Challenges integrating skin sensitization data for assessment of difficult to test substances

Allison Greminger^{1*}, Joseph Frasca², Katy Goyak¹, Colin North^{1,**}

¹ExxonMobil Biomedical Sciences, Inc., Health and Environmental Sciences Division, Annandale, NJ, USA

²ExxonMobil Biomedical Sciences, Inc., Health and Environmental Sciences Division, Spring, TX, USA

****Note:** CN authorship fully occurred during time of employment with ExxonMobil Biomedical Sciences, Inc. Current affiliation is BASF Corporation, 26 Davis Drive, Research Triangle Park, NC.

Keywords: skin sensitization, weight of evidence, ToxPi, new approach methodologies, LLNA

Abstract

Difficult to test substances, such as poorly soluble, mildly irritating, or Unknown or Variable Composition Complex reaction products or Biological Materials (UVCBs), producing weak or borderline results in the Local Lymph Node Assay (LLNA), often benefit from additional data integration in a weight of evidence (WOE) approach. Advances in multiple testing and non-testing methodologies (*in vivo*, *in vitro*, and *in silico*) can provide clarity and confidence in concluding on skin sensitization potential. Here we present several case studies using a WOE approach with difficult to test substances and highlight the utility of Toxicological Prioritization Index™ (ToxPi™) as a comparative visualization and integration tool. The three test chemicals chosen represent poorly soluble substances, tetrakis (2-ethylbutyl) orthosilicate and decyl palmitate, including one UVCB substance, alkylated anisole. Data from *in vivo* and *in vitro* assays representing multiple key events within the skin sensitization adverse outcome pathway (e.g., direct peptide reactivity assay, human cell line activation test, GARD®Skin, LLNA) were gathered from publicly available sources or specifically generated. Incorporating the data on our test chemicals as well as chemicals of a known sensitization class (sensitizer, irritating non-sensitizer, and non-sensitizer) into ToxPi™ revealed biological activity profiles which were used to support class prediction for the three test chemicals. Using this method, the biological activity profiles for all three test chemicals were most consistent with a non-sensitizing class. Visualizing the WOE using mechanistic data maximizes the value of all data to the overall assessment of skin sensitization potential by reducing individual assay associated uncertainty.

Introduction

Assessment of skin sensitization potential is a required hazard endpoint for many chemical registration systems globally. It can also be an important parameter in research and development, as the presence or absence of skin sensitization potential can dictate whether a chemical is suitable in different applications, and consequently whether investment in development is warranted. Hence, correct assessment of chemicals for sensitization has important implications both for product safety/stewardship, and investment decisions for new chemical development. Historically skin sensitization testing has relied on laboratory animal models measuring elicitation, such as the Guinea Pig Maximization test (Magnusson and Kligman 1969) and the Buehler Test (Buehler 1965), or those measuring induction, such as the Local Lymph Node Assay (LLNA) (Kimber, et al. 1994). In some cases skin sensitization testing has been performed in human volunteers via Human Repeat Insult Patch Testing (McNamee, et al. 2008). Over time the LLNA became the predominant test used to assess skin sensitization potential, as it provided multiple advantages over other methods, such as increased sensitivity and specificity, quantitative measures of potency, and reduced animal use (Gilmour, et al. 2019, Kimber, et al. 1995).

Even given these advantages, there are still occasions where sensitization potential is difficult to conclude due to uncertainties around potential false positives, predictive performance vs observed human and guinea pig responses, and technical limitations such as solvent-dependent responses (Anderson, et al. 2011). For instance, the LLNA has been recognized both for its limited ability to distinguish irritants and weak sensitizers (Adenuga, et al. 2012, Basketter, et al. 1998, Montelius, et al. 1998) and for its inability to consistently identify chemicals producing borderline or weak results in which the “true” biological activity is close to the classification threshold (Kolle, et al. 2013). In these borderline instances, an individual LLNA result may fall above or below the classification threshold, potentially leading to discordant conclusions regarding the likelihood to produce a sensitization response (Dimitrov, et al. 2016, Hoffmann 2015, Kolle, et al. 2013). Despite these recognized limitations, the predictive performance of the LLNA has been reported as approximately 70-80% compared to human allergic reactions (ICCVAM 1999, Schneider and Akkan 2004, Urbisch, et al. 2015), providing a solid foundation to assess the skin sensitization potential of the vast majority of chemicals. Additional mechanistic information acquired from New Approach Methodologies (NAM), can be incorporated to increase confidence in characterization of skin sensitization.

Significant progress has been made to develop predictive computational and *in vitro* models based on well-characterized mechanisms of skin sensitization. As a result, there are widely available Organization for Economic Co-operative Development (OECD) guideline tests which address specific key events (KE) in the skin sensitization adverse outcome

pathway (AOP) and, when used in combination, appear to equal the predictive performance of *in vivo* laboratory animal tests (Strickland, et al. 2016). Briefly, the AOP consists of four KEs and captures a simplistic description of the cascade of biological events responsible for chemical mediated skin sensitization characterized by protein binding, cell activation, and subsequent elicitation of an immune reaction (Figure 1). The Direct Peptide Reactivity Assay (DPRA) addresses KE1 in the sensitization pathway, the covalent binding of a hapten to skin proteins (Gerberick 2016, OECD 2020). At the cellular level, KE2 is the activation of a keratinocyte response, which can be assessed in the KeratinoSens™ or LuSens™ assays (Natsch, et al. 2011, OECD 2018a, Ramirez, et al. 2014); while KE3, the activation of dendritic cell response, can be assessed in the human Cell Line Activation Test (h-CLAT), interleukin 8 luciferase (IL-8 LUC), U-SENS™ (Ashikaga, et al. 2006, OECD 2018b) or GARD®Skin (Corsini, et al. 2021) (Figure 1). In addition, model-based screening tools, such as OASIS TIMES or QSAR Toolbox, supplement our understanding of chemicals of interest by identifying structural fragments associated with known skin sensitization mechanisms in parent structures as well as probable metabolites occurring in skin or by autoxidation (Patlewicz, et al. 2007, Roberts, et al. 2007).

Despite the mechanistic understanding of the biological pathway and associated tests, these KE-based tests each have clearly defined requirements for test substance properties that impact feasibility or confound reliance on negative results particularly for difficult to test substances (Table 1). In our experience, difficult to test substances include UVCBs and/or those that have challenging physical chemical properties (e.g. low water solubility, high LogKow). For example, evidence indicates that the h-CLAT assay demonstrates low predictive performance for test chemicals with high octanol-water partition coefficients increasing the likelihood of a false negative result (Takenouchi, et al. 2013); while the DPRA is not applicable to UVCBs due to the need for an exact ratio of test substance to peptide which is made difficult by the unknown or variable molecular weights of UVCB substances (EURL-ECVAM 2012). Additionally, QSAR models typically have applicability domains that do not include UVCBs or substances on the extremes of physical chemical property boundaries (OECD 2020, OECD 2021). Recognizing the uncertainties associated with relying on a single assay to determine skin sensitization potential, two defined approaches (DA) that represent transparent, rule-based systems to data interpretation of KE-based NAMs have been published by the OECD and have been shown to better predict human skin sensitization hazard and potency when compared to the LLNA alone (Hoffmann, et al. 2018, Kleinstreuer, et al. 2018, OECD 2021). Nonetheless, application of the two published DAs remain challenged for difficult to test substances due to the specific test method requirements described above as well as a higher burden of proof for negative results in certain circumstances. Therefore, difficult to test substances which produce borderline or weak results in the LLNA face additional challenges in KE-based *in vitro* assays that ultimately necessitates the application of WOE approaches in order to conclude on skin sensitization potential.

In this study, we present several real world experiences in assessing challenging substances in an overall, flexible WOE utilizing the ToxPi™ framework to compare biological activity as it relates to skin sensitization. The ToxPi™ graphical user interface enables integration of multiple data sources and types, e.g. chemical/physical properties, test data, exposure information, by transforming data into visual rankings (Marvel, et al. 2018). An advantage of this approach is its flexibility in assessing various parameters of biological activity and has been used previously to assist grouping of complex substances (Grimm, et al. 2019, Grimm, et al. 2016), as well as ranking and prioritizing of chemicals for further study (Reif, et al. 2010, Tilley, et al. 2017). This WOE approach facilitated the comparison of several challenging test chemicals to prototypical sensitizers and non-sensitizers and the results demonstrate that a more comprehensive assessment of biological activity allows for fit-for-purpose data-driven decision making.

Materials & Methods

Test Chemicals

Test samples were obtained as follows: decyl palmitate (CASRN 42232-27-9) and alkylated anisole (a reaction mass resulting in anisole attached to one or more variably branched hydrocarbon chains) from ExxonMobil Chemical Company (Houston, TX USA); tetrakis (2-ethylbutyl) orthosilicate (CASRN 78-13-7) from Gelest, Inc. (for LLNA, h-CLAT and DPRA) or

from Sigma Aldrich (for GARD®Skin); hydroquinone (CASRN 123-31-9), p-phenylenediamine (CASRN 106-50-3), dimethylsulfoxide (DMSO, CASRN 67-68-5), acetone (CASRN 98-86-2), and methyl salicylate (CASRN 119-36-8) from Sigma Aldrich for GARD®Skin. Assay data not conducted by ExxonMobil or contracted laboratories were obtained from SkinSensDB (accessible at: <http://cwtung.kmu.edu.tw/skinsensdb>), a curated database of skin sensitization assay data (Wang, et al. 2017). Specifically, data from OECD TG 442C- DPRA, OECD TG 442D- Keratino-Sens™ /U-Sens™, OECD TG 442E- h-CLAT and OECD TG 429/442B-LLNA were incorporated for hydroquinone, p-phenylenediamine, 2-mercaptobenzothiazole (CASRN 149-30-4), and cinnamal (CASRN 67485-29-4) as prototypical skin sensitizers; DMSO, acetone, methyl salicylate, and salicylic acid (CASRN) as prototypical non-sensitizers; and lactic acid (CASRN 50-21-5), hexane (CASRN 110-54-3), nonanoic acid (CASRN 112-05-0), and sodium lauryl sulfate (CASRN 151-21-3) as irritating non-sensitizers. GARD®Skin assay results for 2-mercaptobenzothiazole, cinnamal, hexane, sodium lauryl sulfate, salicylic acid, and lactic acid were obtained from Johansson, et al. (2014).

Local Lymph Node Assay (LLNA)

The LLNA for alkylated anisole and decyl palmitate were conducted at Envigo (Shardlow, UK) under OECD Principles of Good Laboratory Practice (GLP) (ENV.MC/CHEM (98) 17). Both studies were designed and conducted in accordance with OECD TG 429 and the requirements of the United Kingdom Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012. The studies were not reviewed as part of Envigo – Shardlow Ethical Review Process as it was not deemed necessary by the laboratory. Animals (25 total/study) were in the weight range of 15 to 23 g and were 8 to 12 weeks old at study initiation; and randomly assigned to dose group upon receipt using a non-selective allocation to equalize variation in body weight. Prior to the start of treatment (5 days acclimatization), animals were examined to ensure there were no observable skin lesions. Based on initial irritation screening in a single mouse, decyl palmitate was tested at 0%, 10%, 25% and 50% (v/v) in butanone in five female CBA/Ca mice per dose group; alkylated anisole was tested at 0%, 25%, and 50% (v/v) in acetone:olive oil (4:1, AOO) and undiluted (100%) in five female CBA/Ca mice per dose group. Five days following the first topical application of test item, all mice were injected via tail vein with 20 µCi ³H-methyl thymidine (³HTdR). Animals were sacrificed five hours following administration of ³HTdR, and the draining auricular lymph nodes were excised and combined for each animal in a single cell suspension. ³HTdR incorporation was measured by scintillation counting reported as disintegrations per minute (DPM). The Stimulation Index (SI) was calculated by dividing the mean DPM of each animal by the mean DPM of the vehicle control group. A test article was regarded as a sensitizer if any treatment group yielded an SI value of 3 or more according to OECD TG 429.

The LLNA for tetrakis (2-ethylbutyl) orthosilicate was conducted at MB Research (Spinnerstown, PA USA) under GLP standards. The study protocol was reviewed by the Institutional Care and Use Committee of MB Research and found to be in compliance with acceptable ethical standards of animal welfare. The study was designed and conducted in accordance with OECD TG 442B and the Guide for Care and Use of Laboratory Animals: Eighth Edition, National Academies Press. Animals (31 total) were in the weight range of 18.5 to 23 g and were 8 to 12 weeks old at study initiation; and randomly assigned to dose group upon receipt using a non-selective allocation to equalize variation in body weight. Prior to the start of treatment (5 days acclimatization), animals were examined to ensure there were no observable skin lesions. Based on initial irritation screening in 2 mice per dose, the substance was tested at 10%, 25% and 50% (v/v) in acetone:olive oil (4:1, AOO) in five female CBA/J mice per dose group. Five days following the first topical application of test item, all mice were administered BrdU by intraperitoneal injection. Animals were sacrificed 24 hours following administration of BrdU, and the draining auricular lymph nodes were excised and combined for each animals into a single cell suspension. BrdU incorporation was measured using a BrdU-specific Enzyme-Linked Immunosorbent Assay (ELISA) kit and reported as optical density values. The Stimulation Index (SI) was calculated by dividing the mean optical density of each animal by the mean optical density of the vehicle control group. The mean SI and standard deviation (S.D.) were calculated for each group from the individual animal data. A test article was regarded as a sensitizer if any treatment group yielded an SI value of 1.6 or more according to OECD TG 442B. Both OECD TG have undergone extensive inter-laboratory validation and have been shown to reliably detect moderate to strong sensitizers, thus no additional efforts to minimize

potential confounders were applied to the study design. No blinding of group allocations was required for these studies as the data is determined by digital measurement and is not required by the OECD TG.

GARD®Skin

GARD®Skin studies were facilitated by MB Research (Spinnerstown, PA USA) with SenzaGen (Lund, Sweden) acting as lead test laboratory. All protocols associated with the GARD®Skin assay have been previously published (Forreryd, et al. 2016, Johansson, et al. 2013). In short, cultivated human dendritic cell-like SenzaCells (ATCC Depository PTA-123875), are exposed *in vitro* to the test chemical(s) of interest for 24 h. Following dose-response measurements of induced cell toxicity, an appropriate and test chemical specific input concentration is defined at non- to low-toxic levels. Transcriptional levels of the GARD®Skin prediction signature (Johansson, et al. 2011) are quantified using the Nanostring nCounter system. The data are analyzed by a Support Vector Machine (SVM) (Cortes and Vapnik 1995), which has been appropriately trained on samples generated during technology platform transfer, consisting of expressional profiles generated by the reference panel of chemicals, as described in Forreryd, et al. (2016). Final prediction calls are derived from the mean SVM decision value (Johansson, et al. 2019) generated by the biological triplicate samples; any test chemical inducing a mean SVM decision value ≥ 0 is classified as a skin sensitizer. Consequently, any test chemical inducing a mean SVM decision value < 0 is classified as a non-sensitizer.

Direct Peptide Reactivity Assay (DPRA)

The DPRA was facilitated by MB Research Laboratories (Spinnerstown, PA USA) with EAG Laboratories (Easton, MD, USA) acting as lead test laboratory under GLP standards according to OECD TG 442C. Three replicate solutions of tetrakis (2-ethylbutyl) orthosilicate was incubated with cysteine or lysine containing peptides at a molar ratio of 1:10 and 1:50, respectively. Relative peptide concentration was measured by high performance liquid chromatography (HPLC) with gradient elution and ultraviolet (UV) detection at 220 nm following 24 ± 2 hours of incubation with the test chemical at a temperature of $25 \pm 2.5^\circ\text{C}$. The concentration of the cysteine and lysine peptides were determined in the appropriate reference control samples by substituting the peak area responses into the appropriate linear regression equation. The percent peptide depletion was determined in each test substance assay and positive control sample by measuring the peak area of the appropriate peptide and dividing it by the mean peak area of the relevant reference control samples.

Human Cell Line Activation test (h-CLAT)

The h-CLAT was conducted at MB Research Labs (Spinnerstown, PA USA) under GLP standards according to OECD Test Guideline 442E except where noted below. Activation of a human monocytic leukemia cell line (THP-1; ATCC No. TIB-202) was evaluated by measuring changes in the expression of dendritic cell surface markers (CD86, CD54) via flow cytometry. The test guideline was modified to utilize 100% Ethanol (1% final dilution) as the test substance solvent as initial solubility studies indicated tetrakis (2-ethylbutyl) orthosilicate was not soluble in either DMSO or saline (0.9%). Additional solvent controls (1% Ethanol in RPMI-10 culture medium), as well as proficiency controls (isopropanol CARN: 67-63-0 non-sensitizer; 2-mercaptobenzothiazole CASRN: 149-30-4 moderate sensitizer) were tested concurrently with the negative control (Lactic Acid; CASRN 50-21-5, $\geq 85\%$ purity) and both positive controls (1-chloro-2,4-dinitrobenzene DNCB CASRN 97-00-7, $\geq 99\%$ purity and Nickel Sulfate CASRN 10101-97-0, $\geq 99\%$ purity) in the study to confirm no interference of the alternative vehicle. THP-1 cells were cultured, at 37°C with tetrakis (2-ethylbutyl) orthosilicate in concentrations up to 5,000 $\mu\text{g/ml}$. After 24 ± 0.5 hours of exposure, CD54/86 expression is measured by FITC-labelled anti-CD86, anti-CD54 or mouse IgG1 (isotype) antibodies at 4°C for 30 min. Cell viability was measured by propidium iodide staining. Results were quantified as the effective concentration inducing a relative fluorescence intensity (RFI) positive threshold criteria for each cell surface marker (RFI = 150 for CD86, and RFI = 200 for CD54). An RFI greater than either threshold for any test dose (with cell viability $> 50\%$) in at least two independent assays the substance was predicted as positive, otherwise it was considered negative.

In-silico

OECD QSAR Toolbox 4.3 was utilized to check for skin sensitization structural alerts (SA) of all chemicals presented here as well as metabolites by autoxidation or skin metabolism. A QSAR prediction for skin sensitization was generated using the Skin Sensitization with autoxidation v. 20.24 model of the OASIS TIMES v.2.29.1 software. Values for the energy of the Highest Occupied Molecular Orbital (eHOMO, eV) and Softness (σ) were obtained from OASIS TIMES and included in the ToxPi™ data integration under KE1 as these molecular parameters have been shown to inform nucleophile/electrophile interactions as they relate to skin sensitization potential (LoPachin, et al. 2012, LoPachin, et al. 2019, Mekenyan, et al. 1997, Ouyang, et al. 2014).

Data integration in Toxicological Priority Index (ToxPi™)

The data were separated based on their associated key event within the skin sensitization AOP; the framework for profiling and data visualization is detailed in **Figure 2**. ToxPi™ is represented as component slices of a unit circle with each slice representing a piece of information. For each slice, the distance from the origin is proportional to a normalized value, and the width represents the relative weight of that piece of information within the overall ToxPi™ profile. In the implementation presented here, all 12 slices are weighted equally and are normalized so that larger pie slices correlate to increased biological activity as it relates specifically to skin sensitization. For example, in **Figure 2**, where the orange slice represents LLNA results, a smaller EC 3 value (higher potency) correlates to a larger pie slice (further distance from origin). Raw data was log transformed and normalized using the min and max values for each endpoint; missing or inconclusive data points were entered as “NA” and not included in the dataset. See supplemental information for raw data, normalization and scaling of each endpoint. Hierarchical clustering of ToxPi™ profiles was organized using the complete linkage model. The clustering methods in ToxPi™ GUI correspond to those available for the R function hclust using Euclidean distance, and the optimal number of clusters highlighted by color was defined by the number of largest vertical distance dendograms which did not cut any horizontal lines (Marvel, et al. 2018).

Results

WOE for Tetrakis (2-ethylbutyl) Orthosilicate

Problem formulation

The skin sensitization potential of tetrakis (2-ethylbutyl) orthosilicate was assessed to determine the appropriate hazard classification for regulatory registration purposes. Because one objective was regulatory registration, the testing strategy was designed to meet the information requirements mandated by the regional regulatory authority, following the decision logic recommended by the regional guidance for skin sensitization hazard identification (ECHA 2017a, United Nations 2019). Annex VII of the Registration, Evaluation, and Authorization of Chemicals (REACH) regulation stipulates that *in vitro* assays should be considered prior to conducting animal-based assays (European Parliament Regulation EC No 1907/2006).

NAM data generation

No structural alerts for sensitization potential were identified by the OECD QSAR Toolbox (www.qsartoolbox.org). TIMES Skin Sensitization with autoxidation model predicted a negative skin sensitization result (see supplemental data), although it should be noted that the substance is outside of the model domain due to the presence of 30% unknown structural fragments.

Initial data generation focused on the mechanistically-based *in chemico* and *in vitro* methods associated with key events 1 and 3 in the skin sensitization AOP (Figure 1); namely, a DPRA (OECD 2020) to assess covalent interaction with cysteine and lysine peptides in skin and the h-CLAT (OECD 2018b) to assess the activation of dendritic cells through measurement of markers CD86 and CD54. In the DPRA, a precipitate was observed immediately upon addition of tetrakis (2-ethylbutyl) orthosilicate to the peptide solution, due to low aqueous solubility of the test chemical. Although the mean percent depletion values for cysteine and lysine ($1.12 \pm 1.0\%$ (N=3, CV=91.7%) and $0.251 \pm 0.43\%$ (N=3; CV = 173%), respectively)

predict no or minimal reactivity with skin proteins, the observed precipitate decreases confidence in this prediction, as the amount of chemical available in solution to react with the peptides is unknown. As per OECD Test Guideline 442C in this type of case (OECD 2020), no firm conclusion on the lack of reactivity should be drawn from a negative result. Therefore the assay is considered inconclusive and cannot be used as a definitive line of evidence in the overall assessment of skin sensitization potential. In the h-CLAT, tetrakis (2-ethylbutyl) orthosilicate was tested in 100% ethanol, due to insolubility of the test substance in either of the OECD TG validated solvents (saline and DMSO). In this modified version, the EC150 and EC200 for tetrakis (2-ethylbutyl) orthosilicate was >5000 µl/ml (highest dose tested). However, OECD TG 442E notes that substances with high octanol water partition coefficients (Log Kow>3.5) tend to produce false negative results (Takenouchi, et al. 2013). Based on a measured partition coefficient (Log Kow) >6.01, a firm predictive conclusion could not be drawn from this negative result. Consequently this result was considered inconclusive and not included in the WOE. Further analysis using the GARD®Skin assay predicted tetrakis (2-ethylbutyl) orthosilicate to be a non-sensitizer based on a mean SVM decision value of -1.07 (**Table 2**); all three replicates produced SVM decision values below zero, the established threshold for a positive result.

Animal data generation

Because the WOE based on QSAR and *in vitro* assays were inconclusive for classification purposes, tetrakis (2-ethylbutyl) orthosilicate was tested in an LLNA (OECD TG 442B). The test substance was applied to CBA:JN mice at 10%, 25%, and 50%, which produced stimulation indices (SI) of 1.1, 2.5, and 4.9, respectively (**Table 2**). However, it should be noted that the 50% concentration produced a mean ear thickness of >25% indicating “excessive local irritation” according to the OECD TG 442B. Although the substance would be classified as a skin sensitizer based on the results of OECD TG criteria, it seems there was a positive dermal irritation response in the LLNA. It should be considered that the SI value of 4.9 produced by the 50% concentration level may be at least partially influenced by irritation. Considering an SI=1.6 as a positive indication of skin sensitization in this assay, the effective concentration (EC) at which the test substance may produce skin sensitization (i.e. EC 1.6) is calculated to be 15.5%. While GHS criteria does not yet exist for sub-categorization into high potency or low/moderate potency based on an SI of 1.6, the magnitude of the dose response is suggestive of low to moderate potency.

Evidence integration

Tetrakis (2-ethylbutyl) orthosilicate was inconclusive in QSAR, DPRA, and h-CLAT due to technical and predictive limitations of the assays with this specific chemistry. Tetrakis (2-ethylbutyl) orthosilicate was negative in GARD®Skin, corresponding to KE3 in the AOP. The substance met the criteria for a positive response in LLNA corresponding to KE4, T-cell proliferation. Although the LLNA result suggests that tetrakis (2-ethylbutyl) orthosilicate may meet the classification criteria for a skin sensitizer, no additional mechanistic inputs across the AOP for skin sensitization corroborate this conclusion. Taken together with no genomic profile, i.e. GARD®Skin results, or structural alerts associated with known sensitization responses, the entirety of the evidence suggests that tetrakis (2-ethylbutyl) orthosilicate is less likely to have skin sensitization potential.

WOE for decyl palmitate

Problem formulation

The skin sensitization potential of decyl palmitate was assessed to determine priority as a candidate substance for a product development project. Initial screening by computational methods was applied, and recognizing the chemical’s low water solubility may cause negative or inconclusive results to be rejected as a basis for global regulatory hazard identification purposes, no *in vitro* assays were planned for use. The LLNA was utilized for hazard identification due to its wide acceptance globally for that purpose. Due to conflicting *in silico* and *in vivo* results, additional evidence from the GARD®Skin assay was sought to provide additional mechanistic context to clarify known complications associated with the LLNA (e.g. marginal result).

NAM data generation

Initial *in silico* screening using OECD QSAR Toolbox and TIMES models suggested the chemical was unlikely to be a skin sensitizer. An assessment of physico-chemical properties and structural alerts did not give any indication of skin sensitization potential. TIMES Skin Sensitization with autoxidation model predicted a negative skin sensitization result (see supplemental data).

The GARD®Skin assay predicted decyl palmitate to be a non-sensitizer based on a mean SVM decision value of -1.16 (**Table 2**); all three replicates produced SVM decision values below zero, the established threshold for a positive result.

Animal data generation

The [³H]-thymidine-based LLNA (OECD TG 429) was used for the purposes of hazard characterization. No skin irritation was observed in a preliminary evaluation of local skin irritation in a single mouse. Decyl palmitate was tested at 10%, 25%, and 50% (% w/w) in butanone, reaching a stimulation index of 3.18 (threshold SI=3) at the highest tested concentration; the calculated EC3 value was 43% (**Table 2**). There are various thresholds for defining potency based on EC3 values, for instance ECETOC considers an EC3 value <10% as the threshold for strong sensitizers (ECETOC 2003) while ECHA uses a threshold of <2% to indicate strong potency (ECHA 2017b). Under either threshold, decyl palmitate is considered a weak sensitizer.

Evidence integration

Initial *in silico* modeling predictions consistently predicted non-sensitizer, thus a positive LLNA result was unexpected. Individual dose groups exposed to decyl palmitate indicated a monotonically increasing proliferation, and while there were individual animals which were markedly higher in proliferation responses than others within the same group, they were not obvious outliers. GARD®Skin assay results predicted decyl palmitate to be a non-sensitizer, with individual replicate decision values suggesting that the genomic signature was not close to the classification profile. Given both *in silico* and *in vitro* results suggesting decyl palmitate is a non-sensitizer, and borderline *in vivo* results, the overall WOE suggests decyl palmitate is unlikely to be a skin sensitizer.

WOE for alkylated anisole

Problem Formulation

Similar to decyl palmitate above, skin sensitization potential for alkylated anisole was assessed to determine priority as a candidate substance for a product development project. In contrast to decyl palmitate, alkylated anisole is considered a UVCB, whose synthesis results in multiple isomers, and as such cannot be represented by a single molecular structure. Initial screening by computational methods was applied and given that the chemical's low water solubility may cause negative or inconclusive results to be rejected as a basis for global regulatory hazard identification purposes, no *in vitro* assays were planned for use. LLNA was utilized for hazard characterization due to its wide acceptance globally for that purpose. Conflicting *in silico* and *in vivo* results lead to the generation of additional evidence in the form of GARD®Skin assay to inform a WOE.

NAM Data Generation

Initial *in silico* analysis was conducted on a single representative isomer, and was later expanded in OASIS TIMES using a range of hypothetical structures that were considered potentially present in the test material. The single representative parent structure was identified as producing a metabolite with a structural alert for the potential to form protein adducts. The metabolite reflected a step-wise metabolism wherein the anisole organic functional group was demethylated and then oxidized to a benzoquinone functional group. Based on this observation, additional model analysis was applied using a skin sensitization model in OASIS TIMES which accounts for kinetics of protein adduct generation. This model predicted all hypothesized isomers and their metabolites to be non-sensitizers (data not shown).

The GARD®Skin assay predicted alkylated anisole to be a non-sensitizer based on a mean SVM decision value of -1.61 (**Table 2**); all three replicates produced SVM decision values below zero, the established threshold for a positive result.

Animal data generation

The [³H]-thymidine-based LLNA was used for the purposes of hazard characterization. No skin irritation was observed in a preliminary evaluation of local skin irritation in a single mouse. Alkylated anisole was tested at 25%, 50%, and 100% (% w/w) in acetone:olive oil (4:1), and exhibited a non-monotonic dose response reaching a SI of 2.04 at 25%, 5.30 at 50%, and 3.39 at 100% concentration (**Table 2**). One potential outlier was identified in the 50% exposure group, whose decays per minute response was 15,833.75 compared with the group mean (inclusive of potential outlier) 6,455. Influence of the single observation is demonstrated by the difference between the arithmetic group mean and the median: 6,455 decays per minute compared with 3,898 decays per minute, respectively. Exclusion of the potential outlier would have resulted in a marginally lower stimulation index of 3.2 in the 50% group. The calculated EC3 value without outlier exclusion was 47%, whereas calculated EC3 value with outlier exclusion was 70%.

Evidence integration

Similar to decyl palmitate, the conflicting *in silico* and *in vivo* results were unexpected. Deeper review of TIMES *in silico* model predictions for multiple potential isomers identified potential minority metabolites with weak sensitizing potential, though inclusion of kinetic reaction rate information suggested these metabolites would be unlikely to result in skin sensitization. In the case of alkylated anisole, *in vivo* results from a single potential outlier appeared influential on the EC3 result, and consequently the classification category. GARD®Skin results predicted alkylated anisole to be a non-sensitizer, with individual replicate decision values consistently below zero. Given both *in silico* and *in vitro* results suggesting alkylated anisole is a non-sensitizer, and non-monotonic dose response from *in vivo* results, the overall WOE suggests alkylated anisole is unlikely to be a skin sensitizer.

Comparative computational bioactivity profiling

To put the biological activity of our test chemicals into a broader context, evidence associated with KEs in the skin sensitization AOP were identified from the peer-reviewed literature for comparative prototypical sensitizers, non-sensitizers, and irritating non-sensitizers (**Figure 3**). As an example of a prototypical skin sensitizer, hydroquinone consistently activated each key event in the skin sensitization AOP, where structural alerts, electrophilic chemistry and *in silico* peptide binding indicate that covalent binding of skin proteins occurs (KE1), *in vitro* tests indicate activation of keratinocytes (KE2) and dendritic cells (KE3) via h-CLAT and GARD®Skin (sensitizer, average decision value 6.89, see supplemental data), and LLNA studies demonstrate a proliferative T-cell response (KE4: EC3 ranged from 0.07-2% (Kimber et al 1998; Lee et al., 1999)) as well as an allergic response after challenge (AO: Guinea pig maximization test (Basketter and Scholes 1992, Goodwin, et al. 1981)). Additional prototypical sensitizers were included in the comparison to span a range of potencies from low/moderate (2-mercaptobenzothiazole) to extreme (cinnamal) as well as a variety of skin sensitization relevant mechanisms (e.g. SN2, Michael addition, alpha-beta unsaturated ketones, quinones, thiols/disulfides).

A representative irritating non-sensitizer, nonanoic acid, showed evidence of dendritic cell activation in two distinct measures (KE3: h-CLAT and GARD®Skin; weak sensitizer, average decision value of 1.15, see supplemental data) and a weak but dose-responsive T-cell proliferative response in the LLNA (KE4) (see supplemental data). Negative results were obtained in assays assessing the remaining KEs in the skin sensitization AOP (KE1: QSAR) and DPRA; KE2: Keratino-Sens™; AO: Guinea pig maximization assay and human maximization test (see supplemental data). Nonanoic acid produced strong to moderate erythema and edema in all animals during the induction phase of the guinea pig maximization assay (Johnson, et al. 2011).

A representative negative control, methyl salicylate did not activate any of the KE in the AOP: KE1, negative predictions in OASIS TIMES and DPRA; KE2, negative predictions in Keratino-Sens™; KE3, negative predictions in GARD®Skin (average decision value of -0.90, see supplemental data) and h-CLAT; KE4, negative result in; and AO, negative result in guinea pig maximization test and clinical experience suggesting that contact allergy is rare (Basketter, et al. 2014) (see supplemental data). Methyl salicylate does not meet the criteria for skin sensitization in any element of the AOP.

ToxPi™ ranking of biological activity based on normalized ToxPi™ statistics allowed unsupervised hierarchical clustering of chemicals of interest into groups. This revealed 3 main clusters of chemicals, including one cluster of prototypical sensitizers, one of irritating non-sensitizers, and a combined cluster including prototypical non-sensitizers and the three test chemicals (**Figure 3**). Key characteristics of bioactivity profiles within each group were similar with the exception of sodium lauryl sulfate within the irritating non-sensitizers group due to its positive responses in the DPRA, and electrophilic chemistry parameters (purple pie slices). Despite differences in structure, the test chemicals were clustered together, reflective of the qualitative similarities in their bioactivity profiles. Further, a high degree of similarity was observed between biological activity profiles of the test chemicals and prototypical non-sensitizers as well as pronounced differences between the test substances and prototypical sensitizers confirming the classification predictions made through a WOE assessment described above.

Discussion

Classification schemes often give strong preference to single animal studies such as the LLNA or guinea pig maximization assay. Considering multiple inputs collectively in a WOE approach can reduce the uncertainty associated with weakly positive results in individual animal studies through incorporation of mechanistic information. The cases presented here illustrate that consideration of the LLNA assay on its own would have warranted classification as skin sensitizers for each test chemical, however a WOE assessment using unsupervised clustering of AOP-anchored data points supports the lack of skin sensitization potential for these three chemicals.

In the case of the three chemicals of interest (tetrakis (2-ethylbutyl) orthosilicate, decyl palmitate, and alkylated anisole), weak or borderline positive results in the in-vivo LLNA were weighted against other available data. We concluded that the three test chemicals were considered unlikely to have skin sensitizing potential based on no genomic response indicative of skin sensitization, e.g. GARD®Skin, a lack of structural alerts and or applicable QSAR predictions, and the unsupervised clustering conducted in ToxPi™. Without a structural alert that ties to a known skin sensitization mechanism, it is difficult to place the weak positive LLNA results into context; however, several hypotheses were considered for the potential discordant results. The positive LLNA results may be related to non-specific proliferation of lymph node cells representing a false positive. As noted, none of the test chemicals had positive QSAR predictions for parent or possible metabolites by autoxidation consistent with formation of haptens, however neither tetrakis (2-ethylbutyl) orthosilicate nor alkylated anisole fit into the applicability domain for the QSAR models employed. It is possible that a test chemical can be capable of stimulating T-cell proliferation via a non-specific irritation response perhaps in the case of tetrakis (2-ethylbutyl) Orthosilicate which is supported by the indications of excessive local irritation observed. Alternatively, a CD1d lipid antigen type response has been shown to stimulate T-cell proliferation, a potential possibility in the case of decyl palmitate and alkylated anisole. To the extent of our knowledge, neither decyl palmitate nor alkylated anisole or their metabolites have been identified as CD1d ligands, but are expected to be metabolized and glucuronidated. The combination of a lipid moiety attached to a sugar are the organic functional groups that are present in established CD1d ligands (McEwen-Smith, et al. 2015). Alternatively, the ability of the test chemicals to produce an SI value above the respective test method threshold represents a “true” positive but the chemicals are unlikely to be allergenic in humans, possibly due to physiological differences between laboratory animals and man. Certainly, the hydrophobicity of the test chemicals in this case suggests they are more likely to become trapped in the stratum corneum lipids of human skin (which is thicker than in mice) and be unable to reach the epidermal proteins they would need to react with to induce an allergenic response (Tojo, et al. 1988).

To contextualize the weakly positive LLNA results further, additional data from skin sensitization assays were generated or identified in the public domain. A comparative class assessment of bioactivity profiles done via an unsupervised hierarchical clustering approach using ToxPi™, provided a non-expert judgement based class prediction of the skin sensitization potential for each of the three test chemicals. Prototypical chemicals representing classes of sensitizers, irritating/non-sensitizer, and non-irritating/non-sensitizer were included for this comparison. When all AOP-based inputs

were considered equally across multiple classes of skin sensitizers, the three test chemicals were clustered together with the non-irritating/non-sensitizing class of chemicals (Figure 3, purple dendrogram lines).

As part of our analysis, we chose to visualize the missing/inconclusive data as gray pie slices to clearly distinguish data gaps from negative data (Figure 3). It's possible that the higher number of inconclusive data points between the test chemicals and the non-irritating/non-sensitizing chemicals could be a contributing factor to the overall class predictions made using the ToxPi™ visualization. An alternative approach could be to reduce the weighting factor for assays that are more likely to produce false negatives for difficult to test substances, such as the DPRA or h-CLAT. In this way, inconclusive results could still be considered in the overall WOE instead of excluding them altogether. On the other hand, positive outcomes would not be discounted regardless of the likelihood to be a true positive or a false positive. Other options for handling missing data would involve imputing them with the mean, median or mode of each endpoint assay. This method appears to have a larger impact when ToxPi™ is used for prioritization of substances which was not the focus of this work (To, et al. 2018). Furthermore, we did not search for or include data for structural analogues of our test substances in this instance, however this could be easily implemented as additional data points in ToxPi™ given read-across is justifiable. In our experience, using ToxPi™ as a comparative visualization tool provided a flexible approach to integrate distinct lines of evidence to transparently inform class predictions regarding the skin sensitization potential of difficult to test substances in the situation when the LLNA results were weakly positive and the published DAs could not be strictly applied.

Recently published DAs are becoming widely recognized as first-tier testing strategies with equivalent or superior ability to predict skin sensitization potential in animals or in humans when compared to predictions from the LLNA (Kleinstreuer, et al. 2018). This study could not apply DAs due to the difficult to test nature of poorly solubility and UVCB substances in aqueous based test systems (OECD 2002). For example, tetrakis (2-ethylbutyl) orthosilicate was inconclusive both in the DPRA due to formation of precipitate (i.e., technical limitation) and in the h-CLAT due to a tendency for this assay to produce false negatives for chemicals with low aqueous solubility (i.e., predictive limitation) (Takenouchi, et al. 2013). The first published DA, the “2 out of 3” approach, requires that the DPRA and h-CLAT be used as measures of KE1 and KE3 (OECD 2021), respectively, thus precluding consideration of this DA for tetrakis (2-ethylbutyl) orthosilicate. The alternate DA, referred to as the Integrated Testing Strategy (ITS), provides an option to rely on QSAR predictions rather than assays related to KE1 (e.g. DPRA). However, the QSAR predictions were characterized as out of domain by the model for the substances herein. In our experience, the most likely outcome of application of DAs consisting of testing strategies that rely primarily on *in silico* or *in vitro* results to poorly soluble or UVCB is “inconclusive” leading to the need for additional data or modified approaches. Similar issues have been noted for difficult to test cosmetic ingredients, in particular the difficulties in assessing KE1, leading to modified testing strategies specifically for poorly soluble or UVCB substances that avoids reliance on DPRA or QSAR assessment of KE1 (Bergal, et al. 2020).

Careful consideration of DAs and how to integrate non-animal test methods into the current hazard classification scheme for skin sensitization is timely as they are under active discussion at the international level by the United Nations Subcommittee of Experts on GHS (see [UN/SCEGHS/31/INF.27/Rev.2](#)). Currently, classification for skin sensitization is based on a WOE approach of only human experience and animal test methods. Whether single non-animal test methods and/or their combinations as DAs will be adopted in a tiered approach (with less weight compared to human or animal data) or in an overall WOE assessment is unclear. As the UN GHS is meant to inform non-experts on a global scale, simple chemical-agnostic classification schemes are intended to drive harmonization of hazard communication. Whatever scheme is adopted should allow for flexibility to incorporate all available data and overcome chemical-specific challenges or technical limitations of singular assays. Even though GHS has not adopted these concepts yet, some competent authorities already have developed guidance where the totality of available information is considered in order to apply WOE (ECHA 2017a, ECHA 2017b). In the case studies presented here, the final classification would be different if a tiered approach were followed vs a flexible WOE. Recognizing the likely broad implementation of these approaches in the future regulatory landscape, it should be noted that there has been little discussion on the applicability of these approaches to mixtures.

Similar to complex substances, mixtures (e.g. lubricants, greases) can share the same complexity and physical-chemical issues observed for our test cases and deserves further study.

Ultimately our work aims to highlight the challenges of applying context to borderline or weak LLNA responses with additional mechanistic information in the case of difficult to test substances. This initial work could have implications for future regulatory acceptance of broader assessment approaches. Each assay on its own has variability and limitations, but that does not preclude their utility to predict hazard. We demonstrate that unsupervised clustering, where the data inputs are anchored by the biological response pathway, provides further support for the individual chemical WOE prediction for a sensitizing/non-sensitizing class. This combined WOE and clustering approach maximizes the contribution of any data, whether derived from *in vitro*, *in vivo* or *in silico* models, to the overall assessment of skin sensitization potential for difficult to test substances.

DRAFT

Figure 1

463

464 **Figure 1.** Skin Sensitization AOP. The top flow chart conceptualizes the molecular events starting a chemical being
 465 introduced to skin resulting in molecular and cellular events that lead to tissue level manifestation of sensitization. The
 466 middle section highlights the key events (KE) starting with KE1- covalent binding of the chemical to a protein creating an
 467 antigen that is recognized by cells in KE2 and KE3 which elicit a response from T-Cells in KE4. Confounding factors include
 468 physical chemical properties or visually similar pathology that would hinder usefulness of diagnostic tests and
 469 assessment of classification of skin sensitization

470

471 **Table 1.** Advantages/specific limitations of skin sensitization test methods

Key Event	Relevant Test Methods	Advantages	Specific Limitations
Covalent Protein Binding	(Q)SAR/Structural Alerts	Non-test method	Limited Applicability Domain
	OECD442C: DPRA	Non-animal test method	Not suitable for UVCBs
	Electrophilic Chemistry	Non-test method	Limited by water solubility Limited predictability
Keratinocyte Activation	OECD TG 442D: Keratino-Sens™ /LuSens	Non-animal test method	Limited by water solubility
Dendritic Cell Activation	OECD TG 442E: h-CLAT	Non-animal test method	Limited by water solubility
T-Cell Proliferation	OECD TG429: LLNA	In-vivo results with less time/animals	False positives due to irritation or specific chemistry
	OECD TG 442B: LLNA-BrdU		
Elicitation	OECD TG 406: Buehler Assay/ GPMT	Measures adverse outcome	Animal/time intensive

472

473

474

475

Table 2. Stimulation Index (SI) values and average support vector machine (SVM) decision values for case study chemicals in the LLNA and GARD®Skin assay, respectively. Cells highlighted in green are above the associated threshold for a positive response. No data points were excluded from the analyses.

Dose (%)	Decyl Palmitate ^a	Alkylated Anisole ^a	Tetrakis Orthosilicate ^b	Decyl palmitate	Alkylated anisole	Tetrakis Orthosilicate
LLNA SI Mean ± S.D. (N)				GARD®Skin SVM Mean ± S.D. (N)		
10	1.74 ± 0.9 (5)		1.1 ± 0.9 (5)			
25	2.54 ± 1.5 (5)	2.04 ± 0.7 (5)	2.5 ± 0.5 (5)	-1.16 ± 0.4 (3)	-1.61 ± 0.7 (3)	-1.07 ± 0.7 (3)
50	3.18 ± 1.0 (5)	5.30 ± 4.4 (5)	4.9 ± 0.9 (5) ^c	Non-sensitizer	Non-sensitizer	Non-sensitizer
100		3.39 ± 2.0 (5)				

476

a= threshold for OECD TG429: LLNA ³H-methyl thymidine SI = 3

477

b= threshold for OECD TG 442B: LLNA BrdU SI = 1.6

478

c= excessive local irritation observed at this dose; mean ear thickness (day 6-day 1) > 25%

Figure 2

479

480

481

482

483

484

485

Figure 2. Definitions and notation for ToxPi™. Weighted combinations of data were integrated for each chemical for each endpoint with overall ToxPi™ profiles composed of component slices for one or more endpoints. Key events within the Skin Sensitization AOP are depicted by slices of a given color family: Purple, Key Event 1- covalent protein binding; Red, Key Event 2- dendritic cell activation; Blue, Key Event 3- keratinocyte activation; Orange, Key Event 4- T-cell proliferation; Yellow, Adverse Outcome-Elicitation; Green, Confounding Factors. The inner circle represents the percentage of missing data for a given component slice (darker = more missing data).

Figure 3

Figure 3. Hierarchical Clustering of ToxPi™ Profiles. ToxPi™ profiles of case study chemicals (upper left quadrant, purple dendrogram lines) compared to prototypical skin sensitizers (orange dendrogram lines), prototypical non-sensitizers (lower left quadrant, purple dendrogram lines), and irritating non-sensitizers (green dendrogram lines). Missing or inconclusive data values for each substance are represented as gray pie slices; blank pie slices are tested negative data points.

References

- Adenuga D, Woolhiser MR, Gollapudi BB, Boverhof DR. 2012. Differential Gene Expression Responses Distinguish Contact and Respiratory Sensitizers and Nonsensitizing Irritants in the Local Lymph Node Assay. *Toxicological Sciences* **126**:413-425. DOI: 10.1093/toxsci/kfs071.
- Anderson SE, Siegel PD, Meade BJ. 2011. The LLNA: A Brief Review of Recent Advances and Limitations. *Journal of allergy* **2011**:424203. DOI: 10.1155/2011/424203.
- Ashikaga T, Yoshida Y, Hirota M, Yoneyama K, Itagaki H, Sakaguchi H, Miyazawa M, Ito Y, Suzuki H, Toyoda H. 2006. Development of an in vitro skin sensitization test using human cell lines: The human Cell Line Activation Test (h-CLAT): I. Optimization of the h-CLAT protocol. *Toxicology in Vitro* **20**:767-773. DOI: 10.1016/j.tiv.2005.10.012.
- Basketter DA, Alépée N, Ashikaga T, Barroso J, Gilmour N, Goebel C, Hibatallah J, Hoffmann S, Kern P, Martinozzi-Teissier S, Maxwell G, Reisinger K, Sakaguchi H, Schepky A, Tailhardat M, Templier M. 2014. Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis : contact, atopic, occupational, drug* **25**:11-21. DOI: 10.1097/DER.0000000000000003.
- Basketter DA, Gerberick GF, Kimber I. 1998. Strategies for identifying false positive responses in predictive skin sensitization tests. *Food Chem Toxicol* **36**:327-333. DOI: 10.1016/s0278-6915(97)00158-0.
- Basketter DA, Scholes EW. 1992. Comparison of the local lymph node assay with the guinea-pig maximization test for the detection of a range of contact allergens. *Food and Chemical Toxicology* **30**:65-69. DOI: 10.1016/0278-6915(92)90138-B.
- Bergal M, Puginier M, Gerbeix C, Groux H, Roso A, Cottrez F, Milius A. 2020. In vitro testing strategy for assessing the skin sensitizing potential of "difficult to test" cosmetic ingredients. *Toxicology in vitro : an international journal published in association with BIBRA* **65**:104781. DOI: 10.1016/j.tiv.2020.104781.
- Buehler EV. 1965. Delayed contact hypersensitivity in the guinea pig *Archives of dermatology* **91**:171-177. DOI: 10.1001/archderm.1965.01600080079017.
- Corsini E, Clewell R, Cotgreave I, Eskes C, Kopp-Schneider A, Westmoreland C, Alves PM, Navas JM, 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, ISBN 978-92-76-40345-6, doi:10.2760/626728, JRC125963., 10.2760/626728. DOI: 10.2760/626728.
- Cortes C, Vapnik V. 1995. Support-vector networks. *Machine Learning* **20**:273-297. DOI: 10.1007/BF00994018.
- Dimitrov S, Detroyer A, Piroird C, Gomes C, Eilstein J, Pauloin T, Kuseva C, Ivanova H, Popova I, Karakolev Y, Ringeissen S, Mekenyan O. 2016. Accounting for data variability, a key factor in in vivo/in vitro relationships: application to the skin sensitization potency (in vivo LLNA versus in vitro DPRA) example. *Journal of Applied Toxicology* **36**:1568-1578. DOI: 10.1002/jat.3318.

ECETOC. 2003. Technical Report No. 87: Contact Sensitisation: Classification According to Potency. <https://www.ecetoc.org/wp-content/uploads/2014/08/ECETOC-TR-087.pdf>.

ECHA. 2017a. Guidance on Information Requirements and Chemical Safety Assessment: Chapter R.7a: Endpoint specific guidance, Version 6.0. https://www.echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf, 10.2823/337352. DOI: 10.2823/337352.

ECHA. 2017b. Guidance on the Application of the CLP Criteria. Guidance to Regulation (EC) No 1272/2008 on classification, labelling, and packaging (CLP) of substances and mixtures. Version 5.0. ECHA-17-G-21-EN., 10.2823/124801. DOI: 10.2823/124801.

EURL-ECVAM. 2012. Direct Peptide Reactivity Assay (DPRA) ECVAM Validation Study Report. Accessible at <https://tsar.jrc.ec.europa.eu/system/files/Published/DPRA%20Validation%20Study%20Report.pdf>. Last accessed on 13 Aug 2021.

Forreryd A, Zeller KS, Lindberg T, Johansson H, Lindstedt M. 2016. From genome-wide arrays to tailor-made biomarker readout—Progress towards routine analysis of skin sensitizing chemicals with GARD. *Toxicology in Vitro* **37**:178-188. DOI: 10.1016/j.tiv.2016.09.013.

Gerberick GF. 2016. The Use of Peptide Reactivity Assays for Skin Sensitisation Hazard Identification and Risk Assessment. *Alternatives to Laboratory Animals* **44**:437-442. DOI: 10.1177/026119291604400506.

Gilmour N, Kimber I, Williams J, Maxwell G. 2019. Skin sensitization: Uncertainties, challenges, and opportunities for improved risk assessment. *Contact dermatitis* **80**:195-200. DOI: 10.1111/cod.13167.

Goodwin BFJ, Crevel RWR, Johnson AW. 1981. A comparison of three guinea-pig sensitization procedures for the detection of 19 reported human contact sensitizers. *Contact Dermatitis* **7**:248-258. DOI: 10.1111/j.1600-0536.1981.tb04060.x.

Grimm FA, House JS, Wilson MR, Sirenko O, Iwata Y, Wright FA, Ball N, Rusyn I. 2019. Multi-dimensional in vitro bioactivity profiling for grouping of glycol ethers. *Regulatory Toxicology and Pharmacology* **101**:91-102. DOI: 10.1016/j.yrtph.2018.11.011.

Grimm FA, Iwata Y, Sirenko O, Chappell GA, Wright FA, Reif DM, Braisted J, Gerhold DL, Yeakley JM, Shepard P. 2016. A chemical–biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. *Green chemistry* **18**:4407-4419. DOI: 10.1039/c6gc01147k.

Hoffmann S. 2015. LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. *ALTEX-Alternatives to animal experimentation* **32**:379-383. DOI: 10.14573/altex.1505051.

Hoffmann S, Kleinstreuer N, Alépée N, Allen D, Api AM, Ashikaga T, Clouet E, Cluzel M, Desprez B, Gellatly N, Goebel C, Kern PS, Klaric M, Kühnl J, Lalko JF, Martinozzi-Teissier S, Mewes K, Miyazawa M, Parakhia R, van Vliet E, Zang Q,

Petersohn D. 2018. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. *Critical reviews in toxicology* **48**:344-358. DOI: 10.1080/10408444.2018.1429385.

ICCVAM. 1999. The Murine Local Lymph Node Assay: a Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds, The results of an independent peer review evaluation coordinated by the ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). National Institute of Environmental Health Sciences, NIH Publication no. 99-4494, <http://www.iccvam.niehs.nih.gov/> *Fed Regist* **64**:14006-14007.

Johansson H, Albrekt A-S, Borrebaeck CA, Lindstedt M. 2013. The GARD assay for assessment of chemical skin sensitizers. *Toxicology in vitro* **27**:1163-1169. DOI: 10.1016/j.tiv.2012.05.019.

Johansson H, Gradin R, Johansson A, Adriaens E, Edwards A, Zuckerstätter V, Jerre A, Burleson F, Gehrke H, Roggen EL. 2019. Validation of the GARD™skin Assay for Assessment of Chemical Skin Sensitizers: Ring Trial Results of Predictive Performance and Reproducibility. *Toxicological Sciences* **170**:374-381. DOI: 10.1093/toxsci/kfz108.

Johansson H, Lindstedt M, Albrekt AS, Borrebaeck CA. 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, Rydnert F, Kühnl J, Schepky A, Borrebaeck C, Lindstedt M. 2014. Genomic allergen rapid detection in-house validation--a proof of concept. *Toxicological sciences : an official journal of the Society of Toxicology* **139**:362-370. DOI: 10.1093/toxsci/kfu046.

Johnson W, Heldreth, Bart, Bergfeld WF, Belsito DV, Klaassen CD, Hill R, Liebler D, Marks JG, Shank RC, Slaga TJ, Snyder PW, Andersen FA. 2011. Final Report of the Cosmetic Ingredient Review Expert Panel on the Safety Assessment of Pelargonic Acid (Nonanoic Acid) and Nonanoate Esters. *International Journal of Toxicology* **30**:228S-269S. DOI: 10.1177/1091581811428980.

Kimber I, Dearman RJ, Scholes EW, Basketter DA. 1994. The local lymph node assay: developments and applications. *Toxicology* **93**:13-31. DOI: 10.1016/0300-483X(94)90193-7.

Kimber I, Hilton J, Dearman RJ, Gerberick GF, Ryan CA, Basketter DA, Scholes EW, Ladics GS, Loveless SE, House RV, et al. 1995. An international evaluation of the murine local lymph node assay and comparison of modified procedures. *Toxicology* **103**:63-73. DOI: 10.1016/0300-483x(95)03114-u.

Kleinstreuer NC, Hoffmann S, Alépée N, Allen D, Ashikaga T, Casey W, Clouet E, Cluzel M, Desprez B, Gellatly N, Göbel C, Kern PS, Klaric M, Kühnl J, Martinozzi-Teissier S, Mewes K, Miyazawa M, Strickland J, van Vliet E, Zang Q, Petersohn D. 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches (*). *Critical reviews in toxicology* **48**:359-374. DOI: 10.1080/10408444.2018.1429386.

Kolle SN, Basketter DA, Casati S, Stokes WS, Strickland J, van Ravenzwaay B, Vohr H-W, Landsiedel R. 2013. Performance standards and alternative assays: Practical insights from skin sensitization. *Regulatory Toxicology Pharmacology* **65**:278-285. DOI: 10.1016/j.yrtph.2012.12.006.

LoPachin RM, Gavin T, DeCaprio A, Barber DS. 2012. Application of the hard and soft, acids and bases (HSAB) theory to toxicant–target interactions. *Chemical research in toxicology* **25**:239-251. DOI: 10.1021/tx2003257.

LoPachin RM, Geohagen BC, Nordstroem LU. 2019. Mechanisms of soft and hard electrophile toxicities. *Toxicology* **418**:62-69. DOI: 10.1016/j.tox.2019.02.005.

Magnusson B, Kligman AM. 1969. The identification of contact allergens by animal assay. The guinea pig maximization test. *J. Invest. Dermatol.* **52**:268-276. DOI: 10.1038/jid.1969.42.

Marvel SW, To K, Grimm FA, Wright FA, Rusyn I, Reif DM. 2018. ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models. *BMC bioinformatics* **19**:1-7. DOI: 10.1186/s12859-018-2089-2.

McEwen-Smith RM, Salio M, Cerundolo V. 2015. CD1d-dependent endogenous and exogenous lipid antigen presentation. *Current Opinion in Immunology* **34**:116-125. DOI: 10.1016/j.coi.2015.03.004.

McNamee PM, Api AM, Basketter DA, Frank Gerberick G, Gilpin DA, Hall BM, Jowsey I, Robinson MK. 2008. A review of critical factors in the conduct and interpretation of the human repeat insult patch test. *Regulatory Toxicology and Pharmacology* **52**:24-34. DOI: 10.1016/j.yrtph.2007.10.019.

Mekenyan O, Roberts DW, Karcher W. 1997. Molecular Orbital Parameters as Predictors of Skin Sensitization Potential of Halo- and Pseudohalobenzenes Acting as SNAr Electrophiles. *Chemical Research in Toxicology* **10**:994-1000. DOI: 10.1021/tx960104g.

Montelius J, Wahlkvist H, Boman A, Wahlberg J. 1998. Murine Local Lymph Node Assay for Predictive Testing of Allergenicity: Two Irritants Caused Significant Proliferation. *Acta Dermato-Venereologica* **78**:433-437. DOI: 10.1080/000155598442728.

Natsch A, Bauch C, Foertsch L, Gerberick F, Norman K, Hilberer A, Inglis H, Landsiedel R, Onken S, Reuter H, Schepky A, Emter R. 2011. The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers in vitro: Results of a ring-study in five laboratories. *Toxicology in Vitro* **25**:733-744. DOI: 10.1016/j.tiv.2010.12.014.

OECD. 2002. *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures*, OECD Publishing, Paris, , 10.1787/9789264078406-en. DOI: 10.1787/9789264078406-en.

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*

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*

OECD. 2020. *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

OECD. 2021. Guideline No. 497: Defined Approaches on Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 10.1787/b92879a4-en. DOI: 10.1787/b92879a4-en.

Ouyang Q, Wang L, Mu Y, Xie X-Q. 2014. Modeling skin sensitization potential of mechanistically hard-to-be-classified aniline and phenol compounds with quantum mechanistic properties. *BMC Pharmacology and Toxicology* **15**:76. DOI: 10.1186/2050-6511-15-76.

Patlewicz G, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MIH, Aptula AO, Phillips RD, Niemelä J, Madsen C, Wedebye EB, Roberts DW, Bailey PT, Mekenyan OG. 2007. TIMES-SS—A promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. *Regulatory Toxicology and Pharmacology* **48**:225-239. DOI: 10.1016/j.yrtph.2007.03.003.

Ramirez T, Mehling A, Kolle SN, Wruck CJ, Teubner W, Eltze T, Aumann A, Urbisch D, van Ravenzwaay B, Landsiedel R. 2014. LuSens: A keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification. *Toxicology in Vitro* **28**:1482-1497. DOI: 10.1016/j.tiv.2014.08.002.

Roberts DW, Patlewicz G, Dimitrov SD, Low LK, Aptula AO, Kern PS, Dimitrova GD, Comber MIH, Phillips RD, Niemelä J, Madsen C, Wedebye EB, Bailey PT, Mekenyan OG. 2007. TIMES-SS—A Mechanistic Evaluation of an External Validation Study Using Reaction Chemistry Principles. *Chemical Research in Toxicology* **20**:1321-1330. DOI: 10.1021/tx700169w.

Schneider K, Akkan Z. 2004. Quantitative relationship between the local lymph node assay and human skin sensitization assays. *Regulatory toxicology and pharmacology : RTP* **39**:245-255. DOI: 10.1016/j.yrtph.2004.02.002.

Strickland J, Zang Q, Kleinstreuer N, Paris M, Lehmann DM, Choksi N, Matheson J, Jacobs A, Lowit A, Allen D, Casey W. 2016. Integrated decision strategies for skin sensitization hazard. *Journal of Applied Toxicology* **36**:1150-1162. DOI: <https://doi.org/10.1002/jat.3281>.

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. *The Journal of Toxicological Sciences* **38**:599-609. DOI: 10.2131/jts.38.599.

To KT, Fry RC, Reif DM. 2018. Characterizing the effects of missing data and evaluating imputation methods for chemical prioritization applications using ToxPi. *BioData mining* **11**:10. DOI: 10.1186/s13040-018-0169-5.

Tojo K, Chiang CC, Doshi U, Chien YW. 1988. Stratum Corneum Reservoir Capacity Affecting Dynamics of Transdermal Drug Delivery. *Drug Development and Industrial Pharmacy* **14**:561-572. DOI: 10.3109/03639048809151884.

United Nations. 2019. Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Eighth revised edition, New York and Geneva, United Nations Publications. Available at: <https://unece.org/ghs-rev8-2019>.

Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. 2015. Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regulatory Toxicology and Pharmacology* **71**:337-351. DOI: 10.1016/j.yrtph.2014.12.008.

Wang C-C, Lin Y-C, Wang S-S, Shih C, Lin Y-H, Tung C-W. 2017. SkinSensDB: a curated database for skin sensitization assays. *Journal of cheminformatics* **9**:1-6. DOI: 10.1186/s13321-017-0194-2.

DRAFT