be sensitiser comprised between (>30 and < 50%) are predicted “UN GHS cat. 1B” with low confidence;
- Chemicals with probability to be sensitiser ≤ 30% are predicted “UN GHS cat. 1B” with high confidence.

Based on these predictions, the decision rules for a sequential testing strategy are the following:

Tier 1 (Hazard identification):

- Chemicals with a probability to be sensitiser ≤ 30%, are classified “Non-Sensitiser” with high confidence and no further testing is needed.
- Chemicals with a probability > 30%, proceed to Tier 2.

Tier 2 (Potency prediction):

- Chemicals with a probability to be UN GHS cat. 1A ≥ 60 % are classified “UN GHS cat. 1A” with high confidence.
- Chemicals with probability to be sensitiser comprised between ≥50% and <60% are predicted “UN GHS cat.1A” with low confidence;
- Chemicals with probability to be sensitiser comprised between (>30 and < 50%) are predicted “UN GHS cat. 1B” with low confidence;
- Chemicals with a probability ≤ 30% are classified “UN GHS cat. 1B” with high confidence.

Limitations in the application of the defined approach

The strengths and limitations on the 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 on the individual data sources. By integrating the different individual data sources, the stacking meta model minimises individual limitations and allows a correct classification of pre- pro-haptens, dyes and low soluble chemicals. DA can be used for chemicals with defined structures, for which physicochemical properties can be calculated and *in silico* predictions (TIMES-SS, TOXTREE) can be made. Complex mixtures without a defined structure and a defined MW could not be processed in the sequential testing strategy. Overall applicability domain of the defined approach comprises several classes of cosmetic chemicals (fragrances, dyes, preservatives, actives, surfactants and Ultra-Violet filters), and non-cosmetic organic chemicals. Results should be interpreted with care for agrochemicals, metals, nanomaterials, since the representation of these categories in the training set is low or even absent. The DA is not applicable for gases.

Consideration of uncertainties associated with the application of the defined approach

DIP structure

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 far from perfect.

Phase I metabolic pathways are not fully represented.

The information sources used within the defined approach

In the DA, 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, KeratinoSens™ and U-SENS™ assays have been regulatory accepted by OECD (TGs 442C, 442D, 442E respectively).

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

- Inconsistent results in the source data for a given chemical would reduce the confidence in the hazard predictions.
- Volatility was predicted rather than measured.
- Results from TIMES for predicted auto-oxidation products or skin metabolites may rely on those that are not biologically important.

Benchmark data

The benchmark data used to develop the test methods (219 chemicals for Tier1 and 100 chemicals for Tier2) 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% (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 been reported (Dumont et al. 2016). Around those uncertainties, the defined approach was developed using the most prevalent reference result for LLNA hazard (Tier1) and UN GHS Cat. 1A/1B potency classification when multiple data were available.

Impact of uncertainty on the DIP's prediction

The meta stacking model approach is designed to account for uncertainties in the underlying data and is directly incorporated into the final probability prediction (with low and high probability confidence). The user will have higher confidence in predictions by those chemicals:

- That lie within the applicability domain of *in silico* tools, *in vitro* tools.
- That are performed without any technical limitations in the running of *in chemico* or *in vitro* methods.
- That are direct acting in nature and do not require metabolic or chemical activation.

On the other hand, predictions with lower confidence are those where the substances fall outside of the TIMES-SS 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.

If the probability to be a GHS category 1A sensitiser in Tier 2 is very close to the cut offs, the option to integrate a SENS-IS potency class may be used (Table 3). The use of SENS-IS data in Tier 2 is not systematically done, but only in case of borderline predictions, where the inclusion of additional evidence may support classification.

Prediction for Geraniol

The Tier 1 of the DA predicted Geraniol as a skin sensitiser with an 89.63% probability to be a sensitiser. The subsequent Tier 2 predicted GHS category 1B with an 9.69% probability to be a GHS category 1A sensitiser. Because the probabilities in both tiers were well above and below the prediction model cut-offs there was a high confidence in these DA predictions.

6.4. Sensitiser potency prediction based on Key event 1+2+3: Bayesian Network ITS/DS for hazard and potency identification of skin sensitisers (BN-ITS)

Summary

This DA is based on Bayesian Network (BN) 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. Cysteine and Lysine reactivity are treated as two separate, independent molecular initiating events (MIEs). BN ITS uses information on metabolic transformation and auto-oxidation from TIMES-SS in the prediction process. Bioavailability considerations are applied via physicochemical properties, to represent an estimate of the potential to penetrate the stratum corneum and the free concentration respectively. Since the BN-ITS can reason based on partial information, only relevant or available data are used for predictions. This allows explicit consideration of the applicability domains of individual assays. Data outside of domains can be excluded 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: non-sensitisers (NS), weak (W), moderate (M), strong/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 can be cumulative, it can also be used to guide and optimize testing strategies before testing is commenced.

Rationale underlying the construction of the defined approach

The BN-ITS DA uses the following data streams as input data:

- Bioavailability using physicochemical properties (logDpH=7, water solubility WspH=7, fraction ionised (calculated using LogP and LogDpH7), Protein Binding PB– ACDlabs)
- *In silico* metabolism, potential for oxidation, potency prediction (TIMES prediction)
- KE 1: Peptide reactivity [OECD 442 C: Direct peptide reactivity test (DPRA)]
- KE 2: Keratinocyte activation [OECD 442D: ARE-Nrf2 luciferase test method (KeratinoSens™)]
- KE 3: Dendritic cell activation [human cell-line activation test (h-CLAT)].

A Bayesian network is a probabilistic graphical model (a type of statistical model) that graphically represents a set of variables and their conditional dependencies (based on Jaworska et al. 2010, 2013). The structure of the BN ITS DA 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 which allows chemical specific result interpretation in the biological context. The hypothesis generated by the BN ITS model can be explained based on known mechanisms.

Both the construction method and the resulting structure (Figure 5) of the current BN ITS-4 (Kern et al. 2021 in preparation) are identical to the previous version BN ITS-3 (Jaworska et al. 2015, OECD , 2017). For the BN ITS the mechanistic scheme of the skin sensitisation induction process based on the AOP was translated into a Naïve Bayes network structure assuming 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.

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. 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.

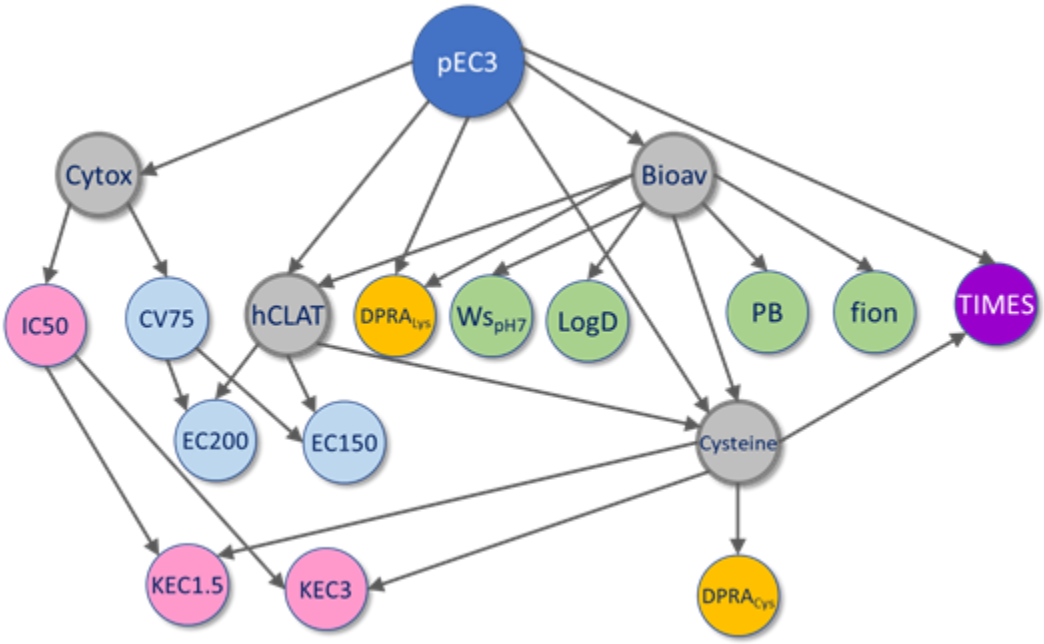
BN ITS also relates to bioavailability and cytotoxicity. To this end, we decided to generalise the bioavailability latent variable to consider skin penetration in the BN ITS framework structure. The bioavailability latent variable is constructed from the following physicochemical properties: water solubility 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.

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.

Figure 5. Structure of the BN ITS.

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.



Description of the individual information sources used

The respective input parameters are described in Table 4 below.

Table 4. Input parameters for the BN-ITS DA.

Input type	Endpoint	Unit
Bioavailability/ physicochemical properties (ACD lab)	1.Ws-Water solubility at pH='7' 2. Log D- Distribution coefficient at pH='7' 3. Plasma protein binding fraction 4. Fraction ionized	M/l [-] [-] [-]
TIMES-SS	1. Prediction of 3 classes (Non-sensitiser (1), weak (2) or moderate/strong (3) based on the most potent among parent and predicted metabolites.	Classes (NS, W, S)
KE 1: DPRACys, DPRALys	% cysteine- (Cys), and lysine- (Lys) peptide remaining in the DPRA	% 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;	µM/l

	200% induction of the cell surface activation marker CD54 in the h-CLAT; 25% reduction in cell viability in the h-CLAT.	

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:

Gathering evidence

a) Calculation of physicochemical properties of chemicals.
b) Prediction of sensitisation potency category using TIMES SS:
   i) Potency is based on the highest potency between the parent molecule and its predicted metabolites
   ii) Review potential of metabolic activation (pro-hapten) and auto-oxidation (pre-hapten) to facilitate interpretation of DPRA, KeratinoSens™ and h-CLAT assay results (not used as data input to the BN-ITS)
c) Collection of data on KE1, 2, 3 and evaluation of the completeness (in particular for MIE if Lys depletion is missing).
d) Assessment of applicability domains:
   i) TIMES-SS: predictions are used if parent is within total domain. If parent is out of total domain, metabolite predictions are only used in case of high transformation map reliability.
   ii) If the chemical is deemed a potential pre- or pro-hapten, then DPRA, KeratinoSens™, and h-CLAT data are examined with caution, against potential conflict with other data. A hypothesis without these data can be considered.
   iii) *in vitro* data are used as input if results are reliable within the limits/ applicability domains of the OECD TGs restrictions.

Integration of all relevant evidence via BN ITS and prediction of the pEC3 probability distribution.

a) Analysis of the cumulative evidence from all relevant input data: review probability distributions over 4 categories. (Highest probability most likely to determine potency category prediction).
b) Conversion of probability distribution to Bayes factors (BF) for final interpretation and acceptance of prediction. Highest BF across the 4 groups defines the final potency category. Magnitude of BF determines confidence in prediction. A flat probability curve distribution might now allow derivation of one potency class (similar BF across some categories). In that case probabilities could be pooled for the weak/ moderate and strong categories and compared with the probability for the NS category to obtain a hazard result as sensitiser/ non sensitiser.

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
P(H=not x) probability of a chemical to not to be in class x given the evidence provided to BN ITS

Table 5. 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

Limitations in the application of the defined approach

BN ITS 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) as described in the respective OECD TGs. Technical limitations when conducting the assays need to be considered. TIMES predictions for chemicals outside applicability domain (as described above) should not be used as input or taken with caution. DA can be used for chemicals with defined structures, for which physicochemical properties can be calculated and *in silico* predictions (TIMES-SS) can be made. Derivation of physicochemical properties of e.g., charged molecules should be carefully reviewed. Physicochemical data should be generated using ACD lab software. Physicochemical data generated via other sources can be used, but the BN ITS prediction should be carefully evaluated for potential impact on the predicted potency classes.

Consideration of uncertainties associated with the application of the defined approach

DIP structure

A Bayesian network approach allows converting integrated predictions to transparent, consistent decisions. It has the following features: a) it is 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 it quantifies confidence in decisions; and allows them to be fit for purpose; c) it 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.

Table 6. Uncertainty assumptions made for the BN-ITS DA.

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 shows a very stable network.
The physicochemical 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	False positive will yield overestimation, false negative will

characterized by TIMES.	yield underestimation.

The information sources used within the defined approach

In the DA, all *in chemico*/*in vitro* methods used have been shown to be reliable and relevant through multicentre studies evaluations. DPRA, KeratinoSens™, h-Clat assays have been regulatory accepted by OECD (TGs 442C, 442D, 442E respectively).

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

- Inconsistent results in the source data for a given chemical would reduce the confidence in the hazard predictions.
- Results from TIMES for predicted auto-oxidation products or skin metabolites may rely on those that are not biologically important.
- Physicochemical properties calculated from non ACD lab would reduce confidence in predictions.

Benchmark data

The benchmark data used to develop the DA are 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). Around those uncertainties, the defined approach was developed using a representative reference result for LLNA potency classification when multiple data were available.

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 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 structure and, in part, uncertainty due to the variability of input information sources as well as the target, i.e., LLNA pEC3.

Prediction for Geraniol

Geraniol was predicted as a skin sensitiser with a moderate potency. The conversion to BF to assess the strength of the evidence showed borderline weak to strong evidence.

6.5. DIP for Skin Allergy Risk Assessment (SARA)

Summary

The SARA model is a DA designed to be applied within the NGRA framework. It utilises a Bayesian statistical approach to infer a human-relevant metric of sensitiser potency and a measure of consumer risk for any given consumer exposure to a chemical of interest. It can utilise any combination of data from a HRIPT (Politano and Api 2008), LLNA (OECD, 2010), DPRA (OECD, 2021a), KeratinoSens™ (OECD 2018a), h-CLAT or U-SENS™ (OECD 2018b) to derive sensitiser potency with explicit

quantification of the uncertainty in the prediction. Furthermore, by incorporation of clinical benchmark exposures which, based upon history of use in consumer product and clinical evidence have been assigned high/low risk for induction of skin sensitisation it is also possible to calculate the probability of whether the given exposure is low risk (SARA risk metric) for a given exposure.

Rationale underlying the construction of the defined approach

The SARA model is a DA which uses Bayesian statistics to infer a human-relevant metric of sensitiser potency and allows explicit quantification of the uncertainty in potency estimates (Reynolds et al. 2019). The available inputs into the model were any combination of HRIPT (Politano and Api, 2008), LLNA (OECD, 2010), DPRA (OECD, 2021a), KeratinoSens™ (OECD, 2018a), h-CLAT or U-SENS™ (OECD, 2018b) data. Briefly, features of this model include:

- a dataset of 30 chemicals
- sensitiser potency metric defined as the maximum HRIPT dose at which there is no chance of inducing sensitisation
- modelling assumptions pertaining to:
    - population variability of thresholds for sensitisation
    - sampling variability in the HRIPT
    - relationships between the potency metric and data generated in the LLNA and *in vitro* assays for skin sensitisation
    - explicit modelling of variability of LLNA data, but not *in vitro* data.

Since publication, the model has undergone substantial revision. The updated SARA model not only provides a human-relevant metric of sensitiser potency but also provides a probability that a consumer exposure scenario can be classified as low risk, hereon termed the SARA risk metric (Reynolds and Gilmour et al. 2021 submitted). Briefly, major updates include:

- Redefining the metric of sensitiser potency from a HRIPT dose with a 0% chance of sensitisation, to a dose with a 1% chance of sensitisation (termed the ED_{01} with units $\mu g/cm^2$). The latter metric can be inferred more reliably from the data.
- Expansion of the dataset underpinning the model, from 30 chemicals to 81 chemicals. The dataset is now restricted to chemicals which are currently, or have historically been, used in consumer (household or personal care) goods.
- Expansion of the model to explicitly account for variability in the DPRA, KeratinoSens™, h-CLAT and U-Sens™. This change was necessary because, alongside the revision of the model, the *in vitro* component of the database underpinning the model was expanded to include multiple *in vitro* study results for many chemicals in the database. Variability in the historical in vivo data was a feature of the original published model.
- Expansion of the model to consider correlations between residuals within the regression components of the model. Two tests for skin sensitisation potential or potency may give a misleading result for the same reasons, for example, lack of metabolism for activation of a pre-hapten. This refinement of the model protects against overconfident estimates of potency by allowing for tests to be correlated in their predictive error.

Expansion of the model to incorporate benchmark exposure information. This component of the model is used to calculate a risk metric for an exposure of interest.

The SARA DA draws upon available information spanning all KEs (i.e. KE1: DPRA; KE2: KeratinoSens™; KE3: h-CLAT; U-Sens™; KE4: historical (LLNA) and the adverse outcome (AO) (i.e. historical HRIPT and clinical benchmarks) to infer the dose with a 1% chance of sensitisation (termed the ED01 with units µg cm-2). The underlying rationale and assumption in the proposed DA is that,

given the coverage of the AOP KEs and the well-characterised assays, these data sources contain mechanistically-relevant information for predicting human skin sensitisation potency.

Description of the individual information sources used

Table 7. Input parameters for the SARA model.

Method	Input
(Historic) Human Repeat Insult Patch Test (HRIPT)	Only studies in which cohort size, number sensitised and applied dose are reported are admissible
(Historic) Local Lymph Node Assay (OECD TG 429)	EC3 values obtained from individual studies are admissible as evidence. Representative or averaged values are not admissible
DPRA (OECD TG 442c)	Maximum % depletion with either cysteine and/or lysine
KeratinoSens™ (OECD TG 442D)	EC1.5 value (M)
h-CLAT (OECD TG 442E)	CD54 EC200, CD86 EC150
U-Sens™ (OECD TG 442E)	Reported CD86 EC150 (g ml-1)
Exposure	Dermal dose applied (µg/cm2)
Clinical benchmark	High / low risk

Data interpretation procedure applied

The SARA DA utilises a probabilistic (technically a Bayesian multilevel (i.e., hierarchical) model to characterise the relationships between data from sources identified above in Table 7.

. The approach makes a quantitative prediction of the dose with a 1% chance of sensitisation in the HRIPT (termed the ED_{01} with units µg/cm^2). The Bayesian model allows to explicitly, and quantitatively, describe the uncertainty in potency estimates, conditional on the model and available data. The ED_{01} (µg cm^2), is inferred as a probability distribution. The variance of the distribution reflects the precision in the estimate, conditional on the chemical-specific data available. The ED_{01} is the measure of hazard potency for a chemical and may be used to derive a point-of-departure for risk assessment. A margin of exposure (MOE), (i.e. the measure of how far away the consumer exposure is from the POD) is calculated and by comparison of this to the MOE for the clinical benchmarks, which are classified as either high or low risk based upon clinical experience a probability that an exposure is low risk is calculated (SARA risk metric). The SARA model structure is shown in

Table 6.

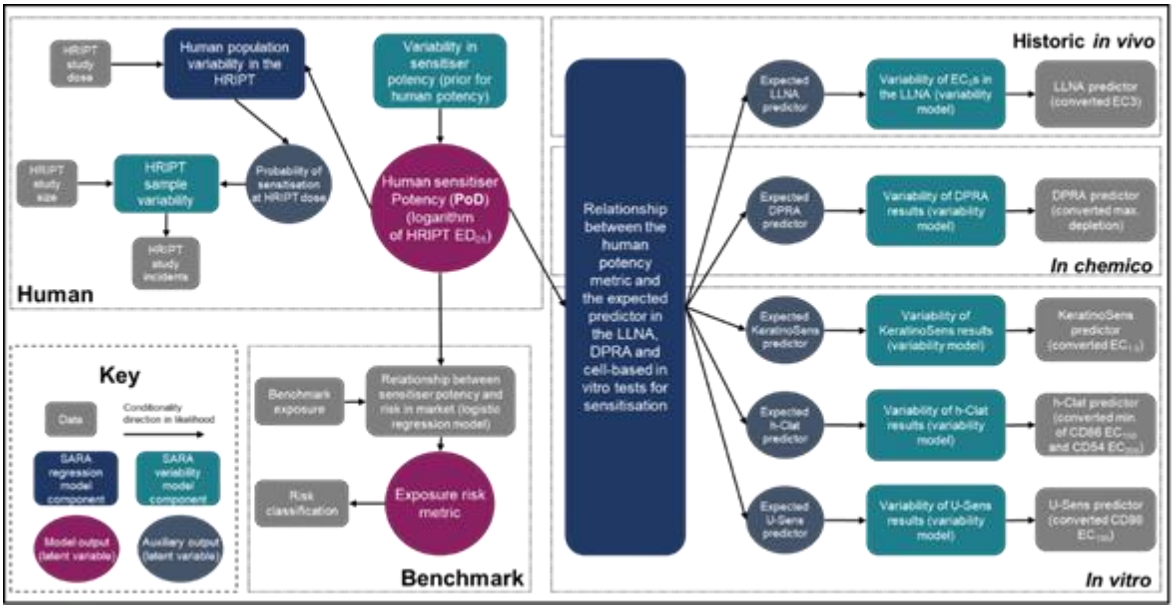
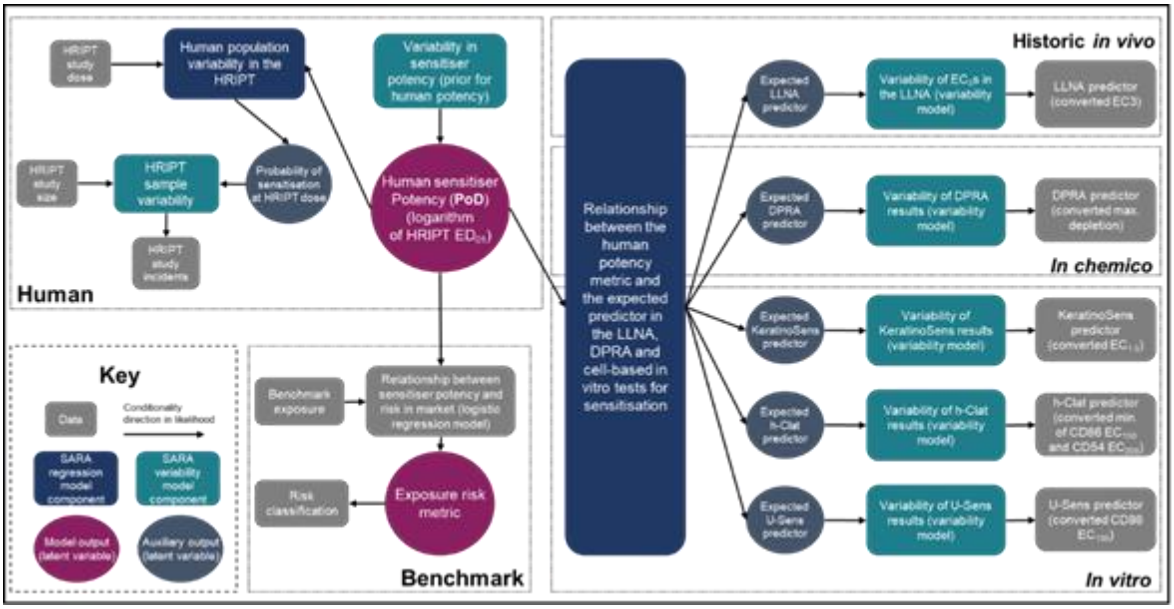


Figure 6. Schematic of the relationships between variables in the SARA DA (see supplementary material for full details of the mathematical model).



Limitations in the application of the defined approach

The various assays have technical applicability domains that limits the number of chemicals which can be tested in them. These same limits clearly apply to the DA as well. Underlying the approach is the assumption that for a chemical of interest, the relationship between the ED_{01} and the various assays is represented by the relationships learnt from the wider dataset. The approach is data driven and therefore estimates of the ED_{01} will be biased if the output of an individual assay is itself biased, i.e., a DPRA assay with 0% depletion because a sensitising chemical requires metabolic activation will bias

towards underestimating potency. It should be noted however that when data from several assays are used, the influence of a single unrepresentative assay result will be mitigated provided the other assay results are representative. Furthermore, the reliability of each assay is learnt from the dataset if there are many chemicals whose assay outcome is unrepresentative of human potency, the influence of that assay will be automatically decreased in the prediction model.

Consideration of uncertainties associated with the application of the defined approach

DIP structure

The SARA DA utilises a Bayesian multilevel modelling approach. The fundamental assumption behind this modelling approach is that there is a quantifiable relationship between the individual data sources. The rationale for this being the case is underpinned by the fact that each of these data sources corresponds to key events of the AOP or to the adverse outcome itself. However, the strength of the relationship and how informative data from one data source (e.g., *in vitro* data sources) is for predicting another (e.g., HRIPT) is determined by updating the model parameters using the underlying dataset and then assessing model fit.

The information sources within the defined approach

The DA utilises a Bayesian multilevel modelling approach that attempts to account for a) the variability in underlying observed data and b) the uncertainty in the strength of relationship that exists between the data sources. Specifically, variability of *in vitro* and historical *in vivo* data is explicitly accounted for in the model.

Benchmark data

The present dataset is limited to 81 chemicals, restricted to chemicals which are currently, or have historically been, used in consumer (household or personal care) goods. There are ongoing efforts to collate additional data. A fundamental uncertainty associated with this model (and other approaches seeking to learn from past data) is whether the data under consideration is representative of future data. For this approach, two key questions are a) is the distribution of sensitiser potency across the set of chemicals historically tested in the HRIPT representative of the distribution of sensitiser potency of future chemicals and b) will the strength of the association between sensitiser potency and the outcome of the various *in vitro* tests be conserved for future compounds. Confidence in the predictions of the SARA DA is conditional on having a reasonable level of belief that these assumptions hold. However, this is no different to any other scientific approach where the knowledge is built around a limited dataset and then extrapolation made to new scenarios.

The SARA risk metric is established based upon a set of 83 clinical benchmark exposures. The benchmark exposures have been selected based upon common established allergens which have been reviewed extensively in the literature, where it is possible to draw some conclusion about whether an exposure is high or low risk based upon epidemiological information and the profile of allergic reactions observed. The current small number of benchmarks is a limitation and increasing the number of materials within the SARA model with both high and low risk exposures will improve confidence in the reliability of the risk metric identification of additional risk benchmarks is ongoing.

Impact of uncertainty on the DIP's prediction

The modelling approach is designed to account for uncertainties in the underlying data and the relationships between the data sources. Thus, the effect of this uncertainty is directly incorporated into the final prediction of the human sensitiser potency (ED01) which is expressed as a probability distribution and SARA risk metric (probability that a given exposure is low risk) The remaining, unquantified uncertainties, relate to those discussed under benchmark data and to whether the dataset used to build the model is representative in general. This would have to be considered on a case-by-case basis and would likely take into account the representativeness of the underlying dataset with respect to the chemistry of the chemical for which a prediction is required.

Prediction for Geraniol

The expected ED_{01} is 4600 µg/cm^2. This ranks geraniol with other weak/moderate skin sensitisers (see Annex II: SARA model predictions). The probability that exposure to 0.1% geraniol in a face cream is low risk is 0.81 (i.e., 81% probability low risk). This ranks with exposures considered low risk for induction of skin allergy (e.g. use of Kathon CG in rinse off products / benzyl alcohol and phenoxyethanol in leave on products), (see Annex II: SARA model predictions).

Individual data sources used in the selected DAs for the geraniol case study

Table 8. Overview of DA input data for geraniol.

Information source	Data	ITSv1	ANN	Seq. Testing Strategy		BN ITS	SARA
				Tier 1	Tier 2		
PC properties	MW: 154.25 Da				√		
	LogP: 3.0		√		√		
	Fraction ionised: 0					√	
	LogD @ pH 7: 3.3					√	
	Volatility[1]: semi-volatile			√	√		
	pH: 8.3			√			
	H_2O solubility @ pH 7: 0.0044					√	
	Plasma protein binding (% bound): 95.5					√	
TIMES-SS	Positive (Strong)			√		√	
TOXTREE	Schiff Base			√			
DEREK Nexus	Positive	√					
DPRA	Negative			√			
	Cys depl: 0%	√	√		√	√	√
	Lys depl: 10%	√	√		√	√	√
KeratinoSens™	Positive			√			
	EC1.5: 110 µM				√	√	√

	EC3: >2000 µM					√	
	Imax : 2,05		√				
	IC50%: 875 µM					√	
U-SENS™	Positive			√			
	CD86 EC150: 53.6 µg/ml				√		√
	CV70: 133.90 µg/ml						
h-CLAT	Positive						
	CD86 EC150: 123 µg/ml	√	√			√	√
	CD54 EC200 - µg/ml	√	√			√	√
	CV75: 139.2 µg/ml					√	
SENS-IS	Moderate				(optional)		

TIER 2

Targeted testing

Based on the existing NAM data availability which allowed a clear hypothesis to be generated in TIER 1 it was decided that there was no need for additional targeted testing or data generation at this stage.

DA risk predictions for geraniol

The skin sensitisation hazard and potency predictions (and for SARA the risk prediction) by the five defined approaches for geraniol can be found in Table 9.

Table 9. Predictions for geraniol by the DAs used

Defined Approach	DA Prediction for geraniol	Comment
ITSv1 DA	GHS Cat. 1B skin sensitiser (ITS score of 3)	The ITS score of 3 was well within the 2-5 GHS Cat. 1B range, and *in chemico*/*in vitro* outcomes are applicable and *in silico* prediction is in domain. Therefore, there was a high confidence in the DA prediction.
ANN	Skin sensitiser (EC3 value: 10.1%)	The predicted EC3 value is 10.1%. All *in chemico*/*in vitro*/*in silico* parameters are applicable. Therefore, there was a high confidence in the DA prediction.
Sequential testing strategy	Tier 1: Skin sensitiser (89.63% probability to be a sensitiser) Tier 2: GHS category 1B (9.69% probability to be a GHS category 1A sensitiser)	As the probabilities in both tiers were well above and below the prediction model cut-offs there was a high confidence in the DA prediction.
BN ITS	Skin sensitiser	Based on the probability distribution, the prediction clearly

	(moderate potency, weak to strong evidence)	indicated a moderate potency. The conversion to BF to assess the strength of the evidence showed borderline weak to strong evidence.
SARA	Human sensitiser potency ED01 = 4600 µg cm^{-2} (95^{th} % CI = 210-99000 µg/cm^2) SARA risk metric (Probability exposure is low risk) = 0.81	This expected ED_{01} is 4600 µg/cm^2. This ranks Geraniol with other weak/moderate skin sensitisers (see Annex II: SARA model predictions). The probability that exposure to 0.1% geraniol in a face cream is low risk is 0.81 (i.e., 81% probability low risk). This ranks with exposures considered low risk for induction of skin allergy (e.g.: use of Kathon CG in rinse off products / benzyl alcohol and phenoxyethanol in leave on products), (see Annex II: SARA model predictions).

Use of DAs and/or WoE for Point of Departure (POD) determination

The predictions from the DAs were converted to points of departure, based on each unique DA approach as shown in Table 10. In brief, the ITS and sequential testing strategy predict a GHS category 1B prediction, corresponding to a default LLNA EC3 value of >2% (ECHA, 2017). The ANN predicts an EC3 value of 10.1%. For the BN ITS the predicted LLNA moderate potency category corresponds to an EC3% value of 1-10%, and then taking the conservative value of 1% to adjust for a low BF. A number of proposals have been published on how to convert LLNA EC3 values into sensitisation potency categories or POD values for risk assessment (summarized in Griem et al., 2003). For this case study we applied a unified approach, by converting the DA predicted LLNA threshold values (EC3%) into a POD [µg/cm^2] as dose per unit area by using a factor of 250 (Robinson et al. 2000), based on the standard LLNA protocol where 25 µL test solution are distributed over a surface of 1 cm^2 per mouse ear (Griem et al. 2003). The SARA DA predicts a human potency (ED_{01} value), which can be directly used as a POD.

Risk Assessment

Compare reference dose to consumer exposure

The information obtained from the individual DAs, derived (where required) POD values, and subsequent risk assessment have been summarised in Table 10. The margins of exposure (MOE) were determined by dividing the POD values by the calculated consumer exposure level of 2.73 µg/cm^2.

The POD derivation is different for each of the 5 DAs, each of which ensures sufficient conservatism when setting the POD. Concluding on the risk for the final decision also varies between DAs, depending on the conservatism already applied in POD derivation, acceptable MOE values can vary. While the SARA DA directly integrates a risk prediction, in this case study, a generic MOE of 100 was chosen in line with e.g., MOEs accepted for systemic toxicity assessments in SCCS dossiers (SCCS NOG, 2021). Only the BN-ITS had a borderline MOE of 91.6.

Table 10. Summary of the output of the five DAs used, POD determination and risk assessment conclusions for the use of geraniol at 0.1% in a face cream.

***According to (Griem at al., 2003). NA: Not applicable**

DA	ITSv1 DA	ANN	Sequential Testing Strategy	BN-ITS	SARA
DA potency	GHS Cat. 1B	Predicted LLNA EC3:	GHS Cat. 1B	Moderate	ED_{01} = 4600 (95^{th}% 210-99000)

prediction		10.1%.			
Method to derive POD	GHS Cat. 1B corresponds to default LLNA EC3 (>2%), which was converted into POD [µg/cm²] by factor 250*.	Predicted EC3 (10.1%) converted into POD [µg/cm²] by factor 250*.	GHS Cat. 1B corresponds to default LLNA EC3 (>2%), which was converted into POD [µg/cm²] by factor 250*.	Default POD based on moderate potency category. (EC3 1-10% converted to 250-2500 µg/cm²*) Lower end picked due to lower BF.	ED_{01} is the POD
Point of Departure (µg/cm²)	>500	2525	>500	250	4600 (95^{th}% 210-99000)
Consumer exposure level (µg/cm²)	2.73	2.73	2.73	2.73	2.73
Margin of exposure (MOE)	>183	924.9	>183	91.6	1700 (95^{th}% 78-36000)
Risk assessment	MOE is considered as sufficiently high.	MOE is considered as sufficiently high.	MOE is considered as sufficiently high.	MOE is considered borderline sufficient.	Probability exposure is low risk = 0.81. This ranks with exposures considered low risk for induction of skin allergy (e.g., use of KathonCG in rinse off products / benzyl alcohol and phenoxyethanol in leave on products).
Risk assessment conclusion	Safe	Safe	Safe	Borderline safe	Safe

7 Uncertainty

The underlying principle for this skin sensitisation IATA is based in the skin sensitisation AOP (OECD, 2014). The mechanistic understanding of skin sensitisation and description of the AOP has enabled the development and regulatory acceptance of a multitude of NAMs, DAs and the NGRA. With that confidence in this IATA approach is high.

The uncertainty of this IATA case study has been qualitatively assessed by reviewing and discussing the uncertainty in each element of the IATA (Figure 2). The Table 11 below describes the uncertainty along each step of the NGRA framework. Where uncertainty needed to be specified for the individual DAs applied, this is clearly indicated in the table below.

Table 11. Uncertainty in the elements of the NGRA case study for geraniol.

NA: not applicable.

NGRA element	Uncertainty (low, medium, high)	Comment
Tier 0		
Consumer exposure	Low	Deterministic worst-case consumer exposure assessment according to the SCCS NoG, derived from well conducted European consumer exposure studies published in peer reviewed journals (Hall et al. 2007, 2011). Exposure from single product only, aggregate exposure was not considered.
Case study chemical	Low	Well characterized chemical, 100% pure
NAM data used in the IATA	Low	Data quality is considered high since the data were generated based on: Validated or scientifically sound methods Within the application domain of *in silico/ in vitro/ in chemico* assays No technical issues conducting assays with geraniol. Data reliability: NAM predictions are reliable, meaning positive and negative data are as expected in the different NAMs. The negative and positive data can be explained based on the structure, physicochemical properties and metabolic capacity of geraniol.
Analogue data/ Read across	NA	Not used, no impact on uncertainty
Tier 1		
Hypothesis	Low	High confidence in the hypothesis that geraniol is a potential sensitiser, due to reliability of NAM information. DPRA result can be explained based on pre-hapten consideration.

Tier 2		
Individual DA outputs	Low to medium	ITSv1: The ITSv1 total battery score was 3, *and in chemico/in vitro* outcomes were applicable and *in silico* prediction was in domain. ANN:EC3 value was predicted and *in chemico/in vitro* parameters and *in silico* prediction were applicable and in domain. Seq. testing strategy: Probabilities in both tiers were well above and below the prediction model cut-off BN-ITS: Probability distribution indicated moderate potency (highest probability). Conversion of to BFs to assess evidence, indicated borderline weak to strong evidence for moderate. SARA: Uncertainty in ED01 provided in SARA predictions. (ED01 expressed as expected and 95th percentile confidence interval)
Extrapolating POD from DA	Low	ITSv1: High confidence of the ITSv1 DA prediction and overall reliability of NAM data. High confidence as POD is derived using the most conservative EC3 value. ANN: High confidence as POD is calculated from predicted EC3. Seq. testing strategy: High confidence in probability value and overall reliability between NAMs data. High confidence as POD is derived using the most conservative EC3 value. BN-ITS: High confidence as POD is derived using the most conservative EC3 value for a moderate potency category. NAM data support decision. SARA: The ED01, the dermal dose that is predicted to sensitise 1% of the HRIPT population, is the POD.
For each DA-based risk prediction	Low	ITSv1: Conservative approach with the selection of the lowest EC3 for the UN-GHS Cat. 1B. ANN: Practical approach with the predicted EC3 value. Seq. testing strategy: Conservative approach with the selection of the lowest EC3 value for the UN-GHS category 1B. BN-ITS: Conservative approach, as conservative POD was chosen. Borderline outcome. SARA: Empirical support for the MOE is obtained by regression against the MOE for established high/low risk benchmark exposures. Uncertainty in potency and risk predations provided as model output
Overall for IATA	Low	All DAs come to a risk assessment conclusion with a low uncertainty. Using the BN-ITS provides a borderline safe result. Refinements could be considered.

8 Strategy and integrated risk assessment conclusion

This illustrative case study was performed to demonstrate the applicability of an NGRA framework to access the use of geraniol at 0.1% in a face cream. The case study followed the tiers and steps outlined in the framework (Figure 2). After the determination of the consumer exposure level, existing information was collected and considered in a WOE approach. Based on this information geraniol was hypothesized to be a skin sensitiser. The NAM data was integrated in five DAs, which predicted geraniol to be a weak, moderate or UN-GHS Cat. 1B sensitiser. The DA outputs were converted into PODs and a MOE was calculated (ratio of consumer exposure to POD) (Table 10).

Risk assessments based upon predictions from four of the applied DAs resulted in a conclusion that use of 0.1% geraniol in a face cream would be safe. The risk assessment based upon the BN-ITS predictions resulted in a borderline conclusion. For such risk assessments, which conclude borderline risk, next steps in a tiered and iterative approach could be to: 1. Reduce the use level and thus consumer exposure, 2. If appropriate generate additional data (not already incorporated into DA) for use as further evidence in a weight of evidence approach to decision making. To note, whilst read across was excluded from this case study, in principle data from read across candidates can be used to refine / inform the derivation of the POD. Overall, as exemplified by this case study, information from NAMs can be applied within a WOE IATA for skin sensitisation to reach a conclusion on consumer risk.

It is of interest to note that geraniol is a fragrance material used in many consumer products (SCCS, 2012). Based upon historical in vivo data it is considered to be a weak skin sensitiser. Use levels have been restricted by IFRA on this basis and with the introduction of the latest IFRA standard (IFRA, 2020) the maximum safe use level for geraniol in a face cream (category 5) is 1.2%.

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. 2006;3(4):189-97.

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. Feb;107(2):451-60. doi: 10.1093/toxsci/kfn243.

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 sensitisation testing. Toxicol in vitro, 30: 373-82.

Alépée N, Piroird C, Nardelli L (2017) U-SENS™: A U937 cell line activation test for skin sensitization., Alternatives for Dermal Toxicity Testing. DOI: 10.1007/978-3-319-50353-0_22.

Alinaghi F, Bennike NH, Egeberg A, Thyssen JP, Johansen JD (2019) Prevalence of contact allergy in the general population: A systematic review and meta-analysis. Contact Dermatitis 80(2):77-85 doi:10.1111/cod.13119.

Api AM, Basketter DA, Cadby PA, et al. (2008) Dermal sensitisation quantitative risk assessment (QRA) for fragrance ingredients. Regul Toxicol Pharmacol 52(1):3-23 doi:10.1016/j.yrtph.2007.10.008.

Aptula AO, Roberts DW (2006) Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. Chem Res Toxicol. Aug;19(8):1097-105. doi: 10.1021/tx0601004.

Ashikaga T, Sakaguchi H, Sono S, et al. (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(4):275-84 doi:10.1177/026119291003800403.

Assaf Vandecasteele H, Gautier F, Tourneix F, van Vliet E, Bury D, Alépée N (2021) Next generation risk assessment for skin sensitisation: A case study with propyl paraben. Regul Toxicol Pharmacol 123:104936. doi: 10.1016/j.yrtph.2021.104936.

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

Casati S, Griesinger C, Whelan M. (2013) EURL ECVAM Recommendation on the Direct Peptide Reactivity Assay (DPRA) for Skin Sensitisation Testing. EUR 26383. Luxembourg (Luxembourg): Publications Office of the European Union;. JRC85936.

Casati S, Whelan M (2015) EURL ECVAM Recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitisation testing. EUR 27022. Luxembourg (Luxembourg): Publications Office of the European Union; JRC93307.

Casati, S., Zuang, V. and Whelan, M (2017) EURL ECVAM Recommendation on the use of non-animal approaches for skin sensitisation testing, EUR 28553 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-79-67786-1 (online),978-92-79-75244-5 (ePub), doi:10.2760/588955 (online),10.2760/08443 (ePub), JRC106410.

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(4):787-802 doi:10.1016/j.tiv.2015.02.012.

Cottrez F, Boitel E, Ourlin JC, et al. (2016) SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitisation potency: Reproducibility and predictivity results from an inter-laboratory study. Toxicol In vitro 32:248-60 doi:10.1016/j.tiv.2016.01.007.

Dean JH, Twerdok LE, Tice RR, Sailstad DM, Hattan DG, Stokes WS (2001) ICCVAM evaluation of the murine local lymph node assay. II. Conclusions and recommendations of an independent scientific peer review panel. Reg Toxicol Pharmacol 34:258-273.

Dearden, J. and Worth, A (2007) In silico Prediction of Physicochemical Properties, chapter 7. Prediction of boiling point. JRC Scientific and Technical Reports EUR 23051 EN – 2007.

Dearden J.C., Netzeva T.I. and Bibby R (2003) A comparison of commercially available software for the prediction of partition coefficient. In Ford M., Livingstone D., Dearden J. and van de Waterbeemd H. (Eds.), Designing Drugs and Crop Protectants: Processes, Problems and Solutions, Blackwell, Oxford, 2003, pp. 168-169.

Del Bufalo A, Pauloin T, Alepee N, Clouzeau J, Detroyer A, Eilstein J, Gomes C, Nocairi H, Piroird C, Rousset F, Tourneix F, Basketter D, and Martinozzi Teissier S (2018) Alternative Integrated Testing for Skin Sensitization: Assuring Consumer Safety. Applied In Vitro Toxicology, Vol. 4, No. 1. doi.org/10.1089/aivt.2017.0023.

Dent M, Amaral RT, Da Silva PA, et al. (2018) Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. Computational Toxicology 7:20-26 doi:https://doi.org/10.1016/j.comtox.2018.06.001.

Desprez B, Dent M, Keller D, et al. (2018) A strategy for systemic toxicity assessment based on non-animal approaches: The Cosmetics Europe Long Range Science Strategy programme. Toxicol In vitro 50:137-146 doi:10.1016/j.tiv.2018.02.017.

Dumont C, Barroso J, Matys I, Worth A, Casati S (2016) Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. Toxicol In vitro 34:220-228 doi:10.1016/j.tiv.2016.04.008.

ECHA (2008) Guidance on information requirements and chemical safety assessment, Chapter R.6: QSARs and grouping of chemicals. http://echaeuropaeu/documents/10162/13632/information_requirements_r6_enpdf

ECHA (2017) CLH report, Proposal for Harmonised Classification and Labelling, Based on Regulation (EC) No 1272/2008 (CLP Regulation), Annex VI, Part 2, International Chemical Identification: Geraniol; (2E)- 3,7-dimethylocta-2,6-dien-1-ol. https://echa.europa.eu/documents/10162/13626/clh_rep_geraniol_en.pdf/784d6935-cbc4-ea48-1df7-90423713b953.

Eilstein, J., Lereaux, G., Budimir, N., Hussler, G., Wilkinson, S., Duche, D (2014) Comparison of xenobiotic metabolizing enzyme activities in ex vivo human skin and reconstructed human skin models from SkinEthic. Arch Toxicol. 88(9):1681-1694.

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. Jun 15;245(3):281-90. doi: 10.1016/j.taap.2010.03.009.

European Commission, Joint Research Centre, EURL ECVAM (2014) Recommendation on the KeratinoSensTM assay for skin sensitisation testing, ISSN 1831-9424 (online), Doi 10.2788/52914.

Ezendam J, Braakhuis HM, Vandebriel RJ (2016) State of the art in non-animal approaches for skin sensitisation testing: from individual test methods towards testing strategies. Arch Toxicol 90(12):2861-2883 doi:10.1007/s00204-016-1842-4.

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 sensitisation in vitro. Arch Toxicol. 2013 Sep;87(9):1683-96.

Gautier, F., Tourneix, F., Assaf Vandecasteele, H., van Vliet, E., Bury, D., & Alépée, N (2020) Read-across can increase confidence in the Next Generation Risk Assessment for skin sensitisation: A case study with resorcinol. Regulatory toxicology and pharmacology, 117, 104755. https://doi.org/10.1016/j.yrtph.2020.104755.

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. 2004 Oct;81(2):332-43. doi: 10.1093/toxsci/kfh213. Epub 2004 Jul 14. PMID: 15254333.

Gerberick GF, Troutman JA, Foertsch LM, et al. (2009) Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. Toxicol Sci 112(1):164-74 doi:10.1093/toxsci/kfp192.

Gilmour N & Kern PS, Alépée N, Boislève F, Bury D, et al. (2020) Development of a next generation risk assessment framework for the evaluation of skin sensitisation of cosmetic ingredients. Regulatory Toxicology and Pharmacology: 104721.

Goebel C, Aeby P, Ade N, et al. (2012) Guiding principles for the implementation of non-animal safety assessment approaches for cosmetics: skin sensitisation. Regul Toxicol Pharmacol 63(1):40-52 doi:10.1016/j.yrtph.2012.02.007.

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), pp. 52-58.

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, Cyprus. 2012: 271-282.

Griem P, Goebel C, Scheffler H (2003) Proposal for a risk assessment methodology for skin sensitization based on sensitization potency data. Regul Toxicol Pharmacol. Dec;38(3):269-90. doi: 10.1016/j.yrtph.2003.07.001.

Hagvall L, Baron JM, Börje A, Weidolf L, Merk H, Karlberg AT (2008) Cytochrome P450-mediated activation of the fragrance compound geraniol forms potent contact allergens. Toxicol Appl Pharmacol. 233(2):308-313. doi:10.1016/j.taap.2008.08.014.

Hagvall L, Bäcktorp C, Svensson S, Nyman G, Börje A, Karlberg AT (2007) Fragrance compound geraniol forms contact allergens on air exposure. Identification and quantification of oxidation products and effect on skin sensitisation. Chem Res Toxicol. 20(5):807-814. doi:10.1021/tx700017v.

Hall B, Tozer S, Safford B, Coroama M, Steiling W, Leneveu-Duchemin MC, McNamara C, Gibney M (2007) European consumer exposure to cosmetic products, a framework for conducting population exposure assessments. Food Chem Toxicol. Nov;45(11):2097-108. doi: 10.1016/j.fct.2007.06.017.

Hall B, Steiling W, Safford B, Coroama M, Tozer S, Firmani C, McNamara C, Gibney M (2011) European consumer exposure to cosmetic products, a framework for conducting population exposure assessments Part 2. Food Chem Toxicol. Feb;49(2):408-22. doi: 10.1016/j.fct.2010.11.016.

Hewitt NJ, Grégoire S, Cubberley R, Duplan H, Eilstein J, Ellison C, Lester C, Fabian E, Fernandez J, Géniès C, Jacques-Jamin C, Klaric M, Rothe H, Sorrell I, Lange D, Schepky A (2020) Measurement of the penetration of 56 cosmetic relevant chemicals into and through human skin using a standardized protocol. J Appl Toxicol. Mar;40(3):403-415. doi: 10.1002/jat.3913.

Hirota M, Ashikaga T, Kouzuki H (2018) Development of an artificial neural network model for risk assessment of skin sensitisation using human cell line activation test, direct peptide reactivity assay, KeratinoSens and in silico structure alert parameter. J Appl Toxicol 38(4):514-526 doi:10.1002/jat.3558.

Hagvall L, Baron JM, Börje A, Weidolf L, Merk H, Karlberg AT (2008) Cytochrome P450-mediated activation of the fragrance compound geraniol forms potent contact allergens. Toxicol Appl Pharmacol. Dec 1;233(2):308-13. doi: 10.1016/j.taap.2008.08.014.

Hagvall L, Bäcktorp C, Svensson S, Nyman G, Börje A, Karlberg AT (2007) Fragrance compound geraniol forms contact allergens on air exposure. Identification and quantification of oxidation products and effect on skin sensitization. Chem Res Toxicol. May;20(5):807-14. doi: 10.1021/tx700017v.

Hoffmann S (2015) LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitisation test methods and strategies. ALTEX 32(4):379-83 doi:10.14573/altex.1505051.

Hoffmann S, Kleinstreuer N, Alepee N, et al. (2018) Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database(). Crit Rev Toxicol 48(5):344-358 doi:10.1080/10408444.2018.1429385.

IFRA, (2020), Amendment 49 for Geraniol. https://ifrafragrance.org/standards/IFRA_STD_037.pdf

ICCVAM, (1999) “Interagency Coordinating Committee on the Validation of Alternative Methods. 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 ICCVAM an the NICEATM,” National Institute of Environmental Health Sciences, NIH Publication no. 99-4494, http://www.iccvam.niehs.nih.gov/.

Jaworska J, Gabbert S, Aldenberg T (2010) Towards optimization of chemical testing under REACH: a Bayesian network approach to Integrated Testing Strategies. Regul Toxicol Pharmacol. Jul-Aug;57(2-3):157-67. doi: 10.1016/j.yrtph.2010.02.003.

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. Nov;33(11):1353-64. doi: 10.1002/jat.2869.

Jaworska JS, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M (2015) Bayesian integrated testing strategy (ITS) for skin sensitisation potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. Arch Toxicol 89(12):2355-83

doi:10.1007/s00204-015-1634-2.
Jeffreys, H (1961) The Theory of Probability (3rd ed.). Oxford, England. p. 432. ISBN 9780191589676.
Jowsey IR, Basketter DA, Westmoreland C, Kimber I (2006) A future approach to measuring relative skin sensitising potency: a proposal. J Appl Toxicol 26(4):341-50 doi:10.1002/jat.1146.
Kern P.S., Deconinck, E., Dameron G, Ryan, C.A., Rathman, J.F., Ribero, V., Mostrag, A., Madriarz T., Hobocienski B., Yang, C (2021) Bayesian Net ITS (BN-ITS4): Update of the Defined Approach for Skin Sensitization. (to be submitted).
Kimber, I., Basketter, D. A., Gerberick, G. F., Ryan, C. A., & Dearman, R. J. (2011). Chemical allergy: translating biology into hazard characterization. Toxicological sciences : an official journal of the Society of Toxicology, 120 Suppl 1, S238–S268. https://doi.org/10.1093/toxsci/kfq346.
Kleinstreuer NC, Hoffmann S, Alepee N, et al. (2018) Non-animal methods to predict skin sensitisation (II): an assessment of defined approaches (*). Crit Rev Toxicol 48(5):359-374 doi:10.1080/10408444.2018.1429386.
Langton K, Patlewicz GY, Long A, Marchant CA, Basketter DA (2006) Structure-activity relationships for skin sensitisation: recent improvements to Derek for Windows. Contact Dermatitis. 55(6):342-347. doi:10.1111/j.1600-0536.2006.00969.x.
Lucas PJ, van der Gaag LC, Abu-Hanna A (2004) Bayesian networks in biomedicine and health-care. Artif Intell Med. Mar;30(3):201-14. doi: 10.1016/j.artmed.2003.11.001.
Luu-The, V., Ferraris, C., Duche, D., Bélanger, P., Leclaire, J., Labrie, F (2007) Steroid metabolism and profile of steroidogenic gene expression in Episkin: high similarity with human epidermis. J Steroid Biochem Mol Biol. 107(1-2):30-6.
Martin SF (2015) Immunological mechanisms in allergic contact dermatitis. Curr Opin Allergy Clin Immunol 15(2):124-30 doi:10.1097/ACI.0000000000000142.
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. Jun;133(2):259-74. doi: 10.1093/toxsci/kft075.
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, pp. 110-119.
Natsch A, Emter R, Ellis G (2009) Filling the concept with data: integrating data from different in vitro and in silico assays on skin sensitizers to explore the battery approach for animal-free skin sensitization testing. Toxicol Sci. Jan;107(1):106-21. doi: 10.1093/toxsci/kfn204.
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. Nov;33(11):1337-52. doi: 10.1002/jat.2868.
Noçairi H, Gomes C, Thomas M, Saporta G (2016) Improving Stacking Methodology for Combining Classifiers; Applications to Cosmetic Industry. Electronic Journal of Applied Statistical Analysis, Vol 9, No 2. DOI Code: 10.1285/i20705948v9n2p340.
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. Dec;65(6):343-53. doi: 10.1111/j.1600-0536.2011.01952.x.
OECD (2004a) Environment Health and Safety Publications Series on Testing and Assessment No. 49. REPORT FROM THE EXPERT GROUP ON (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS [(Q)SARs] ON THE PRINCIPLES FOR THE VALIDATION OF (Q)SARs. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2004)24.
OECD (2004b) Guidelines for the Testing of Chemicals, Section 1 Physical-Chemical properties, Test No. 117: Partition Coefficient (n-octanol/water), High Performance Liquid Chromatography (HPLC) Method – Annex Pow Calculation Methods; OECD publishings ISSN: 2074-5753.
OECD (2010) Test No. 429: Skin Sensitisation: Local Lymph Node Assay, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264071100-en.
OECD (2013) 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 (2014) The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to

Proteins, https://www.oecd-ilibrary.org/content/publication/9789264221444-en.
OECD (2016) Guidance Document for the use of Adverse Outcome Pathways in developing Integrated Approaches to Testing and Assessment (IATA). doi:http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)67&doclanguage=en.
OECD (2017) 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, https://www.oecd-ilibrary.org/content/publication/9789264279285-en.
OECD (2018a) Test No. 442D: In vitro Skin Sensitisation, https://www.oecd-ilibrary.org/content/publication/9789264229822-en.
OECD (2018b) Test No. 442E: In vitro Skin Sensitisation, https://www.oecd-ilibrary.org/content/publication/9789264264359-en.
OECD (2021a) Test No. 442C: In Chemico Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key event on covalent binding to proteins, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264229709-en.
OECD (2021b), Guideline No. 497: Defined Approaches on Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/b92879a4-en.
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.Jul;48(2):225-39. doi: 10.1016/j.yrtph.2007.03.003.
Patlewicz G, Ball N, Booth ED, Hulzebos E, Zvinavashe E, Hennes C (2013) Use of category approaches, read-across and (Q)SAR: general considerations. Regul Toxicol Pharmacol 67(1):1-12 doi:10.1016/j.yrtph.2013.06.002.
Piroird C, Ovigne JM, Rousset F, et al. (2015) The Myeloid U937 Skin Sensitisation Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitisation. Toxicol In vitro 29(5):901-16 doi:10.1016/j.tiv.2015.03.009.
Politano VT, Api AM (2008) The Research Institute for Fragrance Materials' human repeated insult patch test protocol. Regul Toxicol Pharmacol. Oct;52(1):35-8. doi: 10.1016/j.yrtph.2007.11.004.
Reisinger K, Hoffmann S, Alepee N, et al. (2015) Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. Toxicol In vitro 29(1):259-70 doi:10.1016/j.tiv.2014.10.018.
Reynolds J, MacKay C, Gilmour N, Miguel-Vilumbrales D, Maxwell G (2019) Probabilistic prediction of human skin sensitiser potency for use in next generation risk assessment. Computational Toxicology 9:36-49 doi:https://doi.org/10.1016/j.comtox.2018.10.004.
Reynolds J and Gilmour N*, Baltazar MT, Reynolds G, Windebank S and Maxwell G (2021) Decision making in next generation risk assessment for skin allergy: using historical clinical experience to benchmark risk. Submitted.
Robinson, M. K., Gerberick, G. F., Ryan, C. A., McNamee, P., White, I. R., & Basketter, D. A. (2000). The importance of exposure estimation in the assessment of skin sensitization risk. Contact dermatitis, 42(5), 251–259. https://doi.org/10.1034/j.1600-0536.2000.042005251.x
Safford R.J., Aptula A.O., Gilmour N (2011) Refinement of the Dermal Sensitisation Threshold (DST) approach using a larger dataset and incorporating mechanistic chemistry domains. Regulatory Toxicology and Pharmacology 60, 218–224.
SCCS (2012) Scientific Committee on Consumer Safety SCCS, Opinion on Fragrance allergens in cosmetic products. https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_102.pdf
SCCS (2018) Opinion on Skin Sensitisation Quantitative Risk Assessment for Fragrance Ingredients (QRA2). https://eceuropaeu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_211pdf SCCS/1589/17
SCCS (2021) The SCCS Notes of Guidance for the testing of Cosmetic ingredients and their Safety Evaluation. 10th revison. https://ec.europa.eu/health/sites/default/files/scientific_committees/consumer_safety/docs/sccs_o_250.pdf
SCCS (2012) SCCS OPINION on Fragrance allergens in cosmetic products.

https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_102.pdf
Spicer C.W. et al. (2002) Hazardous poluant handbook, measurements, properties and fate in ambiant air. CRC Press LLC, FL, USA; 2002.
Takenouchi O, Fukui S, Okamoto K, 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 sensitising potential and potency of chemicals. J Appl Toxicol 35(11):1318-32 doi:10.1002/jat.3127.
Tollefsen KE, Scholz S, Cronin MT, Edwards SW, de Knecht J, Crofton K, Garcia-Reyero N, Hartung T, Worth A, Patlewicz G (2014) Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). Regul Toxicol Pharmacol. Dec;70(3):629-40. doi: 10.1016/j.yrtph.2014.09.009.
Tourneix F, Alepee N, Detroyer A, et al. (2019) Assessment of a defined approach based on a stacking prediction model to identify skin sensitisation hazard. Toxicol In vitro 60:134-143 doi:10.1016/j.tiv.2019.05.008.
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. doi: 10.2131/jts.38.599.
US EPA (2021) On-line MPBPVPWinTM user's guide, paragraph 7.3: Vapor Pressure. EPI Suite™ software, version 4.1, https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface.
Valkova I, Vracko M, Basak SC (2004) Modeling of structure-mutagenicity relationships: Counter propagation neural network approach using calculated structural descriptors. Analytica Chimica Acta 509(2):179-186. DOI: 10.1016/j.aca.2003.12.035.
Wolpert, D (1992) Stacked Generalization. Neural Networks, 5, 41-259.
Zirwas MJ (2019) Contact Dermatitis to Cosmetics. Clin Rev Allergy Immunol 56(1):119-128 doi:10.1007/s12016-018-8717-9.

Annex I: Individual information sources used

Direct Peptide Reactivity Assay (DPRA)

Name of the information source	Direct Peptide Reactivity Assay (DPRA) - OECD TG 442C (OECD, 2021a)
Mechanistic basis including AOP coverage	The DPRA measures *in chemico* the binding of test chemicals to model synthetic peptides containing either lysine or cysteine (Gerberick et al. 2004, 2009). 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 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. 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 EURL ECVAM validation study for reliability (Casati et al. 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 (Casati et al. 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. The current version of TG 442C contains a kinetic DPRA. - 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) are reported to be in most cases correctly identified. Strict pro-haptens may be underestimated.
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).
Proprietary aspects	The test method does not have proprietary elements.
Predictive capacity (if applicable)	Results generated in the validation study (Casati et al. 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.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within a Defined Approach (Guideline 497), (OECD, 2021b). 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 an IATA	The DPRA is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the DPRA is part of the integrated decision strategy for skin sensitisation hazard or potency identification. In this case study DPRA data is integrated with other information in skin sensitisation defined approaches for potency prediction to derive a POD for risk assessment.

KeratinoSens™

Name of the information source	KeratinoSens™ OECD TG 442D (OECD, 2018a)
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 (Emter et al. 2010). 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 M.
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 Imax is statistically significantly higher than 1.5-fold as compared to the basal luciferase activity and the EC 1.5 value is below 1000 M 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 KeratinoSens™. - Test chemical stressors other than electrophilic chemicals may activate the Keap1-Nrf2-ARE pathway leading to false positive predictions in the KeratinoSens™.
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-sensitising chemicals within a Defined Approach (Guideline 497), (OECD, 2021b). 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 an IATA	The KeratinoSens™ is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the KeratinoSens™ is part of the integrated decision strategy for skin sensitisation hazard or potency identification. In this case study KeratinoSens™ data is integrated with other information in skin sensitisation defined approaches for potency prediction to derive a POD for risk assessment.

Human Cell Line Activation Test (h-CLAT)

Name of the information source	Human Cell Line Activation Test (h-CLAT) - OECD TG 442E (OECD, 2018b)
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) (Ashikaga et al. 2010). 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-a). 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 EURL ECVAM validation study for reliability (Casati et al. 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 P 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 (Casati et al. 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 (Casati et al. 2015).
Predictive capacity (if applicable)	Results generated in the validation study (Casati et al. 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*.
Proprietary aspects	The test method has intellectual property rights protected by Patent N. 4270702 only in Japan.
	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.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within a Defined Approach (Guideline 497), (OECD, 2021b). 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 an IATA	The h-CLAT is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the h-CLAT is part of the integrated decision strategy for skin sensitisation hazard or potency identification. In this case study DPRA data is integrated with other information in skin sensitisation defined approaches for potency prediction to derive a POD for risk assessment.

U-SENS™

Name of the information source	U-SENS™ , OECD TG 442E (OECD, 2018b)
Mechanistic basis including AOP coverage	Skin sensitisers have been reported to induce the expression of cell membrane markers associated with activation of dendritic cells (DC), typically assessed by expression of specific cell surface markers. Dendritic cell activation upon exposure to sensitisers leads to functional changes. For example, there are clear changes in cytokine secretion (e.g. TNF-a IL-1β and in the expression of some chemokine receptors such as CCR7 and CXCR4. Additionally, during dendritic cell maturation, co-stimulatory and intercellular adhesion molecules such as HLA-DR, HLA-ABC, CD40, CD80, CD83, CD86 and ICAM-1/CD54 can be up-regulated. Most of the *in vitro* test methods measure the activation of the cell surface marker CD86, which has been established as mechanistically relevant and predictive (Ade et al. 2006; Piroird et al. 2015). The U-SENS™ assay quantifies the induction of the CD86 protein marker expression, associated with DC maturation *in vivo*. The assay is performed on the human myeloid U937 cell line, closely related to monocytes and dendritic cells. The assay therefore addresses one of the biological mechanisms covered by key event 3 of the skin sensitisation AOP (OECD, 2014; OECD 2016).
Description	The U-SENS™ method is an *in vitro* assay that quantifies changes of CD86 cell surface marker expression on a human histiocytic lymphoma cell line, U937 cells, following 45±3 hours exposure to at least four concentrations of test chemical selected amongst usable concentrations pre-defined in the DB-ALM protocol N°183 (2017). The CD86 surface marker is one typical marker of U937 activation. CD86 is known to be a co-stimulatory molecule that may mimic monocytic activation, which plays a critical role in T-cell priming. The changes of CD86 cell surface marker expression are measured by flow cytometry following cell staining typically with fluorescein isothiocyanate (FITC)-labelled antibodies (CD86-IgG1 percent of positive cells measurement). Cytotoxicity measurement is also conducted (e.g. by using propidium iodide) concurrently to assess whether upregulation of CD86 cell surface marker expression occurs at sub-ytoctoxic concentrations (cell viability = 70%). Each test chemical is tested in at least four concentrations and in at least two independent runs (performed on a different day) to derive a single prediction (NEGATIVE or POSITIVE). The stimulation index (S.I.) of CD86 cell surface marker compared to solvent/vehicle control is calculated and used in the prediction model, to support the discrimination between sensitisers and non-sensitisers.
Response(s) measured	- Cell viability using Propidium Iodide to calculate the concentration at which a chemical reaches the cytotoxicity threshold of 70% (CV70). - CD86 stimulation index: CD86 relative value of intensity in chemical-treated cells compared to

	solvent/vehicle treated cells to calculate the concentration at which a chemical reaches the CD86 positive threshold of 150%(EC150). The EC150 and CV70 values are calculated - for each run: the individual EC150 and CV70 values are used as tools to investigate the concentration response effect of CD86 increase, - based on the average viabilities, the overall CV70 is determined, - based on the average S.I. of CD86 values, the overall EC150 is determined for the test chemical predicted as POSITIVE with the U-SENS™.
Prediction model	The prediction model is described in the below. For CD86 expression measurement, each test chemical is tested in at least four concentrations and in at least two independent runs to derive a single prediction (NEGATIVE or POSITIVE). - The individual conclusion of an U-SENS™ run is considered Negative (hereinafter referred to as N) if the S.I. of CD86 is less than 150% at all non-cytotoxic concentrations (cell viability = 70%) and if no interference is observed (cytotoxicity, solubility or colour regardless of the non-cytotoxic concentrations at which the interference is detected). In all other cases: S.I. of CD86 higher or equal to 150% and/or interferences observed, the individual conclusion of an U-SENS™ run is considered Positive (hereinafter referred to as P). - An U-SENS™ prediction is considered NEGATIVE if at least two independent runs are negative (N). If the first two runs are both negative (N), the U-SENS™ prediction is considered NEGATIVE and a third run does not need to be conducted. - An U-SENS™ prediction is considered POSITIVE if at least two independent runs are positive (P). If the first two runs are both positive (P), the U-SENS™ prediction is considered POSITIVE. - If, in the first run, the S.I. of CD86 is higher or equal to 150% at the highest non-cytotoxic concentration only, the run is then considered to be concluded NOT CONCLUSIVE (NC), and additional concentrations should be tested in additional runs. In case a run is identified as NC, at least 2 additional runs should be conducted, and a fourth run in case runs 2 and 3 are not concordant (N and/or P independently).
Metabolic competence (if applicable)	Characterized (Fabian et al. 2013): although the activities of some xenobiotic metabolizing enzymes are not detected, U937 have functional NAT-1 and esterases.
Status of development, standardisation, validation	Evaluated in a EURL ECVAM validation study for reliability (Casati et al. 2017) and officially OECD 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 always be reliably assessed. 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: Membrane disrupting substances (like surfactants), Theoretically, substances that may interfere with CD86 induction pathways due to their own biological activity.
Weaknesses and Strengths	Strengths: Transferability, intra- and inter-reproducibility demonstrated Available to CROs Possible automation demonstrated for this method Large dataset tested (175 chemicals) publicly available Provide dose-response information Pre or pro-haptens correctly predicte False negatives in majority amongst weak or rare sensitisers. 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 (flow cytometer).
Reliability (within and between laboratories) (if applicable)	A validation study (combining two multicentric studies conducted in 2013 and 2014) including four laboratories and testing up to 38 chemicals designed to assess reliability was carried out according to internationally agreed OECD principles. The level of reproducibility in predictions

	that can be expected from the test method is in the order of 90% and 84% within and between laboratories, respectively (Alépée et al. 2015).	
Predictive capacity (if applicable)	Results generated in the validation study (Alépée et al. 2015) and other published studies (Piroird et al. 2015) overall indicate that, compared with LLNA results, the accuracy in distinguishing skin sensitisers (i.e. UN GHS Cat.1) from non-sensitisers is 86% (N='166)' with a sensitivity of 91% (118/129) and a specificity of 65% (24/37). Compared with human results, the accuracy in distinguishing skin sensitisers (i.e. UN GHS Cat.1) from non-sensitisers is 77% (N='101)' with a sensitivity of 100% (58/58) and a specificity of 47% (20/43). False negative predictions compared to LLNA with the U-SENS™ are more likely to concern chemicals showing a low to moderate skin sensitisation potency (i.e. UN GHS subcategory 1B) than chemicals showing a high skin sensitisation potency (i.e. UN GHS subcategory 1A) (Alépée et al. 2017). It is important to note that the predictive values given here for the U-SENS™ as a stand-alone test method are only indicative, since the test method should be considered in combination with other sources of information in the context of an IATA.	
Proprietary aspects	The test method does not have proprietary elements.	
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within an IATA. For the purpose of certain regulations, a positive U-SENS™ prediction can be used to classify a chemical into UN GHS category 1. U-SENS™ data can be used within IATA to support potency prediction.	
Potential role within an IATA	The U-SENS™ is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the U-SENS™ is part of the integrated decision strategy for skin sensitisation hazard or potency identification. In this case study U-SENS™ data is integrated with other information in skin sensitisation defined approaches for potency prediction to derive a POD for risk assessment.	

SENS-IS

Name of the information source	SENS-IS®
Mechanistic basis including AOP coverage	SENS-IS® test method addresses the biological mechanisms described under key event 1 and 2 of the skin sensitisation AOP (i.e. skin penetration with a covalent modification of epidermal proteins and activation of epidermal keratinocytes) by the analysis of the expression of a panel of 61 genes relevant to the considered biological processes. The panel of genes are categorized in two groups: The first "ARE" group includes a selection of 17 genes that have an anti-oxidant responsive element in their promotor and monitor the Redox protective signals induced through the interaction of sensitisers binding to the cysteine amino acids of Keap1 (Kelch-like ECH-associated protein 1) -Nrf2 (nuclear factor-erythroid 12-related factor 2) complex. The second SENS-IS group includes a selection of 21 genes involved in inflammation, danger signals and cell migration to address the complex cascade of events leading to activation of DCs by a sensitising chemical. The remaining group of 23 genes are used as indicators of skin irritation.
Description	The test method is performed using a reconstructed human epidermis model (Episkin®) as the test system. 30µl of the test item is applied onto the stratum corneum of the epidermis at 4 different concentrations (50% - 10% - 1% - 0.1%) in the appropriate vehicle. After 15min of treatment, reconstructed skin is washed and post incubated during 6 hours. The samples are harvested and frozen in liquid nitrogen before tissue lysing and cDNA preparation. The expression of 61 genes is measured by RT-PCR. Gene expression data was log2 transformed and normalised against the 3 housekeeping genes (Glucurinidase b, b2 microglobuline and non-POU domain containing, octamer-binding: NONO) Moreover, the level of tissue destruction is evaluated with the cycle threshold (CT) value of the HSPAA1 gene (CT value >110% of the corresponding value obtained for the vehicle control indicates an unacceptable level of tissue damage). The irritation signals are analysed by measuring the expression level of 23 genes. If more than 20 on 23 irritation genes are overexpressed (traducing stress cell) the sample is not accepted.
Response(s) measured	Cytotoxicity and stress cell (HSPAA1 gene, irritation genes group. Gene expression of the 2 groups predictive biomarker genes (ARE and SENS-IS groups)
Prediction model	The prediction model is based on the gene expression of 61 genes.

	A test item is considered as a sensitiser (positive) when at least 7 genes from both the ARE and/or the SENS-IS genes lists are over-expressed (> 1.25 x control vehicle value). The lowest test item concentration fulfilling these criteria is then used for potency classification as follow: A chemical is classified as an extreme, a strong, a moderate or a weak sensitiser if found positive at 0.1%, 1%, 10% or 50% respectively. If negative at all tested concentrations, it is considered as a non-sensitiser.
Metabolic competence (if applicable)	The test method is performed using topical application of test item on the stratum corneum of the reconstructed epidermis Episkin®. The metabolic characterization of this model was performed and metabolic functionality toward xenobiotics was find very similar to normal human skin (mRNA and enzymatic activity level of many enzymes like P450, esterases, ADH, ALDH, peroxydases, GST, NAT, UDPGT, SULT…) (Luu-The et al. 2007 ; Eilstein et al. 2014)
Status of development, standardisation, validation	SENS-IS® test method was evaluated through a ring-study performed within three laboratories, using a test set of 19 blind coded chemicals and showed a good reproducibility. (Cottrez et al. 2016). A validation process is ongoing at ECVAM to confirmed the capacity of this method for distinguish NS/Cat1A/cat1B. To date, no OECD guideline on SENS-IS® has been adopted.
Technical limitations and limitations with regard to predictivity	Technical limitations: No technical limitation claimed. Highly cytotoxic chemicals cannot always be reliably assessed. Limitations with regard to predictivity (Cottrez et al. 2016): The predictive capacity of the SENS-IS® assay for hazard prediction was measured using a test set of 150 chemicals and compare to LLNA class. Some chemicals were wrongly predicted in this assay: two false negative (imidazolinyl urea and isopropyl myristate) and three false positive (dimethylformamise, isopropyl alcohol and benzaldehyde). However, SDS, a false-positive in the LLNA, was correctly classified by the gene signature the SENS-IS® assay. The predictive capacity of the SENS-IS® assay for potency prediction was also measured using a test set of 150 chemicals compare to LLNA class: 2% of test set are under predicted and 5% are over predicted.
Weaknesses and Strengths	Strengths: The topical application of test item on the stratum corneum of the reconstructed epidermis Episkin® allows to assess the skin penetration and chemical reactivity. Moreover, this kind of application gives the possibility to test mixtures and hydrophobic compounds. The predictive capacity of the SENS-IS® assay allows potency prediction. 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 (Cottrez & al, 2015). Weaknesses: Keratinocyte response is only assessed at gene level and doesn't take into account protein expression Activation of dendritic cells included in the KE3 is not covered by this model.
Reliability (within and between laboratories) (if applicable)	Reproducibility within and between laboratories for SENS-IS® test method has been assessed through a ring-study performed within three laboratories, using a test set of 19 blind coded chemicals and showed a good reproducibility. (Cottrez at al., 2016).
Predictive capacity (if applicable)	(Cottrez et al. 2016). The predictive capacity of the SENS-IS® assay for hazard prediction was measured using a test set of 150 chemicals and showed a good performance compare to LLNA prediction (according to Cooper statistics). The accuracy of the SENS-IS® assay in discriminating sensitisers (i.e. UN GHS cat. 1)

	from non-sensitisers is 96.6%, with a sensitivity of 97.7% and a specificity of 95.2%. The Positive Predictive Value (PPV) was 96.6% and the Negative Predictive Value (NPV) 96.7%. Some chemicals were wrongly predicted in this assay: two false negative (imidazolinyl urea and isopropyl myristate) and three flase positive (dimethylformamise, isopropyl alcohol and benzaldehyde). However, SDS, a false-positive in the LLNA, was correctly classified by the gene signature the SENS-IS® assay. The predictive capacity of the SENS-IS® assay for potency prediction was also measured using a test set of 150 chemicals and showed a good performance compare to LLNA class: Concordance = 92.66% with 2% under predicted and 5% over predicted.	
Proprietary aspects	While the identity of irritation-associated genes has been published (Cottrez et al. 2015), the identity of the genes of the "SENS-IS" and "Redox" genes has been disclosed in a patent by ImmunoSearch.	
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within a Defined Approach to measure key event 1 and 2. The assay does provide data that can be used for potency assessment.	
Potential role within an IATA	The SENS-IS™ is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the SENS-IS™ is part of the integrated decision strategy for skin sensitisation potency identification. For example, SENS-IS™ data could be optionally used in the sequential testing strategy Tier 2 in case of borderline predictions close to the prediction model cut-off.	

TIMES-SS

Name of the information source	TIMES-SS
Mechanistic basis including AOP coverage	Chemical reactivity of xenobiotics (and their metabolites) 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, 2004a).
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) Accuracy= 87% Sensitivity= 92% Specificity= 78% comp. not in training set of TIMES (92). 80% 86% 70%.
Proprietary aspects	Need for a License; TIMES-SS may be replaced in the defined approach Proprietary aspects 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 an IATA	TIMES-SS is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the TIMES-SS is part of the integrated decision strategy for skin sensitisation hazard or potency identification. In this case study TIMES-SS data is integrated with other information in skin sensitisation defined approaches for potency prediction to derive a POD for risk assessment.

Toxtree

Name of the information source	Toxtree (from Ideaconsult Ltd)
Mechanistic basis including AOP coverage	The classifications that are attributed by the Toxtree Skin Sensitisation Alerts decision tree are 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. (Aptula and Roberts, 2006)
Description	*In silico* prediction software containing Skin Sensitisation Alerts based on the Reaction Mechanistic Domains classification.
Response(s)	Five mechanistic alerts for reactivity. With “SNAr”, “SN2”, “Acyl transfer agent”, “Michel acceptor” and “Shiff base formation” reactivity alerts the chemical is classified as sensitiser, with “no skin sensitisation alert” it is classified as non-sensitiser.
Prediction model	Toxtree’s Skin Sensitisation Alerts decision tree which relies on a Reaction Mechanistic Domains classification, will output alerts for a parent chemical structure. (Aptula and Roberts, 2006)
Metabolic competence (if applicable)	No
Status of information source development, standardisation, validation	Open source software, no official validation. Toxtree’s Skin Sensitisation Alerts follow OECD *in silico* models’ validation principles. 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. Currently there is no definition of model domain integrated.
Weaknesses and Strengths	Strengths: Mechanism based classification; freely available software; transparency of the algorithms used to generate data; the approach is published in peer-reviewed journals. Limitations: Currently there is no definition of model domain integrated.
Reliability	Not applicable
Predictive capacity (if applicable)	Ex. literature study (Safford et al. 2011) : Based on the Reaction Mechanistic Domains classification (Aptula and Roberts, 2006) and the LLNA skin sensitisation classification (Sensitiser/Non Sensitiser) for 363 chemicals, the reaction mechanistic domain classification was said to have a sensitivity of 86% and a specificity of 64% as a predictive tool for skin sensitisation (S/NS).
Proprietary aspects	Open source software from Ideaconsult Ltd.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within an IATA. The alerts can contribute to classification of chemicals into mechanistic domains to support read-across.
Potential role within an	Toxtree is foreseen to be combined with complementary information and evaluated in the

IATA	context of IATA. In such context, the Toxtree is part of the integrated decision strategy for skin sensitisation hazard or potency identification. In this case study Toxtree data was integrated with other information in Tier 1 of the sequential testing strategy.

Derek Nexus

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 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 "certain", "probable", "plausible", or "equivocal" 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, 2004a). 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 "certain", "probable", "plausible", or "equivocal" 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 an IATA	Derek Nexus is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the Derek Nexus is part of the integrated decision strategy for skin sensitisation hazard or potency identification. In this case study Derek Nexus data was integrated with other information in the ITSv1 DA to predict skin sensitisation potency and derive a POD for risk assessment.

Volatility through MPBPVP model in EPI Suite™ software

Name of the information source	**Volatility through MPBPVP model in EPI Suite™ software (from US EPA).**

Mechanistic basis including AOP coverage	"Volatility" expressed through the vapor pressure calculated by the MPBPVP model (based on the structure of a given chemical) was identified by the different statistical models we applied for its informative value to predict skin sensitisation hazard in combination to the *in silico*, *in chemico* and *in vitro* information sources used in the Integrated Testing Strategy. Although there is no evident link between this "volatility" parameter and a chemical/biological mechanism related to skin sensitisation, our hypothesis is that it might have an impact on the stability/bioavailability of the substances in defined test conditions (differences between *in vivo* and *in chemico* / *in vitro* test conditions).
Description	EPI Suite™ is a computer platform that contains over 13 different predictive calculation modules and databases for physicochemical properties (amongst which vapor pressure) and environmental fate.
Response(s)	Measured vapor pressure data at 25°C in mmHg, if available. Calculated vapor pressure at 25°C in mmHg. Next, we transform these values 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).
Prediction model	Based on the structure of a given chemical, the EPI Suite™ MPBPVP model estimates vapor pressure from various physicochemical equations (US EPA, 2021). 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.
Metabolic competence (if applicable)	No
Status of information source development, standardisation, validation	Open source software, no official validation. The calculated vapor pressure model MPBPVP from EPI Suite™ follows OECD *in silico* models' validation principles. The approach is published in peer-reviewed journals. In addition, the EPI Suite™ platform has undergone detailed review by a panel of EPA's independent Science Advisory Board.
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. Currently there is no definition of model domain integrated.
Weaknesses and Strengths	Strengths: the calculated vapor pressure model is freely available; the approach is published in peer-reviewed journals. Limitations: currently there is no definition of model domain integrated; estimation error can be introduced by poor boiling point estimates or values.
Reliability	Not applicable
Predictive capacity (if applicable)	Ex. literature study (Dearden et al. 2003 and 2007) : Results from a 100-compound test set showed that for the MPBPVP vapor pressure model the average absolute prediction error on log VP (Pa) was 0.285; 18 compounds had errors between 0.4 and 0.6, 4 compounds had errors between 0.6 and 1.0, and 2 compounds had errors > 1.0 (the worst was cy-anogen with an error of 3.339). It is noted that estimation error can be introduced by using poor boiling point estimates or values as input data.
Proprietary aspects	Open source software from US EPA.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within an IATA. More in particular integrate stability and/or bioavailability characteristics that could potentially impact *in vitro*/*in vivo* correlation due to different testing conditions.
Potential role within an IATA	The calculated vapor pressure is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the calculated vapor pressure is part of the integrated strategy for skin sensitisation hazard identification (Tier 1) based on *in silico*, *in chemico*, and *in vitro* data in the sequential testing strategy DA.

Measured pH

Name of the information source	Measured pH
Mechanistic basis including AOP coverage	The measured pH was identified by the different statistical models we applied for its informative value to predict skin sensitisation hazard in combination to the *in silico*, *in chemico* and *in vitro* information sources used in the Integrated Testing Strategy. Although there is no evident link between this parameter and a chemical/biological mechanism related to skin sensitisation, our hypothesis is that it might have an impact on the

	stability/bioavailability of the substances in defined test conditions (differences between *in vivo* and *in chemico* / *in vitro* test conditions).
Description	Measured quantitative pH value in water obtained with a method adapted from OECD Guideline for the Testing of Chemicals No. 122 (OECD, 2013). The pH-measurement is done at 21±2°C with a specific combined glass electrode developed for Hamilton pH-module instrument. It is calibrated before each measurement with 3 standard buffers (pH4, pH7 and pH9,2). The pH measurement of a sample is repeated 8 times with 8 electrodes and a mean and standard deviation is calculated.
Response(s) measured	Measured pH (quantitative variable): value between 1 and 14
Prediction model	Not applicable
Metabolic competence (if applicable)	No
Status of information source development, standardisation, validation	The pH measuring method is adapted from the OECD Guideline for the Testing of Chemicals No. 122.
Technical limitations and limitations with regard to applicability	Its domain mostly covers small organics, rarely inorganics.
Weaknesses and Strengths	Strengths: applicable to a wide range of chemicals; simple method. Weakness: contrary to the OECD Guideline, with our adapted method, if the pH is lower than 4 or higher than 10, the pH values are not reconfirmed by a titration method with a standardized strong base/acid.
Reliability (within and between laboratories) (if applicable)	A variation on the absolute value of ±0,3 units could be observed between 2 laboratories and 2 measurement instruments calibrated with the same buffers.
Predictive capacity (if applicable)	Not applicable
Proprietary aspects	L'Oréal internal method adapted from OECD Guideline for the Testing of Chemicals No. 122.
Proposed regulatory use	To support the discrimination between sensitising and non-sensitising chemicals within an IATA. More in particular integrate stability and/or bioavailability characteristics that could potentially impact *in vitro*/*in vivo* correlation due to different testing conditions.
Potential role within an IATA	The pH is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the pH is part of the integrated strategy for skin sensitisation hazard identification (Tier 1) based on *in silico*, *in chemico*, and *in vitro* data in the sequential testing strategy DA.

Molecular weight

Name of the information source	Molecular weight
Mechanistic basis including AOP coverage	The calculated molecular weight (MW) is applied for its informative value to discriminate UN GHS cat 1A from UN GHS cat. 1B in combination to the *in silico*, *in chemico* and *in vitro* information sources used in the TIER 2 of our DIP. Although there is no evident link between this parameter and a chemical/biological mechanism related to skin sensitisation, our hypothesis is that it might have an impact on the bioavailability of the substances and thus potentially impact their sensitising potency.
Description	The MW is calculated using Biovia software.
Response(s)	Molecular weight of the cleaned* structure of a given chemical (quantitative variable). * structure without any counter ion and with neutralized acid/basic functions
Prediction model	Using the cleaned* structure of a given chemical, the MW model estimates its molecular weight based on the constituting atom's measured weight contributions. * structure without any counter ion and with neutralized acid/basic functions
Metabolic competence (if applicable)	Not applicable
Status of information source development, standardisation, validation	The MW calculation is commercially available through software from Biovia.
Technical limitations and limitations with regard to applicability	The method can only be applied to chemicals with a defined structure (no mixtures, no polymers).

Weaknesses and Strengths	Weaknesses: no calculations on structurally unidentified substances and mixtures possible. Strengths: high predictive capacity; 100% reproducibility; fast; no high expertise needed; can be used on any computer; applicable to all structurally defined substances.
Reliability	Not applicable
Predictive capacity (if applicable)	The precision of the molecular weight is determined by the precision of the least precise atomic mass value.
Proprietary aspects	For molecular weight calculations commercially available software from Biovia can be used, but multiple freeware molecular weight calculators are also available.
Proposed regulatory use	To support the discrimination between UN GHS cat. 1A from UN GHS cat. 1B sensitisers within Tier 2. More in particular integrate bioavailability characteristics in it that could potentially impact sensitising potency.
Potential role within an IATA	The MW is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the MW is part of the integrated strategy for skin sensitisation potency prediction (Tier 2) based on in silico, in chemico, and *in vitro* data in the sequential testing strategy DA.

ClogP

Name of the information source	ClogP
Mechanistic basis including AOP coverage	The octanol-water partition coefficient calculated by the ClogP model is applied for its informative value to discriminate UN GHS cat 1A from UN GHS cat. 1B in combination to the *in silico*, *in chemico* and *in vitro* information sources used in the Tier 2 of our DIP. Although there is no evident link between this parameter and a chemical/biological mechanism related to skin sensitisation, our hypothesis is that it might have an impact on the bioavailability of the substances and thus potentially impact sensitising potency.
Description	The ClogP software calculates the octanol-water partition coefficient (Pow).
Response(s)	Logarithm of calculated octanol-water partition coefficient (quantitative variable).
Prediction model	Using the structure of a given chemical, the ClogP model estimates its octanol-water partition coefficient based on the theoretical fragmentation of the structure into suitable substructures for which reliable log Pow increments are known. The calculated log Pow is obtained by summing the fragment values and the correction terms for intramolecular interactions. These fragment values are originally derived from measured octanol-water partition coefficient data.
Metabolic competence (if applicable)	Not applicable
Status of information source development, standardisation, validation	The ClogP software is commercially available from the BioByte Corp. The used approach is described and referenced in the OECD Guideline for the Testing of Chemicals No. 117 (OECD, 2004b).
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. In general, the applicability of the calculation method decreases as the complexity of the compound under study increases. There is an error number with description given as an output to help evaluate the applicability of the model to the given structure.
Weaknesses and Strengths	Weaknesses: no calculations on structurally unidentified substances and mixtures possible. Strengths: high predictive capacity; 100% reproducibility; fast; no high expertise needed; can be used on any computer; error number with description given to evaluate the applicability to a given structure.
Reliability	Not applicable
Predictive capacity (if applicable)	A comparison by Dearden et al. (2003) found that for ClogP, using a 138-chemical test set, the percentage of chemicals with a calculated log Pow predicted within ± 0.5 log units of the measured log Pow value was 88,4 % with a standard deviation of 0,29.
Proprietary aspects	A license agreement is needed for ClogP, commercially available software from BioByte Corp.
Proposed regulatory use	To support the discrimination between UN GHS cat. 1A from UN GHS cat. 1B within the Tier 2 of our DIP. More in particular integrate bioavailability characteristics in it that could potentially impact sensitising potency.
Potential role within an IATA	The calculated octanol-water partition coefficient is foreseen to be combined with complementary information and evaluated in the context of IATA. In such context, the calculated octanol-water partition coefficient is part of the integrated decision strategy for skin sensitisation hazard or potency identification. In this case study calculated octanol-water partition coefficient data is

	integrated with other information in skin sensitisation defined approaches for potency prediction to derive a POD for risk assessment .

Annex II: SARA model predictions

Figure 7. Probabilistic estimates of the ED01 (50/95% centred credible intervals and median) for chemicals in the SARA model. Ranking of chemicals determined on median.

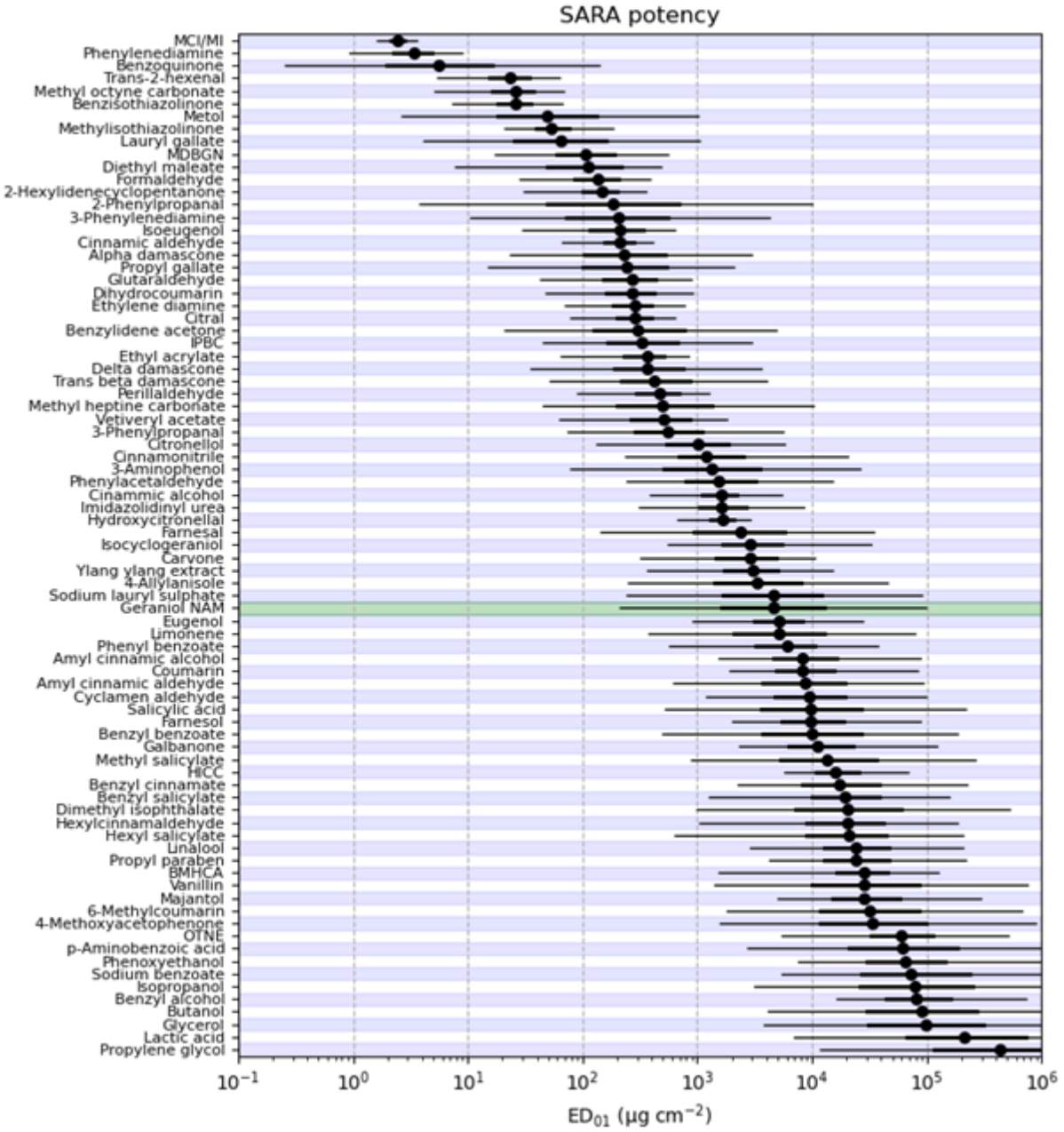
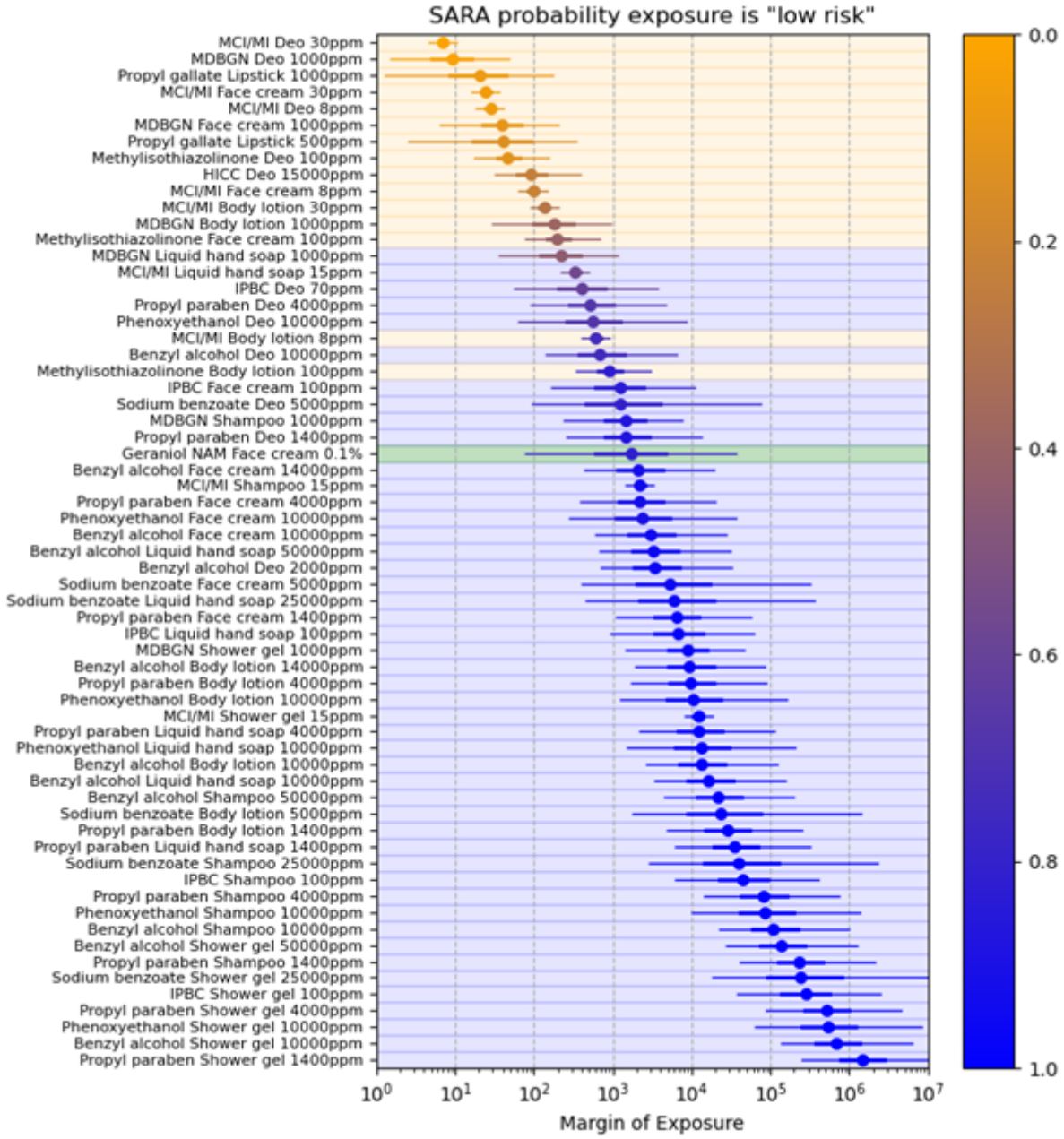


Figure 8. Probabilistic estimates of the margin of exposure corresponding to each benchmark exposure. Background colours indicate assigned risk category (blue – low risk, orange – high risk). Shaded colours indicate the model-inferred risk. Ranking based on the median margin of exposure.



	Expected (50th %ile)	2.5th %ile	97.5th%ile
ED01 (µg/cm2)	4600	210	99000
MOE	1700	78	36000