



Expansion of the Cosmetics Europe skin sensitisation database with new substances and PPRA data

Sebastian Hoffmann^{a,*}, Nathalie Alépée^b, Nicola Gilmour^d, Petra S. Kern^e, Erwin van Vliet^f, Fanny Boislève^g, Dagmar Bury^c, Elodie Clouet^h, Martina Klaric^{i,1}, Jochen Kühnl^j, Jon F. Lalko^k, Karsten Mewes^l, Masaaki Miyazawa^m, Hayato Nishidaⁿ, Mai Thanh Tam Bami^o, Mustafa Varçinⁱ, Anne Marie Api^p, Cosmetics Europeⁱ

^a seh consulting + services, Stembergring 15, 33106, Paderborn, Germany

^b L'Oréal, Research & Innovation, Aulnay-sous-Bois, France

^c L'Oréal, Research & Innovation, Clichy, France

^d Unilever, Colworth Science Park, Bedford, MK44 1LQ, United Kingdom

^e Procter & Gamble Services NV/SA, Temselaan/ Boechoutlaan, 1853, Strombeek-Bever, Belgium

^f Innovito Consulting & Services, Regentenland 35, 3994TZ, Houten, the Netherlands

^g Chanel, 135 Avenue Charles de Gaulle, 92521, Neuilly, France

^h Pierre Fabre, 3 Avenue Hubert Curien, 31100, Toulouse, France

ⁱ Cosmetics Europe, 40 Avenue Hermann-Debroux, 1160, Brussels, Belgium

^j Beiersdorf AG, Unnastraße 48, 20245, Hamburg, Germany

^k Estee Lauder Companies, 155 Pinelawn Road, Suite 300 South, Melville, NY, 11747, USA

^l Henkel, Henkelstraße 67, 40589, Düsseldorf, Germany

^m Kao Corporation, 2606 Akabane, Ichikai, Haga, Tochigi, 321-3497, Japan

ⁿ Shiseido Global Innovation Center, 1-2-11, Takashima, Nishi-ku, Yokohama-shi, Kanagawa, 220-0011, Japan

^o LVMH, 185 Avenue de Verdun, 45804, St Jean de Braye, France

^p Research Institute for Fragrance Materials, Inc, Woodcliff Lake, NJ, USA

ARTICLE INFO

Handling Editor: Dr. Lesa Aylward

Keywords:

Skin sensitisation

LLNA

Human reference data

New approach methodologies

Next generation risk assessment

Database

ABSTRACT

The assessment of skin sensitisation is a key requirement in all regulated sectors, with the European Union's regulation of cosmetic ingredients being most challenging, since it requires quantitative skin sensitisation assessment based on new approach methodologies (NAMs).

To address this challenge, an in-depth and harmonised understanding of NAMs is fundamental to inform the assessment. Therefore, we compiled a database of NAMs, and *in vivo* (human and local lymph node assay) reference data. Here, we expanded this database with 41 substances highly relevant for cosmetic industry. These structurally different substances were tested in six NAMs (Direct Peptide Reactivity Assay, KeratinoSens™, human Cell Line Activation Test, U-SENS™, SENS-IS, Peroxidase Peptide Reactivity Assay).

Our analysis revealed that the substances could be tested without technical limitations, but were generally overpredicted when compared to reference results. Reasons for this reduced predictivity were explored through pairwise NAM comparisons and association of overprediction with hydrophobicity.

We conclude that more detailed understanding of how NAMs apply to a wider range of substances is needed. This would support a flexible and informed choice of NAMs to be optimally applied in the context of a next generation risk assessment framework, ultimately contributing to the characterisation and reduction of uncertainty.

* Corresponding author. seh consulting + services, Stembergring 15, 33106, Paderborn, Germany.

E-mail addresses: sebastian.hoffmann@seh-cs.com (S. Hoffmann), nathalie.alepee@rd.loreal.com (N. Alépée), Nicola.Gilmour@Unilever.com (N. Gilmour), kern.ps@pg.com (P.S. Kern), vanvliete@gmail.com (E. van Vliet), fanny.boisleve@chanel.com (F. Boislève), dagmar.bury@rd.loreal.com (D. Bury), martina.klaric@huntsman.com (M. Klaric), Jochen.Kuehn@Beiersdorf.com (J. Kühnl), jlalko@estee.com (J.F. Lalko), Karsten.Mewes@henkel.com (K. Mewes), miyazawa.masaaki@kao.com (M. Miyazawa), hayato.nishida@shiseido.com (H. Nishida), mttbrami@research.lvmh-pc.com (M.T. Tam Bami), mustafa.varcin@gmail.com (M. Varçin), amapi@rifm.org (A.M. Api), cosmeticseurope@cosmeticseurope.eu (C. Europe).

¹ Current address: Huntsman Europe, Everslaan 45, 3078 Everberg, Belgium.

Abbreviations

AOP	Adverse Outcome Pathway
BP	Boiling Point
CAS	Chemical Abstract Service
CE	Cosmetics Europe
DA	Defined Approach
DPRA	Direct Peptide Reactivity Assay
EC	European Commission
EC3	Effective Concentration (with a three-fold induced stimulation)
GPMT	Guinea Pig Maximisation Test
h-CLAT	human Cell Line Activation Test
IATA	Integrated Approaches to Testing and Assessment
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
KE	Key Event

KOW	Octanol-Water Partition Coefficient
LLNA	(murine) Local Lymph Node Assay
MP	Melting Point
NAM	New Approach Methodologies
NGRA	Next Generation Risk Assessment
OECD	Organisation for Economic Co-operation and Development
PPRA	Peroxidase Peptide Reactivity Assay
PM	Prediction Model
QSAR	Quantitative Structure-Activity Relationship
QSPR	Quantitative Structure–Property Relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
SCCS	Scientific Committee on Consumer Safety
TIMES-SS	Times MEtabolism Simulator platform used for predicting Skin Sensitisation
VP	Vapour Pressure
WS	Water Solubility

1. Introduction

The assessment of skin sensitisation is a key requirement in all regulated sectors, such as industrial chemicals, cosmetics, pesticides, and pharmaceuticals around the globe (Daniel et al., 2018). As for other sectors, there are regional differences in the type of assessment required for cosmetic ingredients, ranging from hazard to potency or from qualitative to quantitative risk assessment. Likewise, differences in the acceptance of data exist, which arise, according to Daniel et al. (2018), from these regional differences. Requirements of the European Union (EU) are likely the most challenging, as it was the first region to act upon public and political pressure by banning the testing of cosmetic ingredients in animals, while still requiring a quantitative assessment for all endpoints, including sensitisation (EC 2009; SCCS 2021).

These developments in Europe have contributed to the fact that the experimental assessment of skin sensitisation is the human health effect that has undergone most changes in this century. The guinea pig-based test methods developed in the 1960s (Buehler 1965; Magnusson and Kligman 1969), which provide hazard information, had been replaced to a large extent by the murine Local Lymph Node Assay (LLNA), which was one of the first test methods that was systematically validated and peer reviewed, leading to its acceptance as an OECD Test Guideline in 2002 (Basketter et al., 2002; Gerberick et al., 2000; OECD, 2010a). One major advantage of the LLNA over the guinea-pig based methods was that it estimated an effective concentration (EC3) as a measure of skin sensitisation potency (Loveless et al., 2010; Ryan et al., 2007; van Loveren et al., 2008). Only some years later the detailed understanding of the biological mechanism involved in skin sensitisation, as described e.g. by Jowsey et al. (2006), as well as ethical concerns about *in vivo* testing in general, led to the development of the first *in vitro* and *in chemico* test methods, referred to here collectively as new approach methodologies (NAMs), for skin sensitisation (Ashikaga et al., 2006; Gerberick et al., 2004; Natsch and Emter 2008; Python et al., 2007). These NAMs modelled specific mechanistic key events (KE), which were later described in an Adverse Outcome Pathway (AOP) (OECD 2014). After successful validation, most of the test methods have also been taken up as OECD Test Guidelines (OECD 2018a; OECD 2018b; OECD 2021a), all of which have been expanded with additional NAMs addressing the same key event (Alepee et al., 2015; Kimura et al., 2015; Ramirez et al., 2016; Fujita et al., 2019a; Natsch et al., 2020).

For all NAMs, large databases have been published by various groups (e.g., Del Bufalo et al., 2018; Hoffmann et al., 2018; Natsch et al., 2021; Urbisch et al., 2015). These databases covered broad ranges of chemical structures and use categories, which provided the basis for the evaluation of the NAMs' applicability domains. Besides substances with

non-cosmetic uses, such as industrial or pharmaceutical, various types of cosmetic ingredients have been included e.g., fragrances and preservatives. However, few more recently developed cosmetic ingredient types, such as actives, surfactants, dyes, and Ultra-Violet (UV) filters were included.

With the goal to harmonise the assessment of the individual NAMs, we compiled a database of substances with well-defined LLNA and human reference data and complete data for the NAMs Direct Peptide Reactivity Assay (DPRA), the KeratinoSens™, the human Cell Line Activation Test (h-CLAT), the U937 cell line activation test (U-SENS™) and the SENS-IS (Hoffmann et al., 2018). This database also provided the basis for a systematic assessment of combinations of NAMs, in particular defined approaches (DA) (Kleinstreuer et al., 2018). Hence, this high-quality database is pivotal for comprehensive evaluations that support broad acceptance of NAMs and DAs, especially for regulatory purposes, as evidenced by the work conducted by the OECD, particularly resulting in the OECD Guideline No. 497 'Defined Approaches on Skin Sensitisation' (OECD, 2021b). In addition, it provides the data and inputs required to explore next generation risk assessment (NGRA) approaches through case studies (Gautier et al., 2020).

To include additional ingredients of high importance to cosmetic industry, we expanded the original Cosmetic's Europe (CE) skin sensitisation database by adding 41 substances. These substances were tested with the same six NAMs, according to the respective OECD test guidelines, or to the published procedures (Cottrez et al., 2016; OECD 2018a; OECD 2018b; OECD 2021a) to assess their performance in terms of applicability and predictivity for this structurally different chemistry. We also expanded the database by adding Peroxidase Peptide Reactivity Assay (PPRA) data for all substances (Ryan et al. (2020)). Here, we describe the expanded database focusing on the substances' uses, their physico-chemical properties, aspects of their applicability domain and the evaluation of the predictivity of the individual NAMs.

We consider the expanded database as a valuable resource to better understand the advantages and limitations of the individual NAMs. Especially when combined with more focused investigations of others, such as Basketter et al. (2019), Mehling et al. (2019) and Bergal et al. (2020), it will further inform the combination of NAMs, e.g. in defined approaches or integrated approaches to testing and assessment (IATA), support uncertainty assessment of the experimental evidence and enable conduct of case studies to build confidence in the application of our next generation skin sensitisation risk assessment (NGRA) of cosmetic ingredients (Gilmour et al., 2020).

2. Materials and methods

2.1. Selection of substances

The following selection criteria were used:

1. Substances had to be of high relevance for cosmetics industries, such as dyes, UV filters, surfactants, and others, which are under-represented in published databases used to assess NAMs (Hoffmann et al., 2018; Urbisch et al., 2015). Due to their specific uses they were expected to be structurally different and to potentially provide useful insight into the applicability domain of the NAMs.
2. LLNA reference data had to be available, specifying at least the test concentrations, the stimulation indices, the vehicle used and the EC3 values. For three UV filters and a dye no LLNA, but guinea pig maximisation test (GPMT) data were available. Due to their importance for the purpose of extending the database to highly relevant substances, we considered them eligible. Based on the reported high concordance of GPMT and LLNA (Haneke et al. 2001), we assumed that the results of both assays in terms of hazard identification would be identical. Note that we will therefore refer to LLNA reference data only.
3. LLNA reference data should cover the entire response spectrum, i.e., from non-sensitiser to extreme sensitisers.
4. *In vivo* and human reference data should cover the entire response spectrum, i.e., from non-sensitiser to extreme sensitisers.
5. The substances had to be clearly defined by name, CAS-registry number, and SMILES code. Consequently, natural extracts or mixtures were not eligible.
6. The substances had to be commercially available.

Based on these criteria, 44 substances were identified that fulfilled the criteria. Of these, two dyes were not available and one fragrance ingredient was available in a very low purity only, so that 41 substances were selected. The independent contract laboratory BioTeSys (Germany) obtained the substances from various suppliers (Sigma-Aldrich, Chemos-group, TCI chemicals, Dragon Chemical Europe, St Cruz Biotechnologies, Syntharo Fine Chemicals, Extrachem, ABCR, Glentham Life sciences and Chimex), and coded and shipped them to the testing laboratories.

2.2. *In vivo* reference data

For most of the 41 substances LLNA data were publicly accessible, i.e. either in opinions of the Scientific Committee on Consumer Safety (SCCS) of the European Commission (EC) or its predecessors, in scientific publications (Ball et al., 2011; De Jong et al., 2007; ICCVAM 2011; Karlsson et al., 2011) or in REACH dossiers disseminated via the website of the European Chemical Agency (www.echa.europa.eu). For eight substances, summary LLNA data according to selection criterion 2 were made available by Cosmetic Europe member companies. For four substances, no LLNA, but GPMT data were found in SCCS opinions or REACH dossiers, which were used instead to determine *in vivo* reference hazard results.

In addition to LLNA data, human data, i.e., human test data, diagnostic patch test data and other relevant published information, were collected for the substances to derive a human potency category, where possible. The retrieved evidence was reviewed and was sufficiently conclusive for 19 substances to assign them to one of six human skin sensitisation potency categories according to the approach described by Basketter et al. (2014) and Api et al. (2017). The data used to derive the potency categories are provided as Supplemental Table 2.

The reference data in the full database, i.e. those of the 41 substances (referred to as set 1) and the 128 substances (referred to as set 2) already published Hoffmann et al. (2018), were compared with data from OECD curated databases (OECD, 2021c (Annex 2)). The comparison showed no

need to amend the reference data in our database, since the LLNA EC3 values and human potency categories were identical or very similar for more than 90% of the overlapping reference substances (102 for LLNA, 58 for human). Remaining differences may partly be explained by differences in inclusion criteria. In particular, the OECD process required that test concentrations were known, while Hoffmann et al. (2018) only required an EC3-value for positive substances.

2.3. Computationally generated data

Several computational tools were used to further characterise the substances within the database. Physico-chemical properties related for example to skin penetration, i.e. molecular weight (MW), octanol-water partition coefficient (KOW), water solubility (WS), vapour pressure (VP), melting (MP) and boiling point (BP) were calculated using the quantitative structure–property relationship (QSPR) developed by Zang et al. (2017). We also calculated these parameters using the OPERA QSPR of Mansouri et al. (2018), which is implemented in the Integrated Chemical Environment on the website of the US National Toxicology Program (<https://ice.ntp.niehs.nih.gov>). However, for our set of 169 substances and the six physico-chemical properties no significant differences between the two QSPR were found (data not shown). Since, the OPERA QSPR did not return results for 17 substances, including eight of the ones presented here for the first time, we present only the output from the QSPR by Zang et al. (2017).

Unless available from Roberts et al. (2007) or Urbisch et al. (2015), mechanistic reactivity domains of the substances were determined with Toxtree v2.6.13 (Patlewicz et al., 2008), freely available at www.toxtree.sourceforge.net, and the OECD QSAR toolbox 4.4 (www.qsartoolbox.org). Toxtree skin sensitisation reactivity domains and protein binding alerts as well as OASIS protein binding alerts for skin sensitisation obtained with the OECD QSAR toolbox 4.4 are provided for all substances in Supplemental Table 1. We also included the skin sensitisation predictions from Derek Nexus (version 6.0.1; Nexus version: 2.2.2), which were dichotomised by considering the prediction ‘non-sensitiser’ as negative and all other predictions as positive, and from TIMES-SS (Tissue Metabolism Simulator model for Skin sensitisation; version 2.30.1.11) by reporting the result for the parent, denoted as TIMES-P, and the most severe result of all metabolites, denoted as TIMES-M. These were generated because some defined approaches use these as input data, such as the integrated testing strategy described by Takenouchi et al. (2015), the Bayesian network of Jaworska et al. (2015), the stacking model by Del Bufalo et al. (2018) and the artificial neural network by Hirota et al. (2018).

Besides the primary use category that was essential for the selection, additional uses were obtained from the CPCat (Chemical and Product Categories) Database of the U.S. EPA, which categorises chemicals based on their usage or function in consumer products (Dionisio et al., 2015).

2.4. NAMs

As we intended to expand our publicly available database, the 41 substances were tested in the DPRA, the KeratinoSens™, the h-CLAT, the U-SENS™ and the SENS-IS, as briefly described by Hoffmann et al. (2018), using either the respective OECD Test Guidelines, where applicable, or the latest protocols of the NAM developers (Alepe et al., 2015; Cottrez et al., 2016; OECD 2018a; OECD 2018b; OECD 2021a).

The PPRA was included as a new NAM, as it was potentially of added value, especially for substances needing oxidative or metabolic activation (Gerberick et al., 2009; Ryan et al., 2020). The PPRA expands the DPRA by i) including horseradish peroxidase to potentially activate substances, ii) by testing several concentrations allowing for a concentration-response analyses and iii) by detecting peptide using mass spectroscopy. It has been described in detail by Gerberick et al. (2009), Troutman et al. (2011) and Ryan et al. (2020). In brief, the PPRA, as the DPRA, determines substance reactivity towards two synthetic peptides

containing either lysine or cysteine. Reactions with the cysteine peptide are performed without and with a horseradish peroxidase-hydrogen peroxide oxidation system (+/-HRP/P) for identifying directly reactive haptens and pre-/pro-haptens, respectively. Cysteine reaction mixtures + HRP/P are post-incubated adding of DL-dithiothreitol (DTT) to reduce dimer formation by reducing the oxidative effects of the peroxide/peroxidase. Lysine peptide reactivity is only measured in the absence of HRP/P. The reaction mixtures for cysteine and lysine peptides with the substances are comparable for DPRA and PPRA under the HRP/P-conditions, except that a different buffer system is used for the lysine reactions.

As of the set 1, two substances were not commercially available (see Table 1), 39 substances were tested with the PPRA. In addition, data for 126 of the set 2 substances were added to the database.

Table 1

Identity, primary use category, NAM results (as 0 (negative) and 1 (positive), reference results of the 41 substances of set 1 (+ median-like location parameter; ni: not interpretable due to co-elution; nt: not testable (due to insolubility); * Tested only up to solubility limit/500 µM; ** flat and no concentration-dependent cytotoxicity curve; *** interpret with caution due strong induction of metabolic activity; na: not available; na': not commercially available at the time of testing; ns: non-sensitiser; # strong sensitiser in GPMT (no LLNA available); "ns in GPMT(no LLNA available)).

Substance name	CAS registry no.	primary use category	NAM						reference data	
			DPRA	PPRA	Keratino Sens TM	h- CLAT	U- SENS TM	SENS- IS	human category	LLNA EC3 ⁺
hydroxytyrosol	10597-60-1	active	ni	1	1	1	1	1	6	1.10
taxifolin	480-18-2	active	1	1	0	1	1	0	6	1.10
dimethyl Caffeic acid phenethyl ester	145551-14-0	active	0	1	1	1	1	0	6	ns
Caffeic acid phenethyl ester (CAPE)	104594-70-9	active	1	1	1	1	1	1	6	0.9
folic acid	59-30-3	active	1	0	0*	0	0	1	na	ns
genistein	446-72-0	active	1	1	1	1	1	0	na	ns
1,3-Benzenediol, 4-(1-phenylethyl)-	85-27-8	active	0	1	1	1	1	1	5	ns
m-aminophenol	591-27-5	dye	1	1	0	1	1	1	3	0.24
4-Amino-2-hydroxytoluene	2835-95-2	dye	1	1	1	1	1	1	5	1.92
2-Methylresorcinol	608-25-3	dye	1	1	0	1	1	1	6	50.00
4-Amino-m-cresol	2835-99-6	dye	1	1	1	1	1	1	na	1.80
2-(4-amino-2-nitroanilino)ethanol	2871-01-4	dye	1	1	0	1	1	1	6	2.20
2-Amino-6-chloro-4-nitrophenol	6358-09-4	dye	ni	1	1	1	1	1	na	3.77
4-Hydroxypropylamino-3-nitrophenol	92952-81-3	dye	ni	1	1	1	1	1	na	ns
3-Amino-2-methylamino-6-methoxypyridin x HCl	83732-72-3	dye	ni	1	1	1	1	1	na	5.60
1-Hydroxyethyl-4,5-Diaminopyrazole Sulfate	155601-30-2	dye	1	1	1	1	1	1	na	ns
2,4,5,6 Tetraaminopyrimidine sulfate	5392-28-9	dye	1	1	0	1	1	0	na	ns
4-Amino-3-nitrophenol	610-81-1	dye	1	1	0	1	1	1	5	0.20
2,4-Diaminophenoxyethanol dihydrochlorid	66422-95-5	dye	1	1	1	1	1	1	na	3.20
1,3-Bis-(2,4-diaminophenoxy)propane 4 HCl	74918-21-1	dye	1	1	1	1	1	1	na	14.70
2-Amino-3-hydroxypyridine	16867-03-1	dye	1	1	0	1	1	1	5	ns
2,7-Dihydroxynaphthalin	582-17-2	dye	1	1	1	1	1	1	6	2.80
2-Chlor-6-methyl-3-aminophenol	84540-50-1	dye	1	1	1	1	1	1	na	ns
1-methyl-2,6-bis-(2-hydroxyethylamino) benzene	149330-25-6	dye	1	1	1	1	1	0	na	ns
1,5-Dihydroxynaphthalin	83-56-7	dye	ni	1	1	1	1	1	na	3.40
n,n'-bis-(2-hydroxyethyl)-2-nitro-p- phenylenediamine	84041-77-0	dye	1	1	1**	1	0	1	na	ns
hydroxyethyl-p-phenylenediamine sulfate	93841-25-9	dye	ni	na^	1	0	1	1	na	0.57
2-[(4-aminophenyl)(2-hydroxyethyl)amino] ethanol, sulfate	54381-16-7	dye	ni	1	1	1	1	1	5	1.04
hydroxypropyl bis(n-hydroxyethyl-p- phenylenediamine HCl	128729-30-6	dye	ni	1	1	1	1	1	na	na [#]
oxazolone	15646-46-5	other	1	1	1	1	1	1	na	0.003
methyl dihydrojasmonate	24851-98-7	fragrance	1	1	0	1	0	0	na	ns
tetraethylene glycol monotetradecyl ether	39034-24-7	surfactant	0	0	0	1	1	0	na	2.80
hexaethylene glycol monododecyl ether	3055-96-7	surfactant	0	0	0	1	1	0	na	7.40
n-heptyl b-d-thioglucofuranoside	85618-20-8	surfactant	0	0	0	1	1	0	na	8.30
1-nonane sulfonic acid sodium salt	35192-74-6	surfactant	1	0	0	0	0	0	na	ns
quaternium 15 cis-isomer	51229-78-8	surfactant	1	1	1	1	1	1	2	ns
octyl-methoxycinnamate	5466-77-3	UV filter	0	0	1	0	1	0	5	ns"
octocrylene	6197-30-4	UV filter	0	0	1***	1	1	1	5	7.70
butyl Methoxydibenzoylmethane (BMDBM)	70356-09-1	UV filter	0	1	0	1	1	0	5	ns"
2-phenylbenzimidazole-5-sulfonic acid (PBSA)	27503-81-7	UV filter	nt	na^	1	0	0	0	5	ns"
2-Hydroxy-4-methoxybenzophenone	131-57-7	UV filter	1	0	1	1	1	0	4	ns

2.5. The database

For completeness, the entire database of 169 substances, i.e., the here introduced 41 substances (set 1) and the original version with 128 substances (set 2), is provided as a [Supplemental Table 1](#), in which substances of set 1 are highlighted. This version also includes a few corrections and updates for substances of set 2, especially for *in silico* software versions. Substances are identified by substance name, CAS-registry number, and SMILES. In addition, primary use categories, which were assigned based on uses by CE member companies, are included. It contains information on the mechanistic reactivity domains, an indication, if a substance requires activation ([Patlewicz et al., 2016](#); [Urbisch et al., 2015](#)), skin sensitisation predictions of DEREK Nexus (version 6.0.1; Nexus version: 2.2.2) and TIMES-SS (version 2.30.1.11),

and six estimated physico-chemical properties (MW, BP, MP, KOW, WS, VP). For the six NAMs the following parameters were included:

- DPRA: cysteine and lysine peptide depletion (%); prediction of reactivity class and potential
- PPRA: Cysteine, Cysteine + HRP/P, Lysine: DPMIN (in %), DPMAX (in %), EC25 (mM); prediction of reactivity class and potential
- KeratinoSens™: EC1.5; EC3; IC50; I_{max} (mM); for natural extracts in mg/mL; prediction as negative or positive
- h-CLAT: CD86 EC150; CD54 EC200, CV75 (μg/mL); prediction as negative or positive
- U-SENS™: CD86 EC150; CV70 (μg/mL); prediction as negative or positive
- SENS-IS: potency prediction, vehicle, and gene expression data; potency and hazard prediction

All newly generated data, i.e., the data of the set 1 substances as well as a substantial part of PPRA data of set 2 substances, were obtained in single experiments.

Human and animal reference results are provided, including experimental details and references, where applicable, for the LLNA (and GPMT). While animal *in vivo* data were available for all substances, we found no or insufficient human data for 22 substances of set 1 to define a potency category.

2.6. Data analysis

We descriptively compared the two sets of substances in terms of their use categories and physico-chemical properties, including a Mann-Whitney test for each property (Graph-Pad Prism 7.0, www.graphpad.com). Distributions of the individual properties for all substances were graphically described by a combination of half-violin, box- and dot-plots (using the R package 'ggplot2' (Wickham 2016), version 3.2.1). We chose the width parameter to obtain a single-peaked violin-shape, as it was assumed that the underlying distributions would be unimodal.

The predictivity of the individual NAMs was evaluated in detail. Applying the standard prediction models (PM) or dichotomised versions thereof, predictivity was described for the six NAMs by specificity, i.e., the proportion of the non-sensitising substances predicted as non-sensitisers, sensitivity, i.e., the proportion of the sensitising substances predicted as sensitisers, accuracy (or concordance), i.e., the proportion of the correctly predicted substances, and balanced accuracy, i.e., the mean of specificity and sensitivity. We also calculated the negative predictive values (NPV), i.e. the proportion of correctly predicted non-sensitising substances among all substances predicted as non-sensitisers, and the positive predictive values (PPV), i.e. the proportion of correctly predicted sensitising substances among all substances predicted as sensitisers, which are dependent of the prevalence and an indication for the reliability of a positive/negative test result (Desprez et al., 2021; Hoffmann and Hartung 2005). Although developed using LLNA reference data, predictivities were not only obtained for LLNA reference data, but also for human reference data. They were calculated for the individual substance sets and their combination.

Mechanistically similar NAMs addressing the same AOP key event, i. e., the h-CLAT and the U-SENS™ as well as the DPRA and PPRA, were compared by calculating Spearman's correlation coefficient, where applicable, and the proportion of concordant predictions.

3. Results

3.1. Description of the expanded database

The expanded database (Supplemental Table 1) contained in total 169 substances, i.e., set 1 with 41 new substances and set 2 with 128 substances (Hoffmann et al., 2018). The two subsets were compared by a descriptive analysis of the use categories. Since the primary use category

(Table 1) was the main selection criterion for set 1, the two datasets differed considerably in this regard. For example, set 1 contained many more dyes (54% vs 5% in set 2) and UV-filters (12% vs 1% in set 2), while there were fewer fragrances (5% in set 1 vs 45% in set 2) and no preservatives and pharmaceuticals (Fig. 1).

This difference is also reflected to some extent in the CPCat use categories, which may assign a substance to several categories (see Supplemental Table 1).

We compared the distributions for the six physico-chemical properties of the two substance sets (Fig. 2A). The mean of the MW distribution of the substances in set 1 was substantially higher compared to that of set 2, as confirmed by a Mann-Whitney test (p-value < 0.001). This difference was also reflected in the distributions of the BP, MP, and VP, which are properties moderately to strongly correlated (data not shown) with the MW. In contrast, a smaller difference was found for KOW (p-value = 0.015) and no difference for WS (p-value = 0.783). The distributions of the combined data sets are shown for each physico-chemical property in Fig. 2B as combined violin-box-dot plots. These plots demonstrate that the substances covered broad ranges of the various properties, with a few extreme values, e.g., for KOW and BP, and that distributions were approximately symmetric for untransformed properties (MW, BP, MP) and for log-transformed properties (KOW, WS, VP). In addition, they provide basic information of particular interest regarding the applicability domains of the NAMs. For example, they indicate that a substantial number of substances with log KOW >3.5 were included, and that the proportion of substances with a high MW (>400) and high VP was limited.

3.2. Reference and NAMs results

The 41 substances of set 1 are listed in Table 1. In addition, their primary use categories, dichotomised results for the six NAMs as well as the human and LLNA reference data are presented. All further information, such as physico-chemical properties and the primary NAM results, are provided in the Supplemental Table 1. No major technical applicability issues were met, with the exception that DPRA results could not be obtained for several substances, primarily dyes, due to cysteine co-elution. Human category indicating a human skin sensitisation risk could be determined for 19 substances. Most of these (16) fell into categories 5 and 6, which we considered as not warranting a classification, i.e., deemed to be non-sensitisers. For 32 substances a single LLNA result was available. For the five substances with two LLNA results, the median-like location parameter was calculated as described by Hoffmann et al. (2018). For the remaining four substances, LLNA results could not be identified. In these cases, results from GPMT were used, assuming that, if tested in the LLNA, the results in terms of hazard would be identical. Results of the *in vivo* tests covered the entire skin sensitisation potency spectrum, including the extremely sensitising oxazolone and 18 non-sensitising substances. Most LLNA-sensitising substances (15) were moderate skin sensitisers (EC3 in the range of 1–10%).

The complete database (set 1 + 2) comprised 169 substances, 165 with LLNA reference data, 4 with GPMT reference data and 141 with human reference data.

3.3. Comparison of NAMs

3.3.1. Basic NAM differences in binary prediction frequencies

After successful development and standardisation, the predictivity of an individual NAM for a specific purpose is usually determined by construction of a PM that condenses the primary NAM outputs into a value supporting decision making. For the six skin sensitisation methods considered here PMs returning a dichotomous result existed, which were obtained by comparison to skin sensitisation hazard reference results as determined by the LLNA. While it is beyond our scope here to reconstruct the PM development, it is important to acknowledge that these PMs have been developed with different substances, different ratios of

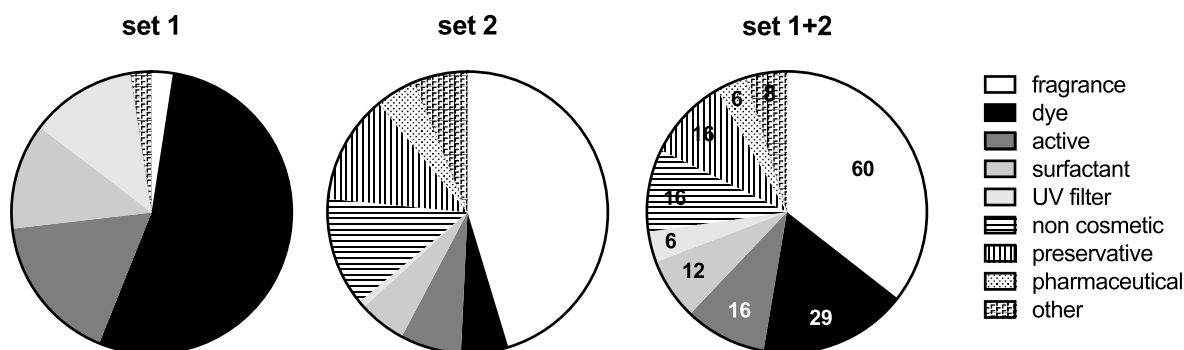


Fig. 1. Primary use categories represented in set 1, i.e., new 41 substances, set 2, i.e., those 128 substances previously published, and the combined set 1 + 2.

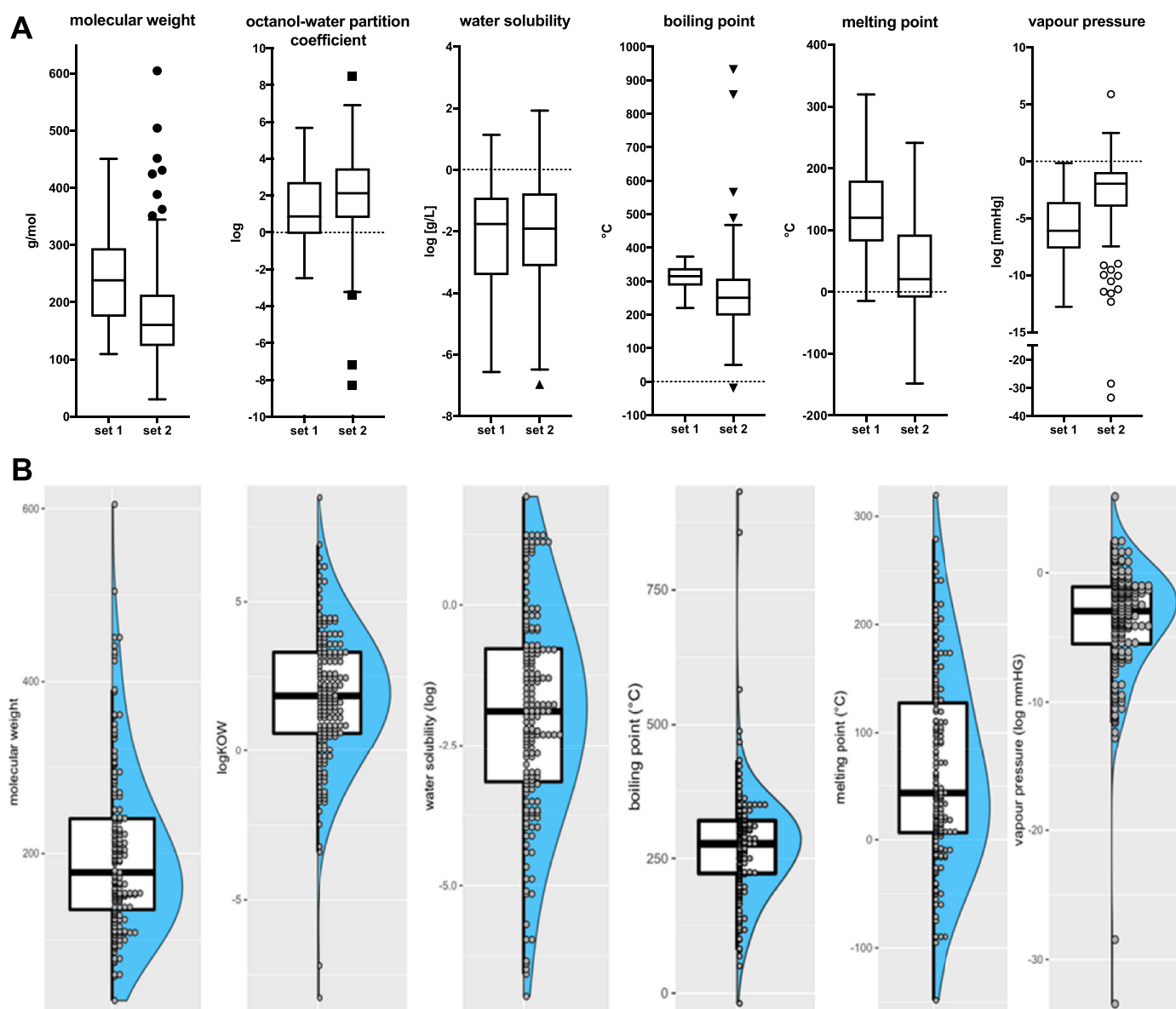


Fig. 2. A: Box-plots comparing the two data sets (set 1: 41 new substances; set 2: 122 substances, excl. Six natural extracts (Hoffmann et al., 2018)); B: Combined half-violin-box-dot plots of the distributions for six physico-chemical properties for the combined set of 163 substances.

substances with a negative and positive reference results, and with different objectives. While e.g., the DPRA PM model was developed to maximise accuracy, the predictive performance of other NAMs suggests

that they were developed favouring sensitivity over specificity, e.g., both test methods addressing the AOP key event 3 (Ashikaga et al., 2010; Piroird et al., 2015). These differences need to be accounted for

when assessing individual NAMs, comparing predictivity among NAMs and when combining them in a testing strategy. Therefore, we calculated the number of positive and negative predictions for both sets, both reference standards and the six NAMs in our database (Table 2). While the ratios of positive to negative prediction over both substance sets was approximately in the range of 1.5–2.5 for both reference standards, the DPRA, KeratinoSens™ and PPRA, this ratio ranged between 3.4 and 5.0 for the h-CLAT, U-SENS™ and the SENS-IS.

In addition, there were clear differences between the two sets of substances. The proportion of LLNA negatives was much higher in set 1 compared to set 2, but the proportion of negative predictions was substantially lower for the DPRA, PPRA, h-CLAT and U-SENS™ (as indicated by a higher ratio). This indicates that substances of set 1 may not have been part of the substance sets used to define the prediction models of these NAMs. Interestingly, the SENS-IS was the only NAM, for which the ratio of positive to negative predictions was lower for set 1 compared to set 2. The main observation that, whilst within the applicability domain of the NAM, *in vivo* negatives in set 1 were often predicted as positive, will be further addressed in the following analysis of the predictive capacity.

3.3.2. Comparison of the predictivity of NAMs

Several parameters were calculated for the two substance sets individually and the combined set to comprehensively describe the predictivity of the NAMs for discriminating sensitisers from non-sensitisers. Therefore, we compared the predictions obtained with the standard PMs of the NAMs, which were developed with LLNA data, against the LLNA reference data and the human potency reference data. In addition, the two references were compared (Table 3). For NAMs with PMs that have more than two categories, i.e., the DPRA, PPRA and the SENS-IS, all categories indicating reactivity and sensitisation potential were combined and considered as positive.

Human predictivity for set 1 needs to be evaluated with care due to the small sample size. The only parameter which allows some interpretation is the specificity (sample size of 13 or 16), which was generally lower as compared to set 2. When combining the sets, this consequently resulted in lower specificity as compared to set 2 alone. Due to the low number of human positives in set 1, sensitivities of the combined set were like those of set 2. All NAMs and the LLNA had similar (balanced) accuracy when compared to the human data of set 2. However, specificities and sensitivities differed, being rather balanced for the DPRA and KeratinoSens™ and skewed towards sensitivity for the h-CLAT, U-SENS™ and SENS-IS, supporting the observation that prediction models were optimised for different objectives (Fig. 3).

The same skewness was observed for the LLNA predictivity. Considering the combined set, the (balanced) accuracy was highest for the SENS-IS, especially as it was the most predictive NAM for set 1.

LLNA predictivity for set 1, which was based on a sample size from 32 (DPRA) to 41, was in generally lower as compared to set 1, except for SENS-IS. The reason for this was that the SENS-IS had a much better

specificity than the other NAMs. However, this result needs to be interpreted carefully due to the very low concordance of human and LLNA predictions for set 1.

3.3.3. Pairwise NAM comparison

We directly compared the NAMs pairwise based on the concordance of their binary predictions, i.e., not considering reference results, to better understand their mutual (in-)dependence. In this context, two NAMs are maximal dependent if they predict the same binary result for all substances, i.e., if they are 100% concordant. This approach to NAM comparison may inform the strategic NAM use, e.g., in a defined approach, as higher independence of two NAMs may be indicative of complementary predictive value, while the combination of two highly dependent NAMs is less promising. For this purpose, we analysed the concordance of predictions of all NAM pairs (Table 4). Focusing on balanced overall concordance, i.e., the mean of positive and negative concordance, the least concordant, i.e., below 60%, and thus most independent pairs, were DPRA & SENS-IS and KeratinoSens™ & h-CLAT, regardless of order. The only pair with a high balanced concordance regardless of order was DPRA & PPRA, which can be explained by the identical purpose to measure reactivity and their experimental similarity. Also, the pairs h-CLAT & PPRA, U-SENS™ & KeratinoSens™, U-SENS™ & h-CLAT and U-SENS™ & PPRA showed balanced concordance of over 70%, but only in the order indicated. The high concordance of the U-SENS™ with the KeratinoSens™ can be explained by the high concordance of negative predictions. The high dependence of the U-SENS™ with the h-CLAT might be due to their mechanistic similarity, addressing the same AOP key event, while potential explanations for the dependence of these KE3-NAMs with the PPRA are less evident.

3.3.4. Comparison of NAMs addressing AOP KE 1: DPRA and PPRA

While NAMs addressing the same AOP event are considered mechanistically similar, we investigated how that similarity is reflected in our database. The first pair of similar NAMs are the DPRA and the PPRA, which model the molecular initiating event of the skin sensitisation AOP, i.e., covalent binding to skin proteins, by measuring peptide depletion *in chemico*. Both NAMs measure the depletion of two model synthetic peptides either containing cysteine or lysine OECD TG 442C (OECD 2021a; Ryan et al., 2020). In addition, the PPRA determines cysteine depletion after a potential activation step using horseradish peroxidase for characterising reactivity of hapten-as well as pre-/pro-hapten sensitisers. Furthermore, the ratios of the peptide amount to the test substance amount and the lysine buffer differ between the two NAMs. This leads to differences in depletion values that limit the direct comparison of depletion values, as evident from moderately correlated cysteine depletion values (Spearman coefficient of 0.585), and very weakly correlated lysine depletion values (Spearman coefficient of 0.201).

Fig. 4A comprehensively summarises the comparison of predictive performance for the DPRA and PPRA based on binary predictions. It

Table 2

Number of positive and negative results and their ratio, for the two references (human and LLNA) and the six NAMs.

set		human	LLNA	DPRA	PPRA	KeratinoSens™	h-CLAT	U-SENS™	SENS-IS
1	no. of substance	19	41	32	39	41	41	41	41
	no. positive predictions (A)	3	23	24	31	27	36	36	26
	no. negative predictions (B)	16	18	8	8	14	5	5	15
	ratio A/B	0.19	1.28	3.00	3.88	1.93	7.20	7.20	1.73
2	no. of substance	128	128	124	126	128	127	126	126
	no. positive predictions (A)	88	95	71	87	75	98	103	103
	no. negative predictions (B)	40	33	53	39	53	29	23	23
	ratio A/B	2.20	2.88	1.34	2.23	1.42	3.38	4.48	4.48
1 + 2	no. of substance	147	169	156	165	169	168	167	167
	no. positive predictions (A)	91	118	95	118	102	134	139	129
	no. negative predictions (B)	56	51	61	47	67	34	28	38
	ratio A/B	1.63	2.31	1.56	2.51	1.52	3.94	4.96	3.39

Table 3

Skin sensitisation potential predictivity of individual NAMs compared to human and LLNA reference data for substance sets 1, 2 and 1 + 2 (ns: non-sensitiser; s: sensitiser; *no. of substances, if reduced; N/PPV: negative/positive predictive value).

reference	substance set (n = ns + s = total)	NAM/LLNA (in grey)	predictive parameters [%]					
			specificity	sensitivity	accuracy	balanced accuracy	NPV	PPV
human	1 (n = 16 + 3 = 19)	DPRA (n = 16)*	38.5	100.0	50.0	69.2	100.0	27.3
		PPRA (n = 18)	13.3	66.7	22.2	40.0	66.7	13.3
		KeratinoSens™	37.5	66.7	42.1	52.1	85.7	16.7
		h-CLAT	12.5	100.0	26.3	56.3	100.0	17.6
		U-SENS™	6.3	100.0	21.1	53.1	100.0	16.7
		SENS-IS	31.3	66.7	36.8	49.0	83.3	15.4
	2 (n = 40 + 88 = 128)	LLNA	37.5	33.3	36.8	35.4	66.7	9.1
		DPRA (n = 124)	76.9	72.9	74.2	74.9	56.6	87.3
		PPRA (n = 126)	55.0	81.4	73.0	68.2	57.9	79.5
		KeratinoSens™	77.5	75.0	75.8	76.3	58.5	88.0
		h-CLAT (n = 127)	50.0	89.7	77.2	69.8	69.0	79.6
		U-SENS™ (n = 126)	42.5	93.0	77.0	67.8	73.9	77.7
	1 + 2 (n = 56 + 91 = 147)	SENS-IS (n = 126)	47.5	94.2	79.4	70.8	79.2	79.4
		LLNA	50.0	85.2	74.2	67.6	60.6	78.9
		DPRA (n = 140)	67.3	73.9	71.4	70.6	60.3	79.3
		PPRA (n = 144)	43.6	80.9	66.7	62.3	58.5	69.9
		KeratinoSens™	66.1	74.7	71.4	70.4	61.7	78.2
		h-CLAT (n = 146)	39.3	90.0	70.5	64.6	71.0	70.4
		U-SENS™ (n = 145)	32.1	93.3	69.7	62.7	75.0	68.6
		SENS-IS (n = 145)	42.9	93.3	73.8	68.1	80.0	72.2
		LLNA	46.4	83.5	69.4	65.0	63.4	71.7
		DPRA (n = 32)	25.0	75.0	50.0	50.0	50.0	50.0
		PPRA (n = 39)	23.5	81.8	56.4	52.7	50.0	58.1
		KeratinoSens™	35.3	65.2	52.5	50.3	42.9	57.7
LLNA	1 (n = 18 + 23 = 41)	h-CLAT	22.2	95.7	63.4	58.9	80.0	61.1
		U-SENS™	27.8	100.0	68.3	63.9	100.0	63.9
		SENS-IS	61.1	82.6	73.2	71.9	73.3	73.1
	2 (n = 33 + 95 = 128)	DPRA (n = 124)	67.7	65.6	66.1	66.7	39.6	85.9
		PPRA (n = 126)	59.4	79.8	74.6	69.6	50.0	85.2
		KeratinoSens™	66.7	67.4	67.2	67.0	41.5	85.3
		h-CLAT (n = 127)	48.5	86.2	76.4	67.3	55.2	82.7
		U-SENS™ (n = 126)	40.6	89.4	77.0	65.0	56.5	81.6
		SENS-IS (n = 126)	46.9	90.4	79.4	68.7	62.5	83.3
	1 + 2 (n = 51 + 118 = 169)	DPRA (n = 156)	53.2	67.0	62.8	60.1	41.0	76.8
		PPRA (n = 165)	46.9	80.2	70.3	63.6	50.0	78.2
		KeratinoSens™	56.0	66.9	63.7	61.5	41.8	78.2
		h-CLAT (n = 168)	39.2	88.0	73.2	63.6	58.8	76.9
		U-SENS™ (n = 167)	36.0	91.5	74.9	63.7	64.3	77.0
		SENS-IS (n = 167)	52.0	88.9	77.8	70.4	66.7	81.3

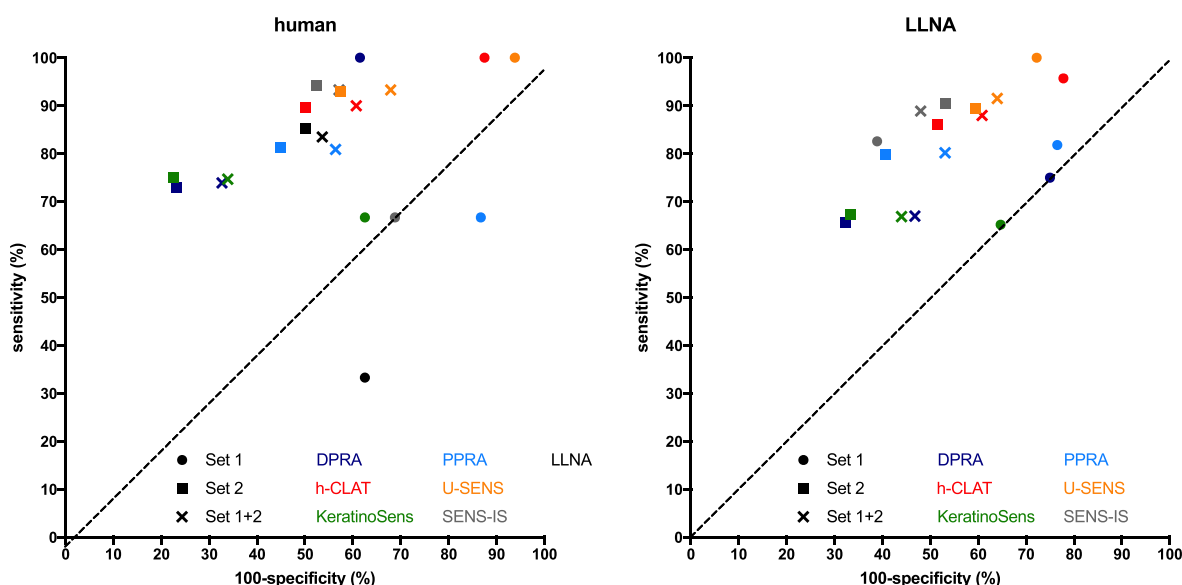


Fig. 3. Dot-plot of the specificity and sensitivity of the individual NAMs and the individual and combined substance sets for both reference standards.

Table 4

Pairwise NAMs' concordance (*first NAM used as denominator, e.g. of 61 DPRA negatives 37 were also negative in the KeratinoSens™).

NAMs pair*		Concordance of predictions [%]			
		negative	positive	all	all balanced
DPRA	PPRA (n = 155)	54.1 (33/61)	88.3 (83/94)	74.8	71.2
	KeratinoSens™ (n = 156)	60.7 (37/61)	71.6 (68/95)	67.3	66.1
	h-CLAT (n = 156)	32.8 (20/61)	88.4 (84/95)	66.7	60.6
	U-SENS™ (n = 155)	27.9 (17/61)	91.5 (86/94)	66.5	59.7
	SENS-IS (n = 155)	32.8 (20/61)	81.9 (77/94)	62.6	57.4
PPRA	DPRA (n = 155)	75.0 (33/44)	74.8 (83/111)	74.8	74.9
	KeratinoSens™ (n = 165)	65.2 (30/46)	68.9 (82/119)	67.9	67.1
	h-CLAT (n = 165)	45.7 (21/46)	90.8 (108/119)	78.2	68.2
	U-SENS™ (n = 165)	37.0 (17/46)	91.6 (109/119)	76.4	64.3
	SENS-IS (n = 165)	39.1 (18/46)	84.0 (100/119)	71.5	61.6
KeratinoSens™	DPRA (n = 156)	57.8 (37/64)	73.9 (68/92)	67.3	65.9
	PPRA (n = 165)	44.8 (30/67)	83.7 (82/98)	67.9	64.2
	h-CLAT (n = 168)	26.9 (18/67)	84.2 (85/101)	61.3	55.5
	U-SENS™ (n = 167)	32.8 (22/67)	94.0 (94/100)	69.5	63.4
	SENS-IS (n = 167)	38.8 (26/67)	88.0 (88/100)	68.3	63.4
h-CLAT	DPRA (n = 156)	64.5 (20/31)	67.2 (84/125)	66.7	65.9
	PPRA (n = 165)	65.6 (21/32)	81.2 (108/133)	78.2	73.4
	KeratinoSens™ (n = 168)	52.9 (18/34)	63.4 (85/134)	61.3	58.2
	U-SENS™ (n = 167)	47.1 (16/34)	91.0 (121/133)	82.0	69.0
	SENS-IS (n = 167)	41.2 (14/34)	82.0 (109/133)	73.7	61.6
U-SENS™	DPRA (n = 155)	68.0 (17/25)	66.2 (86/130)	66.5	67.1
	PPRA (n = 165)	63.0 (17/27)	79.0 (109/138)	76.4	71.0
	KeratinoSens™ (n = 167)	78.6 (22/28)	67.6 (94/139)	69.5	73.1
	h-CLAT (n = 167)	57.1 (16/28)	87.1 (121/139)	82.0	72.1
	SENS-IS (n = 167)	53.6 (15/28)	83.5 (116/139)	78.4	68.5
SENS-IS	DPRA (n = 155)	54.1 (20/37)	65.3 (77/118)	62.6	59.7
	PPRA (n = 165)	48.6 (18/37)	78.1 (100/128)	71.5	63.4
	KeratinoSens™ (n = 167)	68.4 (26/38)	68.2 (88/129)	68.3	68.3
	h-CLAT (n = 167)	36.8 (14/38)	84.5 (109/129)	73.7	60.7
	U-SENS™ (n = 167)	39.5 (15/38)	89.9 (116/129)	78.4	64.7

demonstrates the good (balanced) concordance of the two NAMs, which is higher than the concordance with both reference standards. Predictivity for positive reference results is substantially larger than for negative reference results for both NAMs as indicated by the size of the white squares. In addition, the potential of the horseradish peroxidase in the PPRA to activate pre- and pro-haptens was explored. Of the 23 substances in our database that were identified as requiring activation, the DPRA identified 14 substances and the PPRA classified 21 as reactive. These seven substances explain some of the difference in the discordant predictions for both human and LLNA as graphically represented in Fig. 4A (light grey squares). The two substances geraniol and farnesol, which are structurally related and therefore could be activated in similar ways, were not detected as reactive by any of the two methods.

These findings confirm that the technically simpler DPRA is the obvious default choice to address KE1, e.g., in a DA, and that the PPRA has a particular advantage when assessing suspected pre-/prohaptens.

3.3.5. Comparison of NAMs addressing AOP KE3: h-CLAT and U-SENS™

The h-CLAT and the U-SENS™ address the KE 3 of the AOP for skin sensitisation by modelling dendritic cell activation (OECD 2018b). Both NAMs measure cell viability by propidium iodide staining, determining the concentration (in µg/mL) with 75% or 70% cell viability, respectively, and the cell surface marker CD86, determining the concentration that induces 1.5-fold increased response as compared to the vehicle control. Main differences include *inter alia* the cell lines used (THP-1 vs U937), the maximum test concentration (5000 vs 200 µg/mL), the prediction models, the readout for the cell surface marker (mean fluorescence intensity of positive cells in the h-CLAT vs % of positive cells (among all cells) in the U-SENS™) and that the h-CLAT measures CD54 as a second marker for skin sensitisation. As therefore a correlation analysis of the primary measured data is not meaningful, we compared the dichotomised results calculating the concordance. Focusing on CD86 only, we obtained a concordance of 59.3% (99/167). Comparing the predictions when including CD54 for the h-CLAT, the concordance

increased to 82.0% (137/167), with 16 substances being negative and 121 being positive in both NAMs (Table 3, Fig. 4B). This is the highest concordance of all NAM pairs. In addition, the discordant predictions were rather balanced (h-CLAT positive and U-SENS™ negative: 12; h-CLAT negative and U-SENS™ positive: 18) and both NAMs have similar predictivities for both references (Table 2, Fig. 4B). It needs to be acknowledged, however, that also prediction model development and optimisation may have contributed to this similarity.

Finally, we evaluated the concordance of the h-CLAT and U-SENS™ separately for the sensitiser and non-sensitisers of both reference standards. Whilst developed in comparison to the LLNA, the proportion of substances predicted correctly as positive in both NAMs were very similar (human: 76/89 = 85.4%; LLNA: 95/112 = 81.2%). Two substances per reference standard were concordantly falsely underpredicted (kanamycin sulfate and allyl phenoxyacetate compared to humans; tocopherol and allyl phenoxyacetate compared to LLNA). Compared to the negative reference data, approximately 50% of the substances, half of which were from set 1, were predicted falsely positive. The most obvious explanation for this overprediction is a conservative prediction model optimisation approach favouring sensitivity over specificity.

In addition, less than 30% of the substances were correctly predicted as negative in both NAMs, regardless of the reference standard. This more detailed analysis supports the comparable performance of the h-CLAT and U-SENS™, providing a basis to explore interchangeable use in DAs.

3.3.6. Analysis of the octanol/water partition coefficient

The discussion of the predictive value of the h-CLAT for poorly water-soluble substances as compared to LLNA data initiated by Ashikaga et al. (2010) resulted in a proposal of considering negative predictions for substances with a logKOW >3.5 as unreliable (Takenouchi et al., 2013). This led to the fact that negative h-CLAT results for such substances should not be considered in a regulatory context (OECD, 2018b). Therefore, we investigated negative NAM predictions of

A

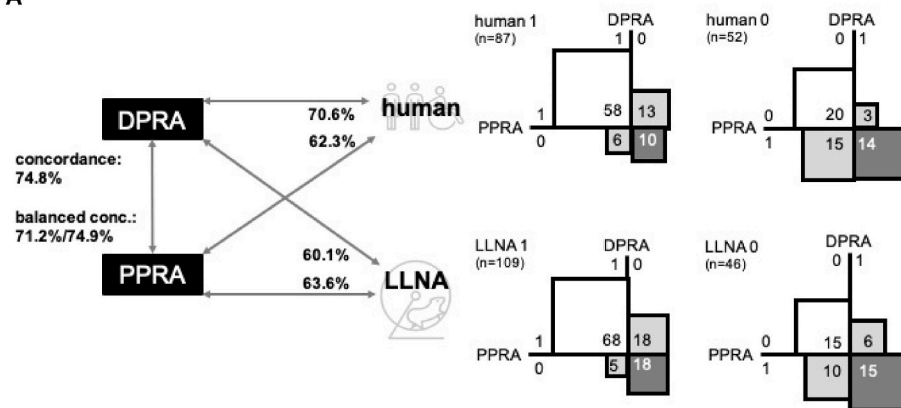
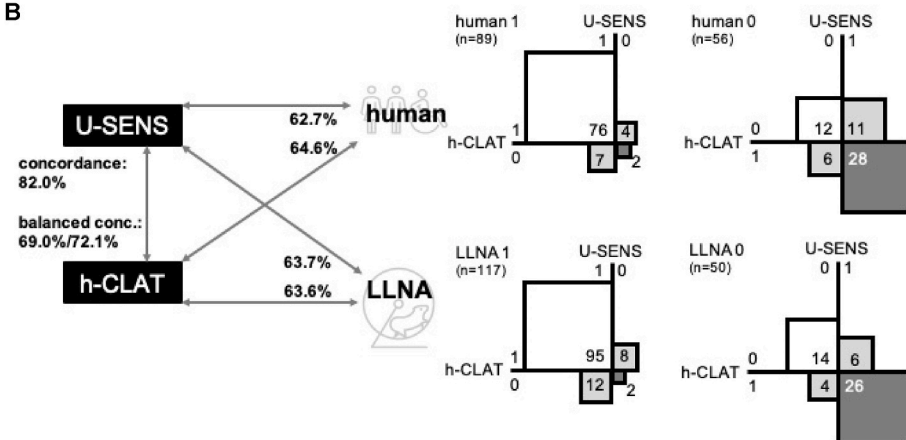


Fig. 4. Comprehensive comparison of A) the DPRA and PPRA, and B) the U-SENSTM and h-CLAT (%-values for comparison with reference standards are balanced concordances; 1: sensitiser (human, LLNA)/reactive (DPRA, PPRA); 0: non-sensitiser (human, LLNA)/negative (DPRA, PPRA) square size reflects respective proportions; number are absolute number of substances; white squares indicate concordance of both NAMs with the reference standard, light grey indicates one NAM concordant with the reference standard and dark grey indicates concordance of NAMs, but discordance with the reference standard).

B



substances in both sets with positive LLNA reference data and of substances with positive human reference data. We determined the relative frequency of these negative predictions in subsets of substances grouped by their logKOW to allow for a comparison of test methods and to identify trends. The grouping was anchored at logKOW of 3.5, used equidistant spacing and aimed at similar group sizes ranging from 20 to 37, except for the group with a logKOW >4.5, which included, depending on the data available per test method, 11 or 12 substances. The under-prediction rate was lower for three NAMs, which were tuned for sensitivity, i.e., the h-CLAT, U-SENSTM and SENS-IS (Fig. 5). In addition, no association of logKOW and the under-prediction rate was observed for these three NAMs. In contrast, this rate increased for substances with a logKOW between 3.5 and 4.5 for the DPRA, PPRA and KeratinoSensTM. Interestingly, the under-prediction rate of the DPRA, PPRA and KeratinoSensTM for the small group of substances with a logKOW >4.5 was high compared to LLNA reference results, but not when compared to human reference data. This difference was due to six substances that were moderately or weakly sensitising in the LLNA, but which were considered not to be human sensitisers. Our analysis indicates that the assumption that a high logKOW may lead to insufficient exposure of the reacting target, i.e., cells or model peptides, cannot be generalised across all NAMs, and that inter-species differences seem to play a role. Therefore, the uncertainty associated with the results of some NAMs for high logKOW substances may translate into a limitation of the applicability for classification and labelling purposes but may be compensated for in a risk assessment context.

Note that the same type of analysis was conducted for the parameter molecular weight without observing any pattern in false prediction rates (data not shown).

4. Discussion and conclusion

Since the early 2000s, many experimental NAMs for skin sensitisation have been proposed to assess the skin sensitisation potential of substances for humans. The development and the assessment of these NAMs have been driven by mechanistic understanding, as summarised in the AOP for skin sensitisation (OECD 2014), and the availability of LLNA reference data of acceptable quality (Basketter and Cadby 2004; Dumont et al., 2016; Hoffmann 2015). The reliance on LLNA data led to a focus on a set of approximately 200 substances, of which subsets have been used to validate the individual NAMs, to develop binary prediction models and to assess the resulting predictive capacities. These prediction models were developed to predict LLNA reference results considering the individual NAMs as stand-alone test methods, while acknowledging that a combination of NAMs addressing different AOP KE is required to reliably assess the skin sensitisation potential of substances (Ezendam et al., 2016; Urbisch et al., 2015). In addition, the PMs were optimised differently with respect to specificity and sensitivity, which severely complicates a comprehensive direct comparison of NAM predictivities. Nor were the PMs optimised to be used in combination with other information, e.g., in DA, so that combination of predictions will likely fall short of the optimal complementary use of the information provided by the individual NAMs.

With a focus on applicability domains, we expanded our database as provided in Hoffmann et al. (2018) to further explore six skin sensitisation NAMs DPRA, KeratinoSensTM, h-CLAT, U-SENSTM, SENS-IS and PPRA (Reisinger et al., 2015). A set of 41 substances (set 1) of high relevance for cosmetic industry, comprising dyes, UV-filters, surfactants, fragrances, and actives, has been selected that were, at least to a large extent, not part of the substance sets to develop and optimise the individual NAMs. These substances, which did not differ in terms of

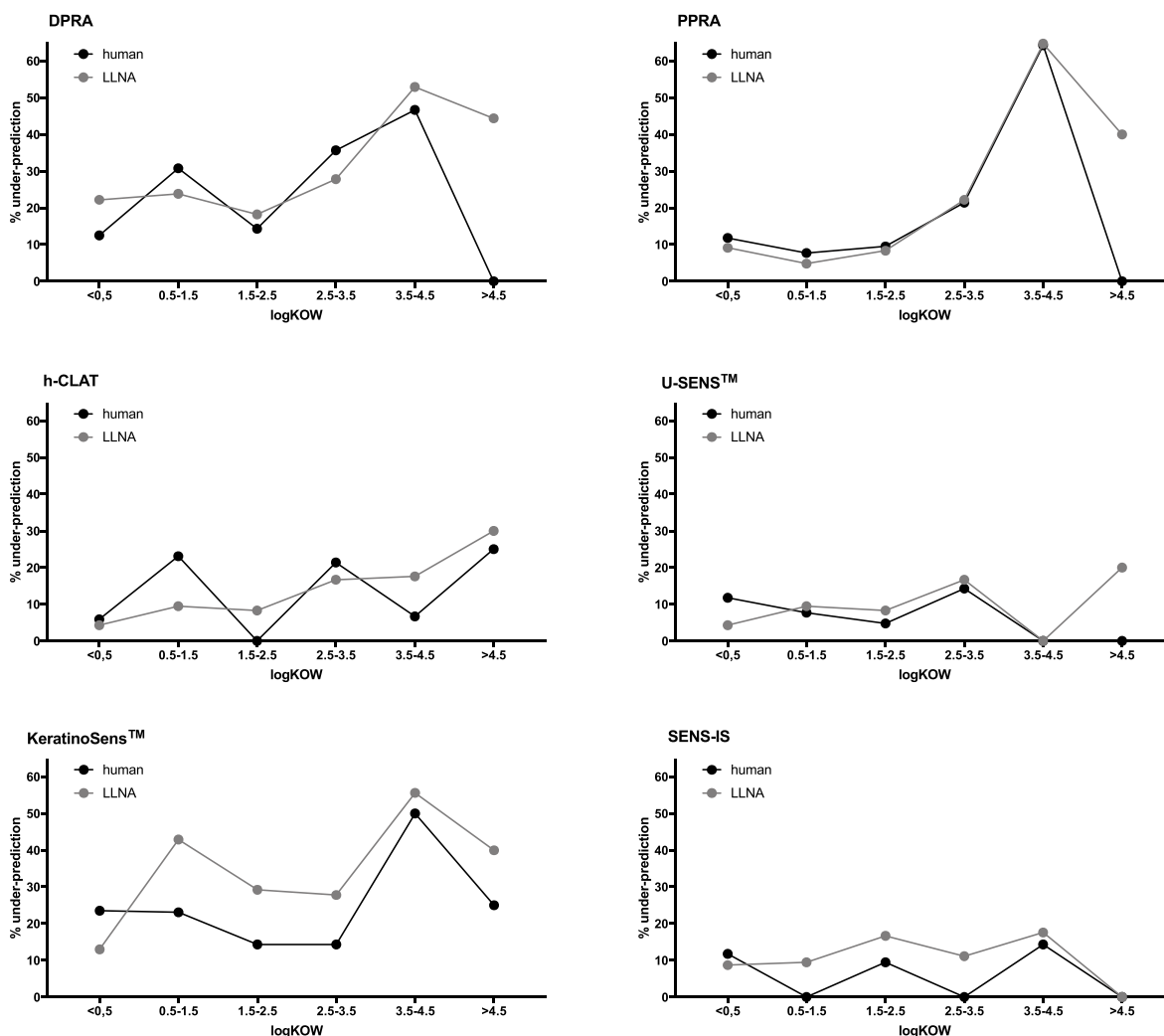


Fig. 5. Under-prediction rates (%) of human (black) and LLNA (grey) sensitising chemicals grouped by the logarithm of the octanol-water partition coefficient (logKOW) for the six NAMs.

physico-chemical properties from those reported earlier, were tested in experienced laboratories under blind conditions in the six NAMs.

In terms of technical limitation, we observed that several dyes co-eluted in the DPRA, which is related to the HPLC detection method and can possibly be overcome by modifications, e.g., using mass spectroscopy as in the PPRA (Fujita et al., 2019b; Ryan et al., 2020). In addition, our analysis of the association of NAM mispredictions with the MW and the logKOW results indicate an increase in false negative predictions for increasing logKOW for some NAMs, i.e., DPRA, PPRA KeratinoSens™, which is generally more pronounced for comparisons with LLNA results. Interestingly, negative concordance of human and LLNA reference data is substantially lower for substances with logKOW >3.5 than for substances with logKOW <3.5, as also alluded to by Natsch et al. (2021). In such cases, the uncertainty related to negative predictions needs to be thoroughly evaluated in the given context, i.e. hazard or risk assessment, and should be appropriately accounted for, e.g. by preferring NAMs without such a correlation, e.g. the h-CLAT, U-SENS™ and SENS-IS.

Except for the SENS-IS, possibly due to the use of a 3D reconstituted human skin with some skin barrier, all NAMs predicted most set 1 substances as sensitizers. Compared to the LLNA, this resulted in a high overprediction rate leading to reduced balanced accuracies when compared to set 2. When focusing on the 22 dyes in set 1, which included seven human and seven LLNA negatives, the highest specificity ($4/7 = 57\%$) was observed for the KeratinoSens™, when compared with human

data. In addition, we observed that three surfactants without human data, which were GPMT negative and have been discussed to be LLNA false positives (Ball et al., 2011), were positive in the h-CLAT and U-SENS™, but negative in the other NAMs. All these differences are likely due to several factors, such as technical applicability issues, LLNA-anchored predictions model development favouring sensitivity, remaining uncertainty in the reference results, both human and LLNA for some substances, and a combination thereof. In addition, and most impactful, these results suggest that insufficient inclusion of more recently designed chemistry in the development and optimisation may have led to an assessment of the NAMs that can, in general, not readily be extrapolated beyond the chemistry tested during development and optimisation. It would be premature to conclude that NAM are not fit for testing such chemistry, but results in a higher uncertainty associated with the NAM results and questions the rigid approach of fixed prediction models for all types of chemistry and all types of purposes. In terms of the NAM applicability domains, our results strongly suggest to ensure that NAMs produce reliable results for the type of chemistry investigated. If unknown, targeted experimental studies may need to be conducted to generate the data required to address this aspect, as e.g., also done by Petry et al. (2017).

These uncertainties add to the general uncertainty related to the NAMs: while there is high confidence in the qualitative skin sensitisation AOP (see <https://aopwiki.org/aops/40>), the uncertainty in how well the NAMs model the KE is less well assessed. With the aim to increase our

understanding of interdependence of test method results, which informs their strategic combination, we therefore compared test methods in detail. Our findings confirmed that similar test methods addressing the same KE, i.e., h-CLAT and U-SENS™ as well as the DPRA and PPRA, are the most concordant, i.e., approx. for 75% of the substances. Discordant cases can be partly explained by differences between the NAMs, such as the increased ability of the PPRA to detect pre-/pro-haptens, and by borderline results (Leontaridou et al., 2019). This suggests that h-CLAT and U-SENS™ might be similar enough to be used interchangeably or as alternatives in DAs. Consequently, the high mutual dependence of the NAMs strongly suggests that inclusion of both h-CLAT and U-SENS™ in a DA is of little added value for its predictive performance. Similarly, a DA that uses the less resource intensive DPRA as a default and the PPRA in cases, when activation of a substance is suspected or a substance co-elutes, appears to be a promising approach. The PPRA could also be used to generate additional information outside a DA, e.g., to demonstrate that substances are not activated.

While prescriptive approaches are needed for some regulatory purposes, e.g., hazard and UN GHS potency classifications, flexibility is needed when assessing skin sensitisation risk to be able to account for substance- and exposure-specific needs. Interchangeability of NAMs in DAs and targeted use of NAM information may contribute to it, but also the freedom to select a suitable DA. Therefore, we proposed an integrative next generation risk assessment (NGRA) framework for skin sensitisation that offers the necessary flexibility for a comprehensive safety evaluation of novel cosmetic ingredients (Gilmour et al., 2020). We will continue to evaluate and evolve the NGRA by case studies to explore how to better address uncertainty and the use of additional information, such as read across information from other NAM data, within a weight of evidence approach.

In conclusion, the testing of non-standard substances revealed that the predictive value of skin sensitisation NAMs associated with standard prediction models was limited. This supports the call for a more detailed understanding of how NAMs apply to a wider range of substances (Basketter et al., 2019). Flexible and informed use of the NAMs is needed, including adjustment of current protocols. Optimal choice of NAMs, either to be used within a DA or to inform other steps in a NGRA, e.g., to support read-across, will best inform the skin sensitisation risk assessment and support the characterisation and reduction of uncertainty.

Funding source

This work was funded through the Long Range Science Strategy (LRSS) programme of Cosmetics Europe (available at <https://www.lrsscosmeticseurope.eu>).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Sebastian Hoffmann and Erwin von Vliet report financial support was provided by Cosmetics Europe.

Acknowledgement

The authors would like to acknowledge the contribution of additional experts (in alphabetical order): Morihiko Hirota (Shiseido), Lily Mijouin (Shiseido) and Dirk Petersohn (Henkel).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.yrtph.2022.105169>.

References

- Alepee, N., Piroird, C., Aujoulat, M., Dreyfuss, S., Hoffmann, S., et al., 2015. Prospective multicentre study of the U-SENS test method for skin sensitization testing. *Toxicol. Vitro* 30, 373–382.
- Api, A.M., Parakhia, R., O'Brien, D., Basketter, D.A., 2017. Fragrances categorized according to relative human skin sensitization potency. *Dermatitis* 28, 299–307.
- Ashikaga, T., Sakaguchi, H., Sono, S., Kosaka, N., Ishikawa, M., 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, 275–284.
- Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., et al., 2006. Development of an in vitro skin sensitization test using human cell lines: the human Cell Line Activation Test (h-CLAT). I. Optimization of the h-CLAT protocol. *Toxicol. Vitro* 20, 767–773.
- Ball, N., Cagen, S., Carrillo, J.C., Certa, H., Eigler, D., et al., 2011. Evaluating the sensitization potential of surfactants: integrating data from the local lymph node assay, Guinea pig maximization test, and in vitro methods in a weight-of-evidence approach. *Regul. Toxicol. Pharmacol.* 60, 389–400.
- Basketter, D., Azam, P., Casati, S., Corvaro, M., Ezendam, J., et al., 2019. Applying non-animal strategies for assessing skin sensitisation report from an EPAA/cefic-LRI/IFRA Europe cross sector workshop, ECHA Helsinki, February 7th and 8th 2019. *Regul. Toxicol. Pharmacol.* 109, 104477.
- Basketter, D.A., Alepee, N., Ashikaga, T., Barroso, J., Gilmour, N., et al., 2014. Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis* 25, 11–21.
- Basketter, D.A., Cadby, P., 2004. Reproducible prediction of contact allergenic potency using the local lymph node assay. *Contact Dermatitis* 50, 15–17.
- Basketter, D.A., Evans, P., Fielder, R.J., Gerberick, G.F., Dearman, R.J., Kimber, I., 2002. Local lymph node assay - validation, conduct and use in practice. *Food Chem. Toxicol.* 40, 593–598.
- Bergal, M., Puginier, M., Gerbeix, C., Groux, H., Roso, A., et al., 2020. In vitro testing strategy for assessing the skin sensitizing potential of "difficult to test" cosmetic ingredients. *Toxicol. Vitro* 65, 104781.
- Buehler, E.V., 1965. Delayed contact hypersensitivity IN the Guinea pig. *Arch. Dermatol.* 91, 171–177.
- Cottrez, F., Boitel, E., Ourlin, J.C., Peiffer, J.L., Fabre, I., et al., 2016. SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitization potency: reproducibility and predictivity results from an inter-laboratory study. *Toxicol. Vitro* 32, 248–260.
- Daniel, A.B., Strickland, J., Allen, D., Casati, S., Zuang, V., et al., 2018. International regulatory requirements for skin sensitization testing. *Regul. Toxicol. Pharmacol.* 95, 52–65.
- De Jong, W.H., De Klerk, A., Beek, M.T., Veenman, C., Van Loveren, H., 2007. Effect of prolonged repeated exposure to formaldehyde donors with doses below the EC3 value on draining lymph node responses. *J. Immunot.* 4, 239–246.
- Del Bufalo, A., Pauloin, T., Alepee, N., Clouzeau, J., Detroyer, A., et al., 2018. Alternative integrated testing for skin sensitization: assuring consumer safety. *Appl. In Vitro Toxicol.* 4, 30–43.
- Desprez, B., Adriaens, E., Alépée, N., 2021. Real world-like simulations show efficient predictive power of in vitro skin corrosion tests used as stand-alone and in combination and how can toxicologists take advantage of them. *Toxicol. Vitro* 70, 105043.
- Dionisio, K.L., Frame, A.M., Goldsmith, M.-R., Wambaugh, J.F., Liddell, A., et al., 2015. Exploring consumer exposure pathways and patterns of use for chemicals in the environment. *Toxicol Rep* 2, 228–237.
- Dumont, C., Barroso, J., Matys, L., 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. Vitro* 34, 220–228.
- Ec, 2009. Regulation (EC) No 1223/2009 of the European parliament and the council of 30 November 2009 on cosmetic products. In: E Commission. Official Journal of the European Union. L342:59–209.
- Ezendam, J., Braakhuis, H.M., Vandebriel, R.J., 2016. State of the art in non-animal approaches for skin sensitization testing: from individual test methods towards testing strategies. *Arch. Toxicol.* 90, 2861–2883.
- Fujita, M., Yamamoto, Y., Watanabe, S., et al., 2019a. The within- and between-laboratory reproducibility and predictive capacity of the in chemico amino acid derivative reactivity assay: results of validation study implemented in four participating laboratories. *J. Appl. Toxicol.* 39 (11), 1492–1505. <https://doi.org/10.1002/jat.3834>.
- Fujita, M., Yamamoto, Y., Wanibuchi, S., Katsuoka, Y., Kasahara, T., 2019b. The underlying factors that explain why nucleophilic reagents rarely co-elute with test chemicals in the ADRA. *J. Pharmacol. Toxicol. Methods* 96, 95–105.
- 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. *Regul. Toxicol. Pharmacol.* 117, 104755.
- Gerberick, G.F., Ryan, C.A., Kimber, I., Dearman, R.J., Lea, L.J., Basketter, D.A., 2000. Local lymph node assay: validation assessment for regulatory purposes. *Am. J. Contact Dermatitis* 11, 3–18.
- Gerberick, G.F., Troutman, J.A., Foertsch, L.M., Vassallo, J.D., Quijano, M., et al., 2009. Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. *Toxicol. Sci.* 112, 164–174.

- Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., Lepoittevin, J.P., 2004. Development of a peptide reactivity assay for screening contact allergens. *Toxicol. Sci.* 81, 332–343.
- Gilmour, N., Kern, P.S., Alépée, N., Boisléve, F., Bury, D., et al., 2020. Development of a next generation risk assessment framework for the evaluation of skin sensitisation of cosmetic ingredients. *Regul. Toxicol. Pharmacol.* 104721.
- Hirota, M., Ashikaga, T., Kouzuki, H., 2018. Development of an artificial neural network model for risk assessment of skin sensitization using human cell line activation test, direct peptide reactivity assay, KeratinoSens™ and in silico structure alert parameter. *J. Appl. Toxicol.* 38, 514–526.
- Hoffmann, S., 2015. LLNA variability: an essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. *ALTEX* 32, 379–383.
- Hoffmann, S., Hartung, T., 2005. Diagnosis: toxic!—trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations. *Toxicol. Sci.* 85, 422–428.
- Hoffmann, S., Kleinstreuer, N., Alépée, N., Allen, D., Api, A.M., et al., 2018. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. *Crit. Rev. Toxicol.* 48, 344–358.
- ICCVAM, 2011. ICCVAM Test Method Evaluation Report: Usefulness and Limitations of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans. National Institute of Environmental Health Sciences, Research Triangle Park, NC.
- Jaworska, J.S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., Miyazawa, M., 2015. Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. *Arch. Toxicol.* 89, 2355–2383.
- Jowsey, L.R., Basketter, D.A., Westmoreland, C., Kimber, I., 2006. A future approach to measuring relative skin sensitizing potency: a proposal. *J. Appl. Toxicol.* 26, 341–350.
- Karlsson, I., Vanden Broecke, K., Martensson, J., Goossens, A., Borje, A., 2011. Clinical and experimental studies of octocrylene's allergenic potency. *Contact Dermatitis* 64, 343–352.
- Kimura, Y., Fujimura, C., Ito, Y., Takahashi, T., Nakajima, Y., et al., 2015. Optimization of the IL-8 Luc assay as an in vitro test for skin sensitization. *Toxicol. Vitro* 29, 1816–1830.
- Kleinstreuer, N.C., Hoffmann, S., Alepee, N., Allen, D., Ashikaga, T., et al., 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches (*). *Crit. Rev. Toxicol.* 48, 359–374.
- Leontaridou, M., Gabbert, S., Landsiedel, R., 2019. The impact of precision uncertainty on predictive accuracy metrics of non-animal testing methods. *ALTEX* 36, 435–446.
- Loveless, S.E., Api, A.M., Crevel, R.W., Debruyne, E., Gamer, A., et al., 2010. Potency values from the local lymph node assay: application to classification, labelling and risk assessment. *Regul. Toxicol. Pharmacol.* 56, 54–66.
- Magnusson, B., Kligman, A.M., 1969. The identification of contact allergens by animal assay. *The Guinea pig maximization test*. *J. Invest. Dermatol.* 52, 268–276.
- Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. *J. Cheminf.* 10, 10.
- Mehling, A., Adriaens, E., Casati, S., Hubsch, B., Irizar, A., et al., 2019. In vitro RHE skin sensitisation assays: applicability to challenging substances. *Regul. Toxicol. Pharmacol.* 108, 104473.
- Natsch, A., Emter, R., 2008. Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitization potential of chemicals. *Toxicol. Sci.* 102, 110–119.
- Natsch, A., Haupt, T., Wareing, B., Landsiedel, R., Kolle, S.N., 2020. Predictivity of the kinetic direct peptide reactivity assay (KDPR) for sensitizer potency assessment and GHS subclassification. *ALTEX* 37 (4), 652–664. <https://doi.org/10.14573/altex.2004292>.
- Natsch, A., Landsiedel, R., Kolle, S.N., 2021. A triangular approach for the validation of new approach methods for skin sensitization. *ALTEX* 38 (4), 669–677. <https://doi.org/10.14573/altex.2105111>.
- 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, 2014. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins, OECD Series on Testing and Assessment. OECD Publishing, Paris. <https://doi.org/10.1787/9789264221444-en>. No. 168.
- OECD, 2018a. Test No. 442D. In: Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. <https://doi.org/10.1787/9789264229822-en>.
- OECD, 2018b. Test No. 442E. In: Vitro Skin Sensitisation: in Vitro Skin Sensitisation Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. <https://doi.org/10.1787/9789264264359-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>.
- OECD, 2021c. Supporting document to the OECD Guideline 497 on defined approaches for skin sensitisation. Series on Testing and Assessment, No. 336. Available at: [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/CBC/MONO\(2021\)11&docLanguage=En](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/CBC/MONO(2021)11&docLanguage=En). (Accessed 24 February 2022).
- Patlewicz, G., Casati, S., Basketter, D.A., Asturiol, D., Roberts, D.W., et al., 2016. Can currently available non-animal methods detect pre and pro-haptens relevant for skin sensitization? *Regul. Toxicol. Pharmacol.* 82, 147–155.
- Patlewicz, G., Jeliakova, N., Safford, R.J., Worth, A.P., Aleksiev, B., 2008. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ. Res.* 19, 495–524.
- Petry, T., Bosch, A., Coste, X., et al., 2017. Evaluation of in vitro assays for the assessment of the skin sensitization hazard of functional polysiloxanes and silanes. *Regul. Toxicol. Pharmacol.* 84, 64–76. <https://doi.org/10.1016/j.yrtph.2016.12.009>.
- Piroid, C., Ovigne, J.M., Rousset, F., Martinozzi-Teissier, S., Gomes, C., et al., 2015. The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. *Toxicol. Vitro* 29, 901–916.
- Python, F., Goebel, C., Aebly, P., 2007. Assessment of the U937 cell line for the detection of contact allergens. *Toxicol. Appl. Pharmacol.* 220, 113–124.
- Ramirez, T., Stein, N., Aumann, A., Remus, T., Edwards, A., et al., 2016. Intra- and inter-laboratory reproducibility and accuracy of the LuSens assay: a reporter gene-cell line to detect keratinocyte activation by skin sensitizers. *Toxicol. Vitro* 32, 278–286.
- Reisinger, K., Hoffmann, S., Alepee, N., Ashikaga, T., Barroso, J., et al., 2015. Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. *Toxicol. Vitro: Int. J. Pub. Ass.BIBRA* 29, 259–270.
- Roberts, D.W., Patlewicz, G., Kern, P.S., Gerberick, F., Kimber, I., et al., 2007. Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chem. Res. Toxicol.* 20, 1019–1030.
- Ryan, C.A., Chaney, J.G., Gerberick, G.F., Kern, P.S., Dearman, R.J., et al., 2007. Extrapolating local lymph node assay EC3 values to estimate relative sensitizing potency. *Cutan. Ocul. Toxicol.* 26, 135–145.
- Ryan, C.A., Troutman, J.A., Kern, P.S., Quijano, M., Dobson, R.L.M., et al., 2020. Refinement of the Peroxidase Peptide Reactivity Assay and Prediction Model for Assessing Skin Sensitization Potential. *Toxicological Sciences*.
- SCCS, 2021. The SCCS notes of guidance for the testing of cosmetic ingredients and their safety evaluation. 11th revision. Available at: https://ec.europa.eu/health/sites/default/files/scientific_committees/consumer_safety/docs/sccs_o_250.pdf.
- Takenouchi, O., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., et al., 2015. Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *J. Appl. Toxicol.* 35, 1318–1332.
- Takenouchi, O., Miyazawa, M., Saito, K., Ashikaga, T., Sakaguchi, H., 2013. Predictive performance of the human Cell Line Activation Test (h-CLAT) for lipophilic chemicals with high octanol-water partition coefficients. *J. Toxicol. Sci.* 38, 599–609.
- Troutman, J.A., Foertsch, L.M., Kern, P.S., Dai, H.J., Quijano, M., et al., 2011. The incorporation of lysine into the peroxidase peptide reactivity assay for skin sensitization assessments. *Toxicol. Sci.* 122, 422–436.
- Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., et al., 2015. Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul. Toxicol. Pharmacol.* 71, 337–351.
- van Loveren, H., Cockshott, A., Gebel, T., Gundert-Remy, U., de Jong, W.H., et al., 2008. Skin sensitization in chemical risk assessment: report of a WHO/IPCS international workshop focusing on dose-response assessment. *Regul. Toxicol. Pharmacol.* 50, 155–199.
- Wickham, W., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
- Zang, Q., Mansouri, K., Williams, A.J., Judson, R.S., Allen, D.G., et al., 2017. In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. *J. Chem. Inf. Model.* 57, 36–49.