

Research Article

Exploration of the GARDskin Applicability Domain: Indirectly Acting Haptens, Hydrophobic Substances and UVCBs

Andy Forreryd¹, Robin Gradin¹, Charles Humfrey², Len Sweet³ and Henrik Johansson¹

¹SenzaGen AB, Lund, Sweden; ²The Lubrizol Corporation, Derby, UK; ³The Lubrizol Corporation, Wickliffe, OH, USA

Abstract

Hazard assessments of skin sensitizers are increasingly being performed using new approach methodologies (NAMs), with several *in chemico*, *in vitro* and most recently also defined approaches (DAs) being accepted for regulatory use. However, keeping track of potential limitations of each method in order to define applicability domains remains a crucial component to ensure adequate predictivity as well as facilitating the appropriate selection of method(s) for each hazard assessment task. The objective of this report is to share test results generated with the GARDTMskin assay on chemicals that have traditionally been considered as difficult to test in some of the conventional *in vitro* and *in chemico* OECD Test Guidelines for skin sensitization. Such compounds may include, for example, indirectly acting haptens, hydrophobic substances, and substances of unknown variable composition or biological substances (UVCBs). Based on the results of this study, the sensitivity for prediction of skin sensitizing hazard of indirectly acting haptens was 92.4% and 87.5%, when compared with LLNA (n=25) and human data (n=8), respectively. Similarly, the sensitivity for prediction of skin sensitizing hazard of hydrophobic substances was 85.1% and 100%, when compared with LLNA (n=24) and human data (n=9), respectively. Lastly, a case study involving assessment of a set of hydrophobic UVCBs (n=7) resulted in a sensitivity of 100, as compared to available reference data. Thus, it was concluded that these data provide support for the inclusion of such chemistries in the GARDTMskin applicability domain, without an increased risk of false negative classifications.

1 Introduction

A chemical substance able to directly or indirectly act as a hapten and induce allergic contact dermatitis (ACD) by the process of skin sensitization is referred to as a skin sensitizer (UN GHS, 2015). The key biological events underlying skin sensitization by small organic molecules have been extensively studied and the existing knowledge has been summarized in the form of an adverse outcome pathway (AOP) (OECD, 2014).

Proactive hazard classification and characterization of skin sensitizers is an important aspect of risk assessment of chemicals, a task which is increasingly being performed by the use of so-called new approach methodologies (NAM), including e.g., *in silico*, *in chemico* and *in vitro* methods. Several such methods have gained regulatory acceptance as OECD TGs and provide acceptable data to support classification (OECD, 2015a,b, 2017a) and subcategorization (OECD, 2021a) in the context of integrated approaches to testing and assessment (IATA) for several chemical classes.

To further facilitate regulatory uptake in specific industry sectors, and to provide guidance for end-users attempting to select the most appropriate assay for their specific chemistry, careful characterization of the applicability domain of such assays is pivotal in providing confidence in classification outcomes. To date, it is recognized that a number of substances, for various reasons, either remain difficult to accurately assess or belong to a chemical space that has hitherto not been thoroughly explored in the existing validated NAMs (Mehling et al. 2019). Such limitations, as far as they have been identified, are incorporated into the individual TGs, and may include for example hydrophobic substances which cannot be tested at sufficiently high concentrations in submerged cell systems, or indirectly acting haptens that are not inherently protein reactive but would require abiotic or biotic activation to initiate the molecular initiating event of skin sensitization. In addition, while testing of complex or undefined test items such as mixtures, formulations and UVCBs may be technically compatible with test methods, they may require customized alterations of protocols, and the appropriate interpretation of the test outcome may not always be straight-forward.

Disclaimer: In relation to CH and LS, the views expressed in this article are the professional opinions of the authors and do not necessarily reflect those of The Lubrizol Corporation.

Received January 28, 2022; Accepted April 13, 2022;
Epub April 21, 2022; © The Authors, 2022.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.

ALTEX 39(##), ###-###. doi:10.14573/altex.2201281

Correspondence: Henrik Johansson, PhD
SenzaGen AB, Bldg 401, Medicon Village, SE-22381 Lund, Sweden
(henrik.johansson@senzagen.com)

The Genomic Allergen Rapid Detection™ (GARD™) assay for assessment of skin sensitizers (GARD™_{skin}) is a NAM-based predictive method addressing KE3 of the AOP. The method is based on test chemical exposure of a surrogate *in vitro* DC-like cell line, followed by quantification of gene expression patterns of endpoint-specific genomic biomarkers. The quantified levels of transcription of the genomic biomarkers are then used for classification of the test chemical with the aid of a machine learning-based prediction algorithm (Forreryd et al., 2016, Johansson et al., 2011, 2019). The GARD™_{skin} method has been validated and reviewed by the EU Reference Laboratory for alternatives to animal testing (EURL ECVAM) Scientific Advisory Committee (ESAC) (Corsini et al., 2021) and is currently progressing towards adoption as an OECD TG. Therefore, the appropriate monitoring of potential limitations, such as those listed above, is a crucial aspect of defining the method's applicability domain.

This report describes retrospective analyses of available GARD™_{skin} data, as well as previously unpublished data, aiming to explore the method applicability in the domains of indirectly acting haptens, hydrophobic substances, and UVCBs.

2 Material and methods

Selection of chemicals and dataset composition

Available GARD™_{skin} data (Corsini et al., 2021, Forreryd et al., 2016, Johansson et al., 2017, 2019), as well as previously unpublished data, were mined for test chemicals identified as indirectly acting haptens in the curated reference dataset compiled by the OECD Expert Group for Defined Approaches for Skin Sensitization (OECD, 2021b). Similarly, the curated reference dataset compiled by the OECD Expert Group for Defined Approaches for Skin Sensitization (OECD, 2021b) was used to identify hydrophobic substances, as defined based on their log P value. Each union set of data was amended with reference classifications based on LLNA and human data extracted from the above-mentioned dataset.

Taken together, a total of 25 indirectly acting haptens with available GARD™_{skin} data and reference classifications could be identified. Annotations as indirectly acting haptens, as well as a systematic assignment as either pro- and/or pre-haptens, were based on the proposed assignments in the respective above-mentioned publications, when applicable. Similarly, a total of 25 hydrophobic substances with available GARD™_{skin} data and reference classifications could be identified.

The performance of the GARD™_{skin} assay was also evaluated on a subset of complex and hydrophobic UVCB substances. A total of seven test items were kindly provided by The Lubrizol Corporation. Approximate molecular weights were available for all substances, except for LUB-4. All test items had either proprietary *in vivo* reference data or human data from human repeated insult patch testing (HRIPT) available.

The identities of each test chemical included in the analyses of this report, along with relevant physico-chemical properties, reference classifications and GARD™_{skin} classifications, are listed in Tables 1-3.

Generation of historical GARD™_{skin} data

The available GARD™_{skin} data for all studies in this report were generated according to the validated GARD™_{skin} assay protocol (EURL ECVAM, 2021) and in compliance with the Draft GARD_{skin} OECD TG. While some of the historical data (Forreryd 2016, Johansson et al., 2017) were generated prior to the establishment of the current GARD™_{skin} assay protocol and the drafting of the GARD™_{skin} OECD TG, they were acquired using identical cellular protocols as described in these documents. Of note, however, the data analysis pipeline was updated with the implementation of a Batch Adjustment by Reference Alignment (BARA) pre-processing and normalization procedure (Gradin et al., 2019), prior to the finalization of the Standard Operating Procedure (SOP) used for method validation. For the purpose of this work, the historical data from the cellular exposure experiments were reanalyzed using the validated and updated data analysis pipeline, in order to generate coherent datasets that were in complete concordance with the GARD™_{skin} assay protocol and the Draft GARD™_{skin} OECD TG.

Generation of novel GARD™_{skin} data

All novel GARD™_{skin} data of studies presented in this report were generated by experiments carried out at SenzaGen's GLP compliant laboratory. When applicable, commercially available test chemicals were purchased from Sigma Aldrich (St. Louis, Missouri). With the exception of the Lubrizol study for testing of UVCB samples, data were collected from experiments conducted according to the GARD™_{skin} assay protocol (EURL ECVAM, 2021) and in compliance with the Draft GARD™_{skin} OECD TG. Data from the Lubrizol study for testing of UVCB substances were generated according to the GARD™_{skin} assay protocol and the GARD™_{skin} draft OECD TG, with two deviations motivated by the solubility properties and complex nature of the test items. Firstly, due to the limited solubility and the complexity of the samples, alternative vehicles, previously not used for method validation, were explored. In addition to DMSO, a mixture of dimethyl formamide (DMF)/glycerol (4/1 (vol%)), LUB-2, -5 and -6) and xylene (LUB-4) were utilized in the study based on expert input from Lubrizol, both at a final in-well concentration of 0.1%. The experimental vehicles were included as additional negative controls at corresponding in-well concentrations. Secondly, the complex nature of the test items motivated approximations of the appropriate molecular weights used for calculations of concentrations. Here, weighted mean molecular weights were approximated, taking into account the relative concentrations of each component of the multiconstituent test items. For one test item, LUB-4, no information regarding the molecular weights or relative concentration of components were available. Based on the approximated molecular weights of similar substances in the study, and in a conservative approach to ensure concentration was sufficiently high, the molecular weight for this test item was approximated to 2000 g/mol.

All test results included in the herein summarized studies met the defined acceptance criteria of the method and were based on a minimum of three replicate biological samples, as defined and described (EURL ECVAM, 2021).

Tab. 1: GARD™ skin assessment of indirectly acting haptens

Test chemical descriptors			Reference classifications ¹		GARD™ classifications	
Name	CASRN	Mechanistic domain ¹	LLNA	Human	#GARD™ studies	Weighted classification
2-Aminophenol	95-55-6	pre	1	NA	5 ^{3,4,6}	1
2-Nitro-p-phenylenediamine	5307-14-2	pre/pro	1	NA	7 ^{3,4,6}	1
2,5-Diaminotoluene sulfate	615-50-9	pre	1	NA	11 ^{3,5,6}	1
3-(Dimethylamino)propylamine	109-55-7	pro	1	NA	11 ^{3,5}	1
3-Aminophenol	591-27-5	pro-MA	1	NA	1 ⁶	1
4-Amino-m-cresol	2835-99-6	pre/pro-MA ²	1	NA	3 ⁷	1
Abietic acid	514-10-3	pre	1	NA	1 ³	1
Aniline	62-53-3	pre/pro	1	1	1 ³	0
Bromothalonil	35691-65-7	pre	1	NA	7 ⁵	1
Chlorpromazine	50-53-3	pre/pro	1	1	1 ³	1
Cinnamic alcohol	104-54-1	pre/pro	1	1	10 ^{3,5}	1
D-Limonene	5989-27-5	pre	1	NA	1 ³	1
Dihydroeugenol	2785-87-7	pre/pro-MA ²	1	NA	3 ⁷	1
Ethylene diamine (free base)	107-15-3	pro	1	NA	10 ^{3,5}	0,1
Eugenol	97-53-0	pre/pro	1	1	10 ^{3,5}	1
Farnesol	4602-84-0	pre/pro	1	1	2 ³	1
Geraniol	106-24-1	pre/pro	1	1	5 ^{3,4,5}	1
Hydroquinone	123-31-9	pre	1	NA	1 ⁴	1
Isoeugenol	97-54-1	pre	1	1	11 ^{3,4,5}	1
Lauryl gallate	1166-52-5	pre	1	NA	10 ^{4,5}	1
Linalool	78-70-6	pre	1	NA	1 ³	1
Metol	55-55-0	pre/pro	1	NA	9 ⁵	1
p-Phenylenediamine ⁸	106-50-3	pre	1	1	20 ^{3,4,5,6,7}	1
Propyl gallate	121-79-9	pre	1	NA	10 ^{3,5}	1
Resorcinol	108-46-3	pre	1	NA	6 ^{3,4,6}	1

¹Values extracted from OECD, 2021b, unless otherwise indicated. Sensitizer and non-sensitizer classifications are denoted 1 and 0, respectively. pre; chemical primarily acting as a pre-hapten. pro; chemical primarily acting as a pro-hapten. pre/pro; chemical acting as both pre- and pro-hapten, or otherwise not able to be subcategorized. MA; chemical primarily reacting as Michael-acceptor. NA; missing value.

²Values extracted from Urbisch et al., 2016.

³Raw data originating from Johansson et al., 2017.

⁴Raw data originating from Forreryd et al., 2016.

⁵Raw data originating from Johansson et al., 2019.

⁶Previously unpublished data.

⁷Previously unpublished data, included in ESAC peer-review (Corsini et al., 2021).

⁸Utilized as GARDskin positive control (EURL ECVAM, 2021).

For additional details regarding GARDskin experimental parameters for each test chemical included in the studies, please refer to Tables S1-S3¹.

Statistics

The predictive performance of GARDskin in each chemical domain was described by Cooper Statistics (Cooper et al., 1979). As certain test chemicals had been assayed in more than one historical study, the results were based on weighted calculations, hindering individual chemicals with multiple test results from biasing the summarizing statistics. For example, benzyl salicylate (Table 2) had been assayed in four independent GARDskin studies, three of which resulted in a non-sensitizer classification and one resulting in a sensitizer classification. The summarized classification used for calculation of Cooper Statistics was therefore reported as 0.25, where the integer 1 denotes a sensitizer-classification and the integer 0 denotes a non-sensitizer classification. This weighted approach was adopted from the methodology that is used by the OECD when developing test guidelines (OECD, 2015c, 2017b) and utilized by the ESAC during the peer-review of the GARD™ skin method (Corsini et al., 2021).

3 Results

The applicability domain of GARD™ skin was evaluated with a specific focus on indirectly acting haptens and hydrophobic substances by combining historical and novel data generated from each domain. In addition, the GARD™ skin functionality in

¹ doi:10.14573/altex.2201281s

Tab. 2: GARD™ skin assessment of hydrophobic substances

Test chemical descriptors			Reference classifications ¹		GARD™ classifications	
Name	CASRN	LogP ¹	LLNA	Human	#GARD™ studies	Weighted classification
Abietic acid	514-10-3	3.92	1	NA	1 ²	1
alpha-Amylcinnamaldehyde	122-40-7	3.99	1	NA	1 ²	1
alpha-Amylcinnamic alcohol	101-85-9	3.68	1	NA	1 ²	1
alpha-Hexylcinnamaldehyde	101-86-0	4.34	1	NA	3 ^{2,6}	1
alpha-Tocopherol	59-02-9	9.41	1	NA	1 ²	0
BADGE	1675-54-3	3.36	1	NA	1 ²	1
Benzyl benzoate	120-51-4	3.97	1	0	10 ^{2,3}	0,5
Benzyl cinnamate	103-41-3	3.62	1	NA	3 ⁴	0,67
Benzyl salicylate	118-58-1	3.11	1	NA	4 ^{2,4}	0,25
Chlorpromazine	50-53-3	5.41	1	1	1 ²	1
Citral	5392-40-5	3.00	1	1	7 ^{2,3}	1
Citronellol	106-22-9	3.91	1	0	1 ²	1
D-Limonene	5989-27-5	4.51	1	NA	1 ²	1
Dibenzoyl peroxide	94-36-0	3.46	1	1	4 ^{2,4}	1
Farnesol	4602-84-0	4.91	1	1	2 ²	1
Geraniol	106-24-1	3.56	1	1	5 ^{2,5,6}	1
Hexyl salicylate	6259-76-3	5.50	1	0	1 ²	1
Isopropyl myristate	110-27-0	6.90	1	NA	1 ⁶	0
Lauryl gallate	1166-52-5	4.64	1	NA	10 ^{3,5}	1
Lilial	80-54-6	3.94	1	1	1 ²	1
Methyl oct-2-ynoate	111-12-6	2.80	1	1	1 ²	1
n-Hexane	110-54-3	3.90	0	0	10 ^{3,5}	0
Pentachlorophenol	87-86-5	5.12	1	NA	1 ²	1
Phenyl benzoate	93-99-2	3.59	1	1	1 ²	1
Tetrachlorosalicylanilide	1154-59-2	5.11	1	1	4 ^{2,4}	1

¹Values extracted from OECD, 2021b. Sensitizer and non-sensitizer classifications are denoted 1 and 0, respectively. NA; missing value.

²Raw data originating from Johansson et al., 2017.

³Raw data originating from Johansson et al., 2019.

⁴Previously unpublished data, included in ESAC peer-review (Corsini et al., 2021).

⁵Raw data originating from Forreryd et al., 2016.

⁶Previously unpublished data.

Tab. 3: GARD™ skin assessment of UVCBs

Test chemical descriptors			Reference classifications ²		GARD™ classifications
Name	CASRN	LogP ¹	WoE	Data source(s)	
LUB-1	-	11.7	1	LLNA/Buehler	1
LUB-2	-	2.7	1	LLNA/Buehler	1
LUB-3	-	4.5	1	LLNA	1
LUB-4	-	10.88	0	Buehler/HRIPT	0
LUB-5	-	10.88	1	LLNA/Buehler/HRIPT	1
LUB-6	-	3.69	1	Buehler	1
LUB-7	-	4.13	0	LLNA	1

¹Values based on proprietary Lubrizol data.

²Values based on Lubrizol expert judgement, using weight of evidence (WoE) based on individually indicated data sources. Sensitizer and non-sensitizer classifications are denoted 1 and 0, respectively.

the domain of hydrophobic substances was further supported by the assessment of a set of UVCBs, specifically selected due to their high log P values and expected incompatibility with submerged cell systems.

The individual test results are presented in Tables 1, 2 and 3, for indirectly acting haptens, hydrophobic substances and UVCBs, respectively. The performance statistics from the studies involving indirectly acting haptens and hydrophobic substances are summarized in Table 4. Lastly, individual test results for each test chemical, including a graphical representation of obtained GARD™ skin Decision Values (DV) from each individual replicate sample from each included study is compiled in Supplementary Materials.

It was concluded that the indirectly acting haptens included in investigated studies were concordant with LLNA and human classifications (OECD, 2021b) in 92.4% and 87.5% of the cases, respectively. Similarly, it was concluded that the

hydrophobic substances included in investigated studies were concordant with LLNA and human classifications in 85.7% and 80.8% of the cases, respectively. Lastly, GARDTMskin results of hydrophobic UVCBs were 85.7% concordant with proprietary *in vivo* and human reference classifications.

Taken together, it was found that these estimates of performance are in line with previously reported estimates of GARDTMskin's accuracy, which are typically in the range of ~80-95%, depending on the chemical subset when considering a wide chemical space. Therefore, these data support the inclusion of indirectly acting haptens and hydrophobic substances in the GARDTMskin applicability domain.

Tab. 4: Confusion matrices and performance metrics

Dataset		Indirectly acting haptens				Hydrophobic substances			
Reference		LLNA ¹		Human ¹		LLNA ¹		Human ¹	
Reference classifications		NS (0)	S (25)	NS (0)	S (8)	NS (1)	S (24)	NS (4)	S (9)
GARD TM skin classifications	NS	0	1.90	0	1	1	3.58	1.50	0
	S	0	23.1	0	7	0	20.4	2.50	9
Accuracy		92.4%		87.5%		85.7%		80.8%	
Sensitivity		92.4%		87.5%		85.1%		100%	
Specificity		-		-		100%		37.5%	
Balanced accuracy		-		-		92.5%		68.8%	

¹Values extracted from OECD, 2021b.

4 Discussion

The field of *in vitro* toxicology has seen a great surge of innovation in the last decade, allowing NAMs to become a new normal when performing hazard assessment of potential chemical skin sensitizers. Several methods have been validated and adopted into OECD TGs, with the most recent addition of TG 497, providing a guideline for the use of several DAs for skin sensitization (OECD, 2021a). Still, regardless of the regulatory context, or if DAs or stand-alone screening methods are used in non-regulatory settings, the appropriate selection of methods used for assessment of specific test chemicals is highly dependent on an understanding of the applicability domain of each test method. To this end, this report describes the use of the GARDTMskin assay for hazard assessment of indirectly acting haptens and hydrophobic substances, both of which are known to be potentially incompatible with individual methods (Mehling et al., 2019). Thus, the major aim of this report was to provide complementary GARDTMskin data to support the inclusion of the above-mentioned chemical space in the applicability domain of the GARDTMskin method, and to serve as a guidance for end users in the selection of the most appropriate assay for their specific chemistry.

Indirectly acting haptens, which are not inherently electrophilic but require either abiotic or biotic activation to gain peptide reactivity, was initially considered to be outside the applicability domain of currently OECD adopted *in vitro* assays for skin sensitization. However, studies have demonstrated that the majority of such compounds can be accurately detected in at least one of these assays, but with important differences in efficacy between individual assays (Patlewicz et al., 2016, Urbisch et al., 2016). Based on the results presented in this report, the sensitivity of GARDTMskin for prediction of skin sensitizing hazard of indirectly acting haptens was 92.4% (23.1/25.0) compared with LLNA references, and 87.5% (7/8) compared with human data. Furthermore, attempts were made to specifically characterize the subgroup of indirectly haptens requiring metabolic activation (pro-haptens), since the metabolic capacity of the *in vitro* cell system has currently not been fully characterized. The available data included a subset of at least three indirectly acting haptens, which were considered to act exclusively as pro-haptens. One of these chemicals, ethylenediamine, was misclassified in 9 out of 10 studies. Ethylenediamine is assumed to act as a pro-schiff base electrophile by conversion of the amine entity to the corresponding aldehyde. The same mechanism, involving the conversion to the corresponding aldehyde, has also been assumed for the chemical 3-dimethylaminopropylamine, which was correctly classified, indicating that the misclassification of ethylenediamine cannot solely be attributed to a general lack in the capacity of detecting chemicals acting via a pro-schiff base reaction mechanism (Patlewicz et al., 2016). Furthermore, it can be assumed that additional pro-haptens were present in the current dataset, but based on their structures, it was not possible to determine if the main pathway of activation required abiotic or biotic activation. In addition, the chemical aniline was misclassified (false negative) in one GARDTMskin study. Considering the otherwise concordant classifications with reference data in the chemical space, no immediate explanation for this misclassification has been identified.

Hydrophobic substances have also been considered as challenging to accurately assess in the currently OECD validated *in vitro* assays, which are largely based on submerged cell-cultures (Mehling et al., 2019, Takenouchi et al., 2013). In this aspect, the major theoretical and regulatory concern is related to the requirement of the test substance to be soluble to a sufficiently high concentration in the aqueous cell media to exceed the limit of detection in the assay, and hence prevent a false negative classification due to testing at a too low concentration. Based on the herein reported results, the sensitivity for detecting hydrophobic substances was 85.1% (20.4/24.0) compared with LLNA references and 100% (9/9) compared with human data, indicating a low risk of false negative classifications because of limitations in solubility of test chemicals. While it is acknowledged that limited substance solubility may indeed be challenging in submerged cell culture systems, the capacity of an assay to detect such compounds is dependent on assay sensitivity, defined as the lowest detectable concentration, and the soluble concentration in the surrounding media, likely resulting in different performances for such compounds for various assays. Considering GARDTMskin, recent data suggest that the majority of sensitizers, irrespective of sensitizing potency, are

detected at concentrations below 100 μM (Gradin et al., 2021). Thus, if a substance is soluble to this level in the cellular media, GARDTMskin is expected to provide accurate predictions, avoiding false negatives. In addition, while not specifically applied in the above-described testing, but as illustrated in the UVCB study discussed below, potential solubility issues may be further mitigated by selection and evaluation of experimental and less polar solvents.

The available data indicate that specificity when compared to human data is notably low, in contrast to the expectation that sensitivity may be lacking when assessing hydrophobic substances. It should be recognized that the subset of expected non-sensitizers is low ($n=4$), obstructing any decisive conclusions from the herein investigated data. Indeed, additional testing of chemicals with negative reference values may be warranted, in order to further and more accurately estimate specificity, and thereby also the overall accuracy, within the applicability domain. Nonetheless, it is evident that LLNA and human references, both of which were extracted from the reference dataset provided by TG 497, are in conflict for the three false positives, as compared to human references: benzyl benzoate, citronellol and hexyl salicylate. Additionally, they are all considered weak human sensitizers (categories 4-5) according to the human potency categories suggested by Basketter et al. (2014). Thus, these GARDskin data are in line with comparable sources of information, corroborating the inherent borderline nature of the compounds, with different results obtained depending on the considered reference classification. Of important note, however, the sole chemical considered as a non-sensitizer by both LLNA and human data references included in the dataset, n-hexane, is also accurately classified as such by GARDskin.

For most investigated test chemicals, GARDskin classifications are unambiguous and reproducible. However, exceptions do exist. The most notorious borderline classifications include those for test chemicals benzyl benzoate, benzyl cinnamate and benzyl salicylate, which is evident both from individual replicate samples within studies (Fig. S2¹) as well as the observed reproducibility between studies (Table 2), indicating the difficulty to reach a conclusive result for these test chemicals also in GARDskin. Additional borderline classifications have been obtained also for other test chemicals; however, they have had little to no impact on the correct final classification, which is based strictly on the mean DV. Furthermore, to the extent where repeated studies are available, also the majority of borderline classifications are reproducible between studies, e.g., 2,5-diamino toluene sulfate, ethylene diamine, geraniol and n-hexane (Fig. S1-S2¹).

Here, the term borderline classification is used to describe a result which is based on individual replicate samples from both sides of the classification threshold ($DV=0$). It should be stressed, however, that a formal procedure for identification of inconclusive results has not been implemented in the validated and the herein used GARDskin protocols. All figures of reproducibility and predictive capacity, in the herein reported studies as well as in the draft GARDskin TG, have been generated without acknowledging so-called borderline classifications. For future work, however, attempts to further define a borderline range from which conclusive results cannot be obtained may indeed be a relevant adaptation of GARDskin protocols.

In an attempt to put the obtained GARDskin results into context, corresponding data from the three validated *in vitro* methods included in the TG 497 was extracted (OECD, 2021b). Compiled data for indirectly acting haptens and hydrophobic substances are summarized in Tables S4-S5¹, respectively. Both datasets are complete, without missing values, allowing for a robust method comparison based on harmonized underlying data. The performances of each method are summarized with confusion matrices and Cooper statistics in Tables S6-S8¹, corresponding to the GARDskin statistics summarized in Table 4.

For the subset of indirectly acting haptens, the accuracy was for DPRA 56.0% and 50.0%, for KeratinoSens 72.0% and 62.5% and for hCLAT 88.0% and 87.5%, to be compared with GARDskin estimates of 92.4% and 87.5%, using LLNA and human references, respectively. For the subset of hydrophobic substances, the accuracy was for DPRA 48.0% and 69.2%, for KeratinoSens 60.0% and 61.5% and for hCLAT 72.0% and 76.9%, to be compared with GARDskin estimates of 85.7% and 80.0%, using LLNA and human references, respectively.

Overall, the summarized results indicate that these chemical subsets may indeed be regarded as difficult to accurately assess in one or several methods. These concerns are well recognized and, when applicable, also specifically addressed in the respective Test Guidelines. For these very reasons, an important conclusion from the herein reported results may be that GARDskin may contribute with important properties of predictive capacity in a wide chemical space where complementary methods may be lacking, highlighting its possibly advantageous use in such contexts.

Lastly, a set of UVCB substances, comprising generally hydrophobic specialty chemicals from the petroleum industry, was investigated. For these substances, the main challenge in terms of *in vitro* assessment was associated with their limited water solubility, as well as their complex and undefined nature, which in turn was hypothesized to increase the risk of false negative classifications. Therefore, the majority of the test items in this study were indeed sensitizers, and few non-sensitizers were evaluated. However, based on the results from this study, the limited solubility for some of the test items did not appear to have a major impact on the potential of the GARDTMskin assay to accurately identify them as skin sensitizers, and no false negative classification was observed in the dataset.

Of important note, while these substances may support GARDTMskin functionality in a chemical space similar to that which was investigated, i.e., hydrophobic substances originating from the petroleum industry, this dataset is too small and homogeneous to draw any conclusions regarding GARDTMskin applicability in the domain of UVCBs at large. Indeed, the term UVCB may represent a vast chemical space, and care should be taken when considering the selection of an appropriate test method, depending on the specific test chemical chemistry at hand. However, in addition to supporting the GARDTMskin applicability in the domain of hydrophobic substances, the examination of this small dataset serves to illustrate two important aspects of the GARDTMskin protocol and its adaptability. Firstly, it highlights the opportunity to explore experimental vehicles in order to increase compatibility with otherwise insoluble test items. While complex and highly hydrophobic substances such as the here invested UVCBs may remain partly insoluble, different solvents may be advantageously explored to generate better dispersions, facilitating enhanced transfer of test chemical molecules from the dispersion to the hydrophobic membranes of the cells, thereby enhancing bioavailability. Secondly, the UVCB case study exemplifies a possible protocol adaptation for complex mixtures for which a molecular weight is not defined. Of important note, however, any such adaptations of the method protocol should be scientifically justified, either by estimates of similarity to compatible sources of information, as was done here, or by any other rationale which may be found scientifically appropriate.

While considered a non-animal method, similar to methods currently included in TGs 442D-E, GARDskin is currently dependent on animal components, including Fetal Calf Serum (FCS) and monoclonal antibodies. However, important work has demonstrated the feasibility of completely animal component-free adaptations of both KeratinoSens (Belot et al., 2017) and h-CLAT (Edwards et al., 2018). For future work, exploring similar adaptations of GARDskin may be a way to further push the boundaries of NAMs, allowing for a sensitive and animal component-free method for assessment also of test chemicals typically considered difficult to test.

In conclusion, based on the herein presented datasets, estimates of GARDTMskin's predictive performance when evaluating indirectly acting haptens, as well as substances with limited water solubility, are in line with previously reported estimates for other datasets comprising organic low-molecular-weight chemicals from a wider chemical space. Importantly, reported data indicate that the rate of false negative classifications associated with the investigated chemistries is relatively low, suggesting that negative GARDTMskin results for such chemistries can be used for decision making without compromising safety. Thus, available data supports the inclusion of indirectly acting haptens and hydrophobic substances into the applicability domain of the test method.

References

- Basketter, D. A., Alepee, N., Ashikaga, T. et al. (2014). Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis* 25, 11-21. doi:10.1097/DER.0000000000000003
- Belot, N., Sim, B., Longmore, C., Roscoe, L. and Treasure, C. (2017). Adaptation of the KeratinoSensTM skin sensitization test to animal-product-free cell culture. *ALTEX* 34, 560-564. doi:10.14573/altex.1701311
- Cooper, J. A., Saracci, R. and Cole, P. (1979). Describing the validity of carcinogen screening tests. *British Journal of Cancer* 39,87-9. doi:10.1038/bjc.1979.10
- Corsini, E., Clewell, R., Cotgreave, I., Eskes, C., Kopp-Schneider, A., Westmoreland, C., Alves, P.M., Navas, J.M. and Piersma, A. (2021). ESAC Opinion on the Scientific Validity of the GARDskin and GARDpotency Test Methods. Asturiol Bofill, D., Casati, S. and Viegas Barroso, J.F. editor(s), Publications Office of the European Union, Luxembourg, 2021. doi:10.2760/626728
- Edwards, A., Roscoe, L., Longmore, C., Bailey, F., Sim, B. and Treasure, C. (2018). Adaptation of the human Cell Line Activation Test (h-CLAT) to Animal-Product-Free Conditions. *ALTEX* 35, 477-488. doi:10.14573/altex.1710051
- EURL ECVAM - EU Reference Laboratory for alternatives to animal testing (2021). GARDskin Assay Protocol. <https://tsar.jrc.ec.europa.eu/system/files/Published/GARDskin%20Assay%20Protocol%20TSAR.pdf>
- Forreryd, A., Zeller, K. S., Lindberg, T. et al. (2016). From genome-wide arrays to tailor-made biomarker readout - Progress towards routine analysis of skin sensitizing chemicals with GARD. *Toxicol In Vitro* 37, 177-88. doi:10.1016/j.tiv.2016.09.013
- Gradin, R., Lindstedt, M. and Johansson, H. (2019). Batch adjustment by reference alignment (BARA): Improved prediction performance in biological test sets with batch effects. *PLoS One* 14, e0212669. doi:10.1371/journal.pone.0212669
- Gradin, R., Forreryd, A., Mattson, U., et al. (2021). Quantitative Assessment of Sensitizing Potency using a dose-response adaptation of GARDskin. *Sci Rep* 176, 423-432. doi:10.1038/s41598-021-98247-7
- Johansson, H., Lindstedt, M., Albrekt, A. S. and Borrebaeck, C. A. (2011). A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. *BMC Genomics* 12, 399. doi:10.1186/1471-2164-12-399
- Johansson, H., Gradin, R., Forreryd, A., Agemark, M. et al. (2017). Evaluation of the GARD assay in a blind Cosmetics Europe study. *ALTEX* 34, 515-523. doi:10.14573/altex.1701121
- Johansson, H., Gradin, R., Johansson, A., et al. (2019). Validation of the GARDskin Assay for Assessment of Chemical Skin Sensitizers: Ring Trial Results of Predictive Performance and Reproducibility. *Toxicol Sci* 170, 374-381. doi:10.1093/toxsci/kfz108
- Mehling, A., E. Adriaens, S. Casati, B. et al. (2019). In vitro RHE skin sensitisation assays: Applicability to challenging substances." *Regul Toxicol Pharmacol* 108, 104473. doi:10.1016/j.yrtph.2019.104473
- OECD - Organisation for Economic Cooperation and Development (2014). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. Series on Testing and Assessment No. 168. Available at: doi:10.1787/9789264221444-en
- OECD - Organisation for Economic Cooperation and Development (2015a). OECD Guideline for the Testing of Chemicals No. 442C: In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA). Paris, France: Organisation for Economic Cooperation and Development. doi:10.1787/9789264229709-en
- OECD - Organisation for Economic Cooperation and Development (2015b). OECD Guideline for the Testing of Chemicals No. 442D: In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. Paris, France: Organisation for Economic Cooperation and Development. doi:10.1787/9789264229822-en
- OECD - Organisation for Economic Cooperation and Development (2015c). Performance standards for assessment of proposed similar or modified in vitro skin sensitisation ARE-NRF2 luciferase test methods. Series on Testing and Assessment, No. 213. Paris, France: Organisation for Economic Cooperation and Development. doi:10.1787/9789264229822-en
- OECD - Organisation for Economic Cooperation and Development (2017a). OECD Guideline for the Testing of Chemicals No. 442E: In Vitro Skin Sensitisation assays addressing the AOP Key Event on Activation of Dendritic Cells. Organisation for Economic Cooperation and Development, Paris. doi:10.1787/9789264264359-en
- OECD - Organisation for Economic Cooperation and Development (2017b). Revised performance standards for the assessment of proposed similar or modified in vitro Reconstructed human Cornea-like Epithelium (RhCE) test methods for eye hazard. Series on Testing and Assessment, No. 216. Paris, France: Organisation for Economic Co-operation and Development. Available at:

- [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO\(2015\)23&doclanguage=en](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2015)23&doclanguage=en)
- OECD - Organisation for Economic Cooperation and Development (2021a). OECD Guideline for the Testing of Chemicals No. 497: Defined Approaches on Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. doi:10.1787/b92879a4-en
- OECD - Organisation for Economic Cooperation and Development (2021b). Series on Testing and Assessment No. 336: Annex 2 of the Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: <http://www.oecd.org/env/testguidelines>
- Patlewicz, G., S. Casati, D. A. Basketter, D. et al. (2016). Can currently available non-animal methods detect pre and pro-haptens relevant for skin sensitization? *Regul Toxicol Pharmacol* 82, 147-155. doi:10.1016/j.yrtph.2016.08.007
- Takenouchi, O., Miyazawa, M., Saito, K., et al. (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. doi:10.2131/jts.38.599
- UN GHS - United Nations, Globally Harmonized System of Classification and Labelling of Chemicals 2015. Sixth revised edition. New York & Geneva: United Nations Publications. doi:10.18356/591dabf9-en
- Urbisch, D., M. Becker, N. Honarvar, S. N. et al. (2016). Assessment of Pre- and Pro-haptens Using Nonanimal Test Methods for Skin Sensitization. *Chem Res Toxicol* 29, 901-13. doi:10.1021/acs.chemrestox.6b00055.

Data availability statement

Data used to arrive at the conclusions presented in this manuscript are available with permission from the authors upon request.

Conflict of interest

AF, RG, and HJ are employees of SenzaGen AB, which offers GARDskin as a commercial service.

Acknowledgements

This work was financed in its entirety by SenzaGen AB and The Lubrizol Corporation.